

Genomewide transcriptional signatures of migratory flight activity in a globally invasive insect pest

CHRISTOPHER M. JONES,* ALEXIE PAPANICOLAOU,† GEORGE K. MIRONIDIS,‡
JOHN VONTAS,‡ § YIHUA YANG,¶ KA S. LIM,* JOHN G. OAKESHOTT,** CHRIS BASS† †¹ and
JASON W. CHAPMAN* † †¹

*AgroEcology, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK, †Hawkesbury Institute for the Environment, University of Western Sydney, Richmond, NSW 2753, Australia, ‡Institute of Molecular Biology & Biotechnology, Foundation for Research & Technology Hellas, 100 N. Plastira Street, GR-700 13 Heraklion Crete, Greece, §Laboratory of Pesticide Science, Department of Crop Science, Agricultural University of Athens, 75 Iera Odos Street, GR-11855 Athens, Greece, ¶College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China, **CSIRO Ecosystems Sciences, Black Mountain, Clunies Ross Street, Canberra, ACT 0200, Australia, ††Biological Chemistry and Crop Protection, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK, †††Environment and Sustainability Institute, University of Exeter, Penryn, Cornwall TR10 9EZ, UK

Abstract

Migration is a key life history strategy for many animals and requires a suite of behavioural, morphological and physiological adaptations which together form the ‘migratory syndrome’. Genetic variation has been demonstrated for many traits that make up this syndrome, but the underlying genes involved remain elusive. Recent studies investigating migration-associated genes have focussed on sampling migratory and nonmigratory populations from different geographic locations but have seldom explored phenotypic variation in a migratory trait. Here, we use a novel combination of tethered flight and next-generation sequencing to determine transcriptomic differences associated with flight activity in a globally invasive moth pest, the cotton bollworm *Helicoverpa armigera*. By developing a state-of-the-art phenotyping platform, we show that field-collected *H. armigera* display continuous variation in flight performance with individuals capable of flying up to 40 km during a single night. Comparative transcriptomics of flight phenotypes drove a gene expression analysis to reveal a suite of expressed candidate genes which are clearly related to physiological adaptations required for long-distance flight. These include genes important to the mobilization of lipids as flight fuel, the development of flight muscle structure and the regulation of hormones that influence migratory physiology. We conclude that the ability to express this complex set of pathways underlines the remarkable flexibility of facultative insect migrants to respond to deteriorating conditions in the form of migratory flight and, more broadly, the results provide novel insights into the fundamental transcriptional changes required for migration in insects and other taxa.

Keywords: insect migration, migratory genomics, tethered flight, transcriptomics

Received 4 June 2015; revision received 30 July 2015; accepted 24 August 2015

Introduction

The ability to initiate and sustain periods of long-distance flight is a prerequisite for the billions of insects

that migrate each year (Chapman *et al.* 2012, 2015). The behavioural, physiological and morphological adaptations necessary to undertake such flights form part of a much larger inherited ‘migratory syndrome’ which is present not only across the Insecta but also in other animal migrants such as birds and fish (Roff & Fairbairn 2007; Dingle 2014). Quantitative trait analyses in a wide

Correspondence: Christopher M. Jones, Fax: +44(0)1582 760981; E-mail: +christopher.jones@rothamsted.ac.uk

¹These authors contributed equally.

range of taxa have shown that there is significant phenotypic and genetic variation in the individual traits that comprise the migratory syndrome (Pulido *et al.* 1996; Roff & Fairbairn 2007). The genes and associated biochemical pathways that underpin this variation, however, remain poorly understood.

A handful of comparative transcriptomic and genomic analyses have started to uncover differences in gene expression and single nucleotide polymorphisms (SNPs) between migratory and nonmigratory populations of the same species (Jones *et al.* 2008; Zhu *et al.* 2009; Postel *et al.* 2010; Zhan *et al.* 2014; McKinney *et al.* 2015). For example, signatures of positive selection acting upon flight muscle genes were detected in migratory populations of the monarch butterfly, *Danaus plexippus* (Zhan *et al.* 2014). Few studies have applied next-generation sequencing to migratory insects from the same population displaying intraspecific variation in a behavioural migratory trait. Such an approach requires a clearly defined and accurate quantification of the focal trait (Liedvogel *et al.* 2011). The propensity to engage in long-distance flight, a crucial proxy for migratory potential, can be quantified under controlled conditions using computerized tethered flight mills. Flight activity has been characterized in several insect species in this way (Colvin & Gatehouse 1993; Kent & Rankin 2001; Dorhout *et al.* 2008; Liu *et al.* 2011), but no study has yet utilized this experimental system to provide robust migratory phenotypes and then use them for downstream genotyping and gene expression analysis.

Comparative methods for determining migration-associated genes rely on a model species that exhibits a variety of migratory phenotypes. Members of the noctuid moth family are some of the most economically damaging pests of agriculture with the capacity for long-range movement a major contributing factor to their pest status. By far the most important species in this group is the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae), which has an extremely large host range and is distributed globally. *H. armigera* is a facultative migrant, moving in response to deteriorating local conditions and, like other noctuid moths (Chapman *et al.* 2008, 2010), they utilize favourable high-altitude wind currents to optimize their movement direction and ground speed (Feng *et al.* 2009). Evidence for these movements is supported from entomological radar, insect trapping and geneflow studies (Zhou *et al.* 2000; Scott *et al.* 2005; Feng *et al.* 2009). Not all populations of *H. armigera* are migratory, however, and the migratory syndrome may only be expressed in certain years depending on the local conditions (Scott *et al.* 2005). This plasticity makes *H. armigera* an attractive species for studying the genetic basis of migration as a

variety of phenotypes is likely to persist in natural populations.

To dissect the molecular basis of flight in a migratory insect, we focus here on the propensity for long-distance flight and use an unique integrative approach of tethered flight and transcriptomics to (i) quantify the intraspecific variation in flight activity during the migratory phase of recently collected field populations of *H. armigera* from different geographic origins, (ii) identify the differentially expressed suite of genes between flight phenotypes that contribute to the migratory capacity of *H. armigera* and (iii) validate the expression of these genes in independently flown *H. armigera* adults.

Materials and methods

H. armigera collections

The adult *H. armigera* used in the flight mill, RNA-seq and qPCR experiments were collected from Bt cotton from five populations in China [Dafeng (Jiangsu province), Anyang (Henan province), Jingzhou, (Hubei province), Qiuxian (Hebei province) and Wanjiang (Anhui province)] and a single site in northern Greece during the summer of 2013 and shipped to Rothamsted Research, UK (Table S1, Supporting information). Eggs were collected from the field populations of Dafeng and Jingzhou and reared for a further two generations. Anyang and Qiuxian adults were collected in light traps and reared for one subsequent generation in the laboratory. *H. armigera* from China were considered separate 'populations' for the study although gene flow between these areas may exist given the long-distance migratory ability of *H. armigera* in China (Feng *et al.* 2009). *H. armigera* from cotton fields in northern Greece (41°N, 023°E) were collected as 4th-5th larval instars and reared for one generation in the laboratory. An additional long-term laboratory strain, Bayer (courtesy of the Max Planck Institute, Jena, Germany), was used in the qPCR experiments. Insects were reared under a constant light regime of L:D 14:10 at 26 ± 1 °C, and the flight mill experiments conducted under the same conditions. Larvae were reared individually in 37-ml clear plastic polypots containing a wheat germ artificial diet (Vanderzant *et al.* 1962).

Tethered flight mills

Flight experiments were conducted in October and November 2013 using newly built computerized tethered flight mills designed and housed at Rothamsted Research (Fig. 1A). In brief, the flight arm is made from twisted wire with the vertical axis secured between two

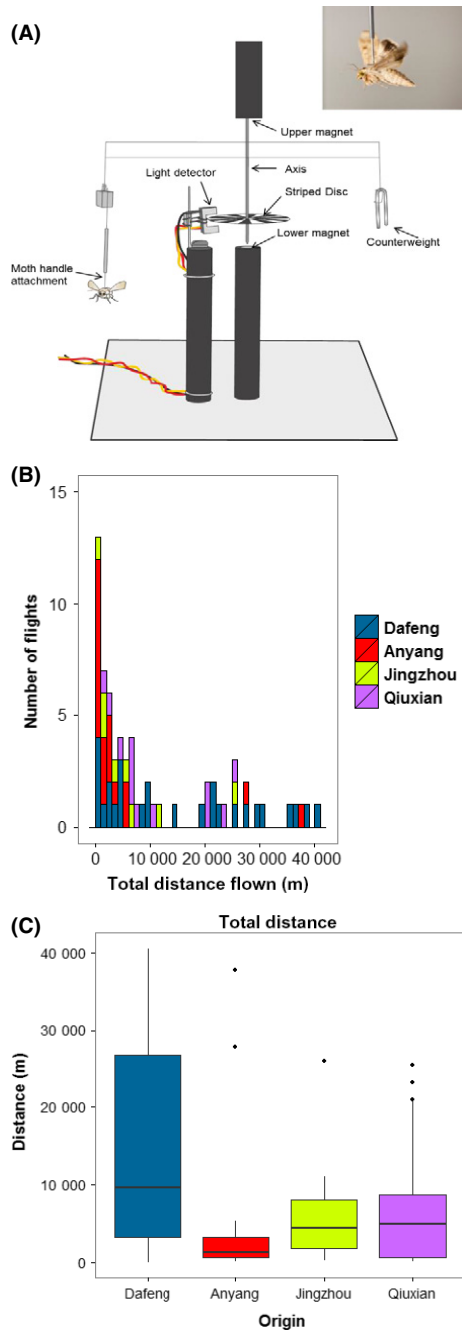


Fig. 1 Variation in flight activity of adult *H. armigera* from China characterized using tethered flight. (A) Diagram of tethered flight mill system for quantifying flight behaviour in noctuid moths. Upper panel shows single *H. armigera* female attached to the handle 'in-flight'. (B) Variation in the total distance flown by the four Chinese populations. Each bin represents 1000 metres flown during the course of a single night. The origin of each insect is given according to the colour code highlighted. (C) Box plots of the total distance flown by the four populations from China (Dafeng, $n = 27$; Anyang, $n = 19$; Jingzhou, $n = 8$; Qiuxian, $n = 19$).

magnets. The insect is attached to a pin that fits into a sleeve suspended from the flight arm. This allows the insect to fly rotationally in a horizontal plane. A sensor detects a rotational disc so that the speed and time of the rotation (and hence the powered flight of the insect) are recorded. The time period between insect collection and quantifying flight activity (3 months) ensured that we captured natural flight behaviour in the shortest time possible. Up to 16 insects were flown during a single night. Each morning pupae were checked for emergence, and any adults were put aside for flight mill experiments. Any moths with damaged wings or unhealthy were not chosen for experiments. All moths were flown the night post-eclosion (1 day old) to capture prereproductive behaviour (Colvin & Gatehouse 1993). Experimental insects were sedated at 4 °C for a minimum of 2 h prior to tethering. Scales were removed from the thorax and a small amount of adhesive glue applied to both a c.60-mg pin and the exposed thorax. Prior to the flight, each insect was provided with a c.20% honey solution ad libitum. All insects were randomly assigned to a flight mill before 1900 each evening. A paper platform was provided to allow the moth to rest prior to undertaking its initial flight. Data were collected between 1900 and 0830 with the 10-h dark/night cycle running between 2000 and 0600. At 0900, the following day, the insects were taken off the flight mills and allowed to recover for 1–2 h before being snap-frozen in liquid nitrogen and stored at –80 °C until RNA extraction. Any insects which looked damaged, unhealthy or had escaped from their pin were disregarded from further analyses. In an initial experiment, male and female adult *H. armigera* from four Chinese populations (Dafeng, Jingzhou, Qiuxian and Anyang) were flown simultaneously over the course of a single week to determine intraspecific population differences. Insects from Wangjiang, Greece and Bayer were flown in independent experiments.

REML analysis of total distance flown

To explore the relationship between the total distance flown by Chinese *H. armigera* and the explanatory variables, *population* and *sex*, we computed variance components using a restricted maximum likelihood (REML) approach. *Population* and *sex* were considered fixed effects, while the night the insect was flown (*night*) and the mill used (*mill*) were treated as random effects. Using the initial (full) fixed model, the individual random components were sequentially dropped from the model and the deviance between subsequent 'simpler' models compared. After determining the variance

model, the fixed terms were sequentially dropped to provide the explanatory variables that were associated with the total distance flown (F statistic, $P < 0.05$). Following the observation that insects from Dafeng flew the furthest distance on the flight mills and nonoverlapping confidence intervals from the model predictions (Fig. S1, Supporting information), a similar REML approach was used to determine whether insects from Dafeng flew significantly further than the three other populations. Total distance flown was squared-root-transformed prior to the statistical analysis. The outputs from the full REML models are provided in Table S2 and Table S3 (Supporting information). All analyses were conducted in GENSTAT 16th edition.

RNA extraction and sequencing

Total RNA was extracted from three whole individual adult females per sample using the ISOLATE II RNA Mini Kit (Bioline). Each flight phenotype consisted of three biological samples (nine insects per phenotype) which were chosen prior to extraction according to the total distance flown. RNA was sent to The Genome Analysis Centre (TGAC, BBSRC, Norwich, UK) for library construction (Illumina TrueSeq Library construction) and sequencing.

Mapping RNA-seq reads to *H. armigera* genome

Paired-end (PE) read 100-bp sequencing was performed on six samples from each experiment on half a flow cell of the Illumina HiSeq 2500. This yielded $c.14.4$ – 27.8 million PE reads per sample. Alignment, correction of illumina and isoform bias, FPKM estimation and TMM normalization were performed using DEW (<http://dew.sourceforge.net/>). Reads from each sample were mapped to the genes of the *H. armigera* genome to find those transcripts with $\geq 95\%$ amino acid identity and no more than 10% length difference (CD-HIT -s 0.90 -c 0.95). The raw reads were then realigned to these transcripts using GSNAP (Wu & Nacu 2010) and the alignment edited to correct for illumina bias and multi-isoforms using eXpress (Roberts & Pachter 2012). A total of 13 776 transcripts had at least 4 mapped reads in at least one sample (80.97%). The number of reads mapping to Greece and China samples was 12,255 (72.09%) and 13,289 (78.17%), respectively.

Differential expression

To detect differential gene expression between flight phenotypes, we used two expression packages edgeR and DESeq2 (Anders *et al.* 2013). Both packages use the

negative binomial model for analysing RNA-seq count data but differ in their estimation of gene dispersal. Using both methods, we aimed to capture as many differential transcripts as possible. The China and Greece data sets were analysed separately and all genes that were statistically significant at a false discovery rate (FDR) < 0.1 were considered candidates for flight performance. For edgeR, any transcript with less than five reads for all samples were removed from the analysis. All analyses were performed in R software (R Core Development Team 2014). An enrichment analysis (Fisher's exact test) of GO terms for differentially expressed transcripts (upregulated and downregulated genes tested separately) was performed against the reference gene set in Blast2GO at an FDR < 0.05 (Götz *et al.* 2008).

Quantitative PCR

A subset of differentially expressed genes ($N = 16$) were tested using qPCR to validate the RNA-Seq data and to independently test expression in additional *H. armigera* flown on the flight mills. The same samples from RNA-seq were used for Dafeng, Anyang and Greece. For the additional samples, total RNA was extracted from flight phenotypes from Wangjiang and Bayer as described above. Total RNA (600 ng) was converted into cDNA using oligo(dT)₂₀ (Invitrogen) and SuperScript III (Invitrogen). Exon-exon spanning primers for target genes were designed using Primer-BLAST (NCBI) (Ye *et al.* 2012). Two previously designed control genes, *elongation factor 1- α* and *β -actin* (Wang *et al.* 2013; Yan *et al.* 2013), were used for internal normalization. The specificity, dynamic range and PCR efficiency of each primer set were determined by testing against a fivefold serial dilution of cDNA (from 1/10th to 1/6000th). Details of all primers are given in Table S4 (Supporting information). All PCRs (20 μ l) were run on the RotorGene 6000 (Qiagen) using 300 nm of each primer, 10 μ l SYBR Green JumpStart Taq ReadyMix (Sigma Aldrich) and 5 μ l of cDNA diluted 50-fold. PCR conditions used throughout were 95 °C for 2 min followed by 40 cycles of 95 °C for 10 s, 57 °C for 15 s and 72 °C for 10s. Melt curves were run after amplification to check for specificity. Reactions were run in duplicate with no template controls (distilled water) for each primer set. Data were preprocessed to remove outliers (SD of technical replicates > 0.5) and late Ct values (> 35). Expression levels were calculated according to the ddCt method following correction for PCR efficiency and normalization for the two control genes (Schmittgen & Livak 2008). qPCR experiments followed best practices according to the MIQE guidelines (Bustin *et al.* 2009).

Results and discussion

Continuous variation in flight activity of H. armigera

The nocturnal flight activity of *H. armigera* was characterized using a newly developed state-of-the-art tethered flight mill system (UK Patent Application No. 1314415.9; Fig. 1A). Each insect was flown for a single night (from 19:00 to 08:30 GMT) in a controlled environment with the speed, distance and duration of all flights electronically recorded. In the first series of experiments, adult moths originating from four field populations from across China were flown shortly after collection ($N = 73$; Table S1, Supporting information). Assessment of the total distance flown indicated there was no delineation between short- and long-distance fliers but rather continuous variation (Fig. 1B), similar to other flight mill studies (Colvin & Gatehouse 1993; Schumacher *et al.* 1997). The maximum distance covered by a single individual during one night was 40.6 km. Restricted maximum likelihood (REML) modelling explored the association of population and sex on the total distance using the mill and night flown as random effect variables. There was strong evidence for population differences for the distance flown ($F_{3,61} = 4.73$, $P = 0.005$) with no evidence that the sex of the insect had any effect ($F_{1,32} = 0.48$, $P = 0.492$; Table S2, Supporting information). Of the four Chinese populations flown, adult moths from Dafeng flew the furthest mean distance (mean distance = 15,430 m; Fig. 1C) and their total distance was significantly greater than the three other populations ($F_{1,57} = 13.47$, $P = <0.001$; Table S3, Supporting information) consistent with predictions from the REML and nonoverlapping 95% confidence intervals (Fig. S1, Supporting information). We also wanted to capture the degree of variation in flight activity within a single *H. armigera* population using a single sex, and for this, we used female insects originating from northern Greece ($N = 28$). The mean distance covered by females from Greece was 17 350 m ranging from 457.8 m to 35 430 m (Fig. S2, Supporting information).

All *H. armigera* adults were flown during the prereproductive period (PRP) (the night following eclosion) when migratory flight activity is known to occur in this species (Colvin & Gatehouse 1993). Seasonal migrations are reported across central and northern China (Feng *et al.* 2009), while in northern Greece, there is evidence for the existence of a local overwintering population supplemented with immigrants in the spring, possibly from North Africa (Mironidis *et al.* 2010). On this basis, it is likely that the *H. armigera* used in our experiments possess, to at least some degree, the genetic adaptations necessary for migratory flight. Heritability estimates

indicate a strong genetic component for flight activity in prereproductive *H. armigera* (Colvin & Gatehouse 1993), and the variation in the propensity for long-distance flight observed both within and between populations (Fig. 1B and Fig. S2, Supporting information) suggests that fundamental physiological and behavioural differences underlie the flight potential of insects used in this study.

Genomewide transcription profiles of H. armigera flight phenotypes

Next, we compared genomewide transcription profiles in short- and long-distance fliers characterized using the flight mills. Two separate RNA-seq analyses were performed both within and between *H. armigera* populations. The within-population comparison was performed on Greek moths, while the between-population comparison was conducted on Dafeng and Anyang from China. To capture true differences in flight behaviour, RNA was extracted from whole insects that represent the opposite ends of the flight mill data distribution and were categorized as short- and long-distance phenotypes (Fig. 2A). Total RNA was subject to Illumina cDNA library preparation and sequencing. A total of *c.*286 million paired-end 100-bp reads were sequenced with an average yield of 14.4–26.8 million reads per sample.

Following quality control, reads were aligned to the annotated *H. armigera* genome, isoform variation was accounted for and a combination of FPKM and TMM normalization using Differential Expression on the Web (DEW) (<http://dew.sourceforge.net>) produced robust gene expression estimates. PCA showed distinct transcriptional signatures for each flight phenotype (Fig. 2B). Differential expression between flight phenotypes was determined at a false discovery rate (FDR) of $P < 0.1$ using a union of two separate methods as implemented in edgeR and DESeq2 (Anders *et al.* 2013). Any gene significantly expressed in either the China or Greece experiment was considered a candidate gene for flight. Using this approach, we identified 215 differentially expressed genes from the estimated 17 001 genes in the *H. armigera* genome (1.26%). (Fig. 2C; Data S1, Supporting information). Of the 215 candidates, the majority come from the analysis between short- and long-distance fliers from China ($N = 191$) compared to the intrapopulation study from Greece ($N = 26$). This could reflect the geographical separation between Dafeng and Anyang (*c.* 650 km) or the greater contrast in flight activity between the two phenotypes (Fig. 2A). There was a balance in the share of upregulated and downregulated genes (up = 55.8%; down = 44.2%), and of the two common genes expressed in both data sets

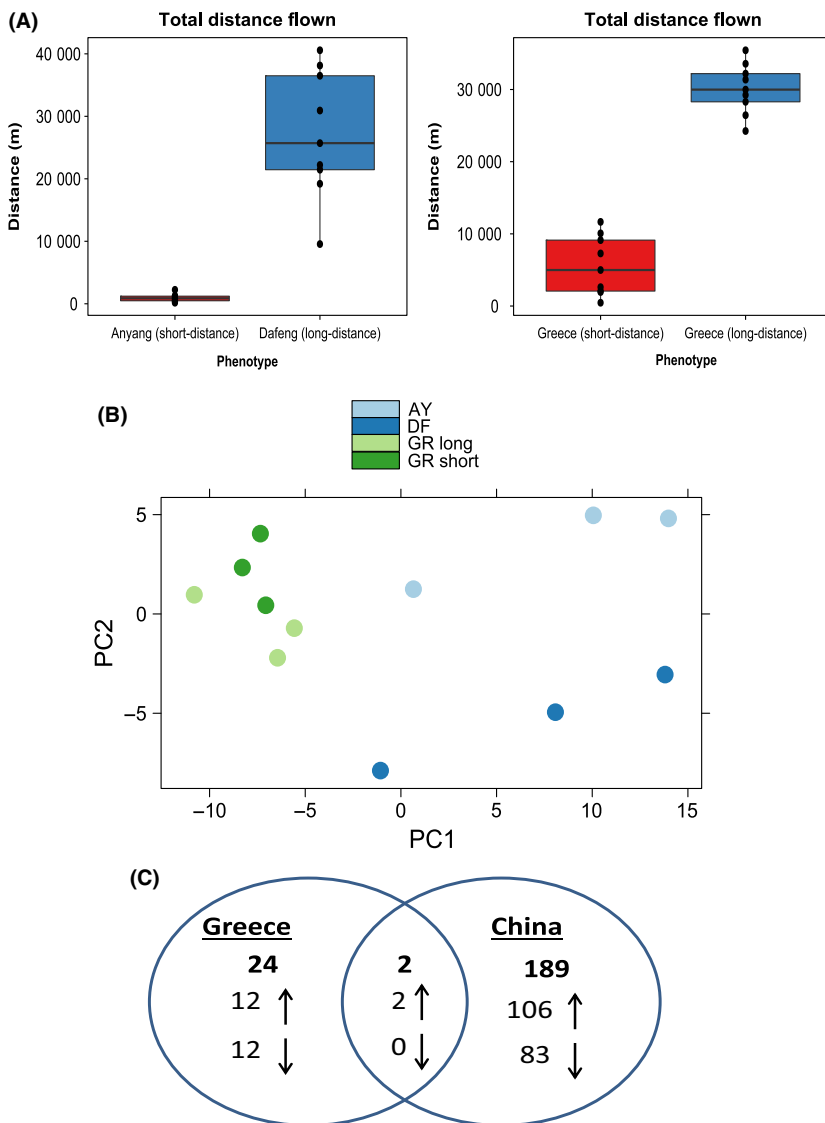


Fig. 2 RNA-seq of *H. armigera* flight phenotypes from China and Greece. (A) Box-plots of total distance flown by moths used in the RNA-seq comparisons between short- and long-distance fliers of *H. armigera* from Greece and China. The magnitude of difference between short- and long-distance fliers is much greater in China than Greece (30.6-fold vs. 4.1-fold). (B) PCA plot of \log_2 -transformed read counts for each sample used in RNA-seq. (C) Venn diagram showing the number of differentially expressed genes between flight phenotypes from Greece and China. All genes were significant at $FDR < 0.1$ using the edgeR and/or DESeq2 package.

(OBP6 and vanin-like protein 3), the fold-change direction was consistent. To validate our expression data, qPCR was used to measure the expression of a subset of genes using the same RNA samples. There was a good agreement between the RNA-seq and qPCR data with a stronger correlation between fold-changes observed for DESeq2 ($R^2 = 0.69$) (Fig. S3, Supporting information) as reported elsewhere (Anders *et al.* 2013).

Expression patterns reflect metabolic demands of prolonged flight in H. armigera

As evidenced by the functional annotation (Fig. 3), a striking proportion of the 215 differentially expressed genes are directly associated with flight physiology. The evolution of prolonged flight in migratory insects, as well as in birds, has led to substantial physiological

adaptations including the efficient mobilization of energy reserves, morphological changes for flight mechanics, coordinated hormonal control and the ability to cope with the stresses imposed by extreme energetic demands. The transcriptional profiles of several genes expressed between the *H. armigera* flight phenotypes reflect these adaptations and are discussed below.

Insect flight is the most energy demanding activity known in the animal kingdom requiring metabolic rate increases of 50- to 100-fold (Arrese & Soulages 2010). The expression data are consistent with the extreme metabolic loads placed on migratory insects during flight. First, we observed several functional GO terms in an enrichment analysis of overexpressed genes in long-distance fliers related to energy production including the purine nucleotide and inosine monophosphate (IMP) pathways (Table S5, Supporting information).

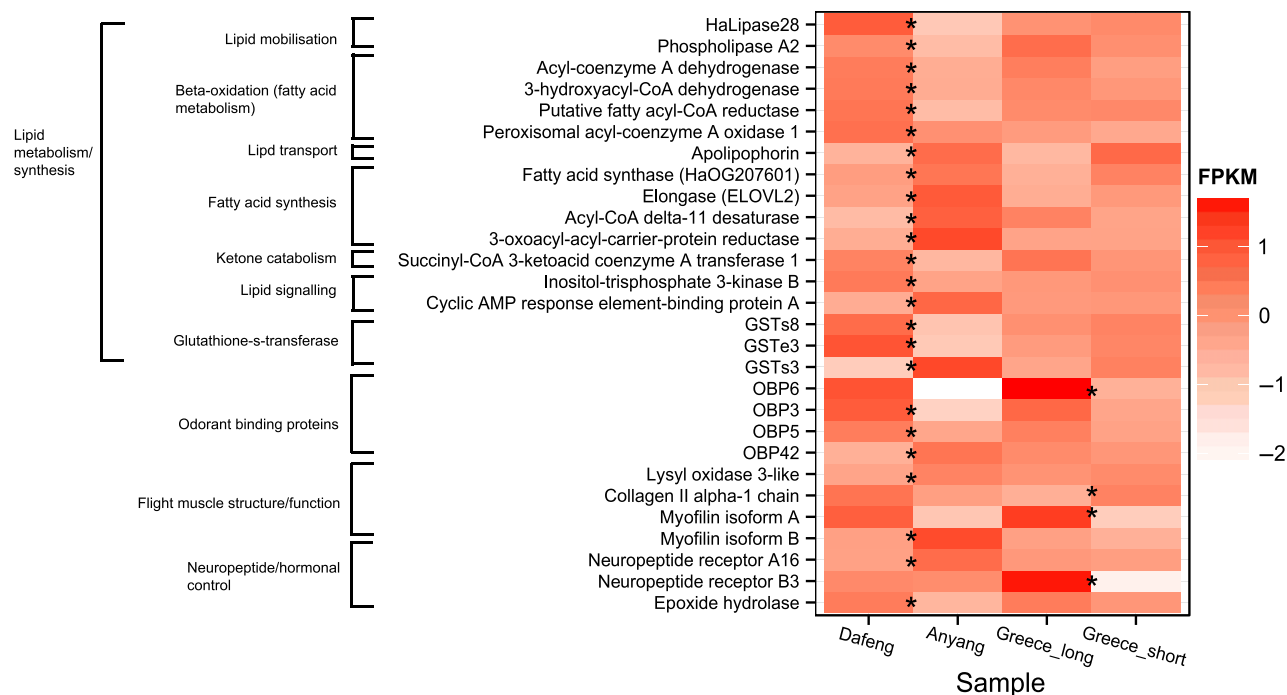


Fig. 3 Differential expression of core genes with an identified role in flight physiology. Heat map of absolute expression measured as normalized FPKM values. Values are given for all genes with differential expression between the flight phenotype comparisons (FDR < 0.1) represented by *. The two flight phenotypes from Greece are marked as 'Greece long' and 'Greece short'.

Both of these pathways are essential for ATP turnover which can increase several hundred-fold in working insect flight muscles (Candy *et al.* 1997). Increases in AMP and IMP have been recorded in locust flight muscle (Weyel & Wegener 1996), and the majority of upregulated transcripts in a microarray comparison of winged versus unwinged pea aphids (*Acyrtosiphon pisum*) were involved in energy production (Brisson *et al.* 2007).

Migratory insects such as locusts and noctuids need to power flight muscle activity for extended periods and do so mainly via the aerobic breakdown of high-energy lipids in the form of triacylglycerol (TAG). The metabolism of fatty acids (FA) in the mitochondria of insect flight muscle cells occurs through β -oxidation. The enzymes that catalyse the first and third reaction of the β -oxidation pathway (fatty acyl-CoA dehydrogenase and 3-hydroxyacyl-CoA dehydrogenase) were upregulated in Dafeng compared to Anyang (1.81–1.83-fold) (Fig. 3). A copy number expansion of genes encoding β -oxidation enzymes was described in the migratory locust (*Locusta migratoria*) providing some evidence that adaptations in this pathway contribute to long-distance flight capacity (Wang *et al.* 2014). β -Oxidation also occurs in peroxisomes and is required to oxidize very long-chain fatty acids that cannot be broken down by mitochondria (Wanders *et al.* 2001). The first step of this

reaction is catalysed by peroxisomal acyl-CoA oxidase which was significantly upregulated in Dafeng moths (c.1.4-fold). Transcriptional regulation of lipid metabolism is supported by the concomitant downregulation of enzymes catalysing the reverse process, fatty acid synthesis (Fig. 3). These include fatty acid synthase, 3-oxoacyl-ACP-reductase, acyl-coA δ desaturase and an elongase (*ELOVL2*). 3-oxoacyl-ACP-reductase and elongase were significantly downregulated in an independent qPCR assessment of additional *H. armigera* flight phenotypes (Fig. 4). Inositol-triphosphate kinase is associated with the inactivation of the first catalytic step of FA synthesis and was also upregulated in Dafeng moths.

Further evidence for lipid-fuelled flight is the expression of two lipases (HaLipase28 and phospholipase A2) and an isoform of the lipid transport protein apolipoprotein. Lipases are essential for mobilizing lipids from TAG stores, and the insect adipose triglyceride lipase, an important enzyme in *Drosophila* metabolism, is a member of the phospholipase A2 family (Arrese & Soullages 2010). Insect flight muscles do not directly store lipids and require an efficient system to transport hydrophobic fatty acids to the muscle cell which in locusts is mediated by apolipoproteins (Hauerland 1997). Finally, there was evidence that *H. armigera* use ketone bodies as supplementary fuel for flight with the

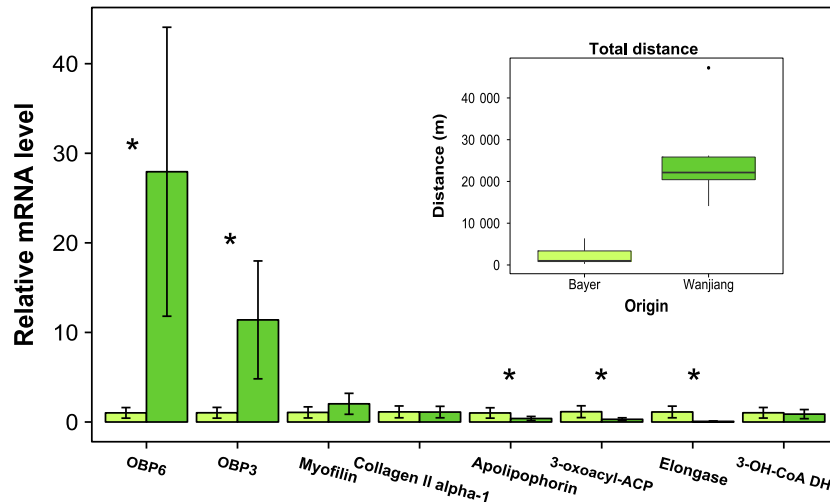


Fig. 4 Transcription of candidate genes for flight in an independent sample set. Expression levels of eight genes identified from RNA-seq were tested using qPCR in *H. armigera* flown in an independent flight mill study. Boxplots (inset) show the total distance flown of a subsample of field-collected adult females from Wanjiang (China) and a laboratory strain (Bayer). The eight genes tested were picked to cover a range of physiological processes contributing to flight (odorant binding proteins (OBP3 and OBP6), flight muscle structure (myofillin and collagen II α -1), fatty acid synthesis (3-oxoacyl-ACP-reductase, elongase), fatty acid metabolism (3-hydroxyacyl-CoA dehydrogenase) and lipid transport (apolipoporphin)). qPCR values are relative to the average of the three biological replicates from Bayer, and significant expression is denoted by * at $P < 0.05$ (t -test of \log_2 -transformed values).

upregulation of succinyl-CoA: 3-ketoacid coenzyme A transferase 1 (*SCOT*), a key enzyme for mediating energy production from ketone. Ketone bodies are present in both the fatty tissues and flight muscles of moths and locusts and have been implicated as an important substrate for flight fuel (Candy *et al.* 1997).

A negative outcome of high ATP turnover and lipid-fuelled flight activity is oxidative damage (Magwere *et al.* 2006). The expression of gene families possessing antioxidant properties, such as glutathione S-transferases (GSTs), protects insects from oxidative stress. Three GSTs were among our expressed data set two of which are members of the sigma class of GSTs and are isoform orthologues of *Drosophila melanogaster* *GSTs1* (Fig. 3). *GSTs1* is highly abundant in *D. melanogaster* indirect flight muscle and protects against 4-hydroxynonenal (4-HNE), a derivative of lipid peroxidation (Singh *et al.* 2001). The expression of sigma GSTs in the insect fat body and copy number expansion in *L. migratoria* (Wang *et al.* 2014), coupled with our data presented here, suggests that *H. armigera* may have evolved a similar adaptation to coping with oxidative stress from prolonged flight.

Odorant binding proteins (OBPs) are small extracellular transporter proteins (13–16 kDa) which capture and transport hydrophobic ligands (e.g. pheromones) (Zhou 2010). Our data set contained four OBPs, one of which, OBP6, was significantly upregulated in moths displaying long-distance phenotypes from both China and Greece (Fig. 3). OBP6 and OBP3 were two of the most

highly expressed genes in long-distance moths from China (Fig. 3) confirmed in the independent assessment of additional flight phenotypes (Fig. 4). While many insect OBPs are antennal specific, both OBP6 and OBP3 are expressed in the insect body (Gu *et al.* 2014) suggesting a role other than olfaction. All moths were flown in the prereproductive period (<1 day old) with no access to host plants so it is highly unlikely that the expression of OBPs represents host-seeking or mating behaviour. The hydrophobic binding capacity of OBPs suggests that this group of proteins may act as lipid transport proteins in *H. armigera*.

Flight activity associated with genes encoding flight muscle proteins

We identified differentially expressed genes strongly associated with flight muscle structure. The α 1-subunit of collagen type II was downregulated in long-distance flying moths from Greece (3.10-fold). Whole-genome sequencing of migratory and nonmigratory monarch butterflies (*D. plexippus*) revealed positive selection surrounding the collagen type IV, subunit α 1, in migratory North American populations (Zhan *et al.* 2014). Collagen IV forms the central component of basement membranes essential for muscle structure and function and the downregulation of this gene in migratory monarchs, coupled with decreased metabolic rates during flight, suggested the evolution of greater metabolic efficiency in the migratory forms (Zhan *et al.* 2014). Collagen II is

the main structural component of cartilage in humans (Garofalo *et al.* 1993) but virtually no information is available on the role of this protein in insects; however, it is possible the downregulation of this gene represents a similar adaptation for greater flight efficiency. The concomitant downregulation of lysyl oxidase supports this finding as this enzyme is essential for maintaining the integrity of collagen structure (Li *et al.* 1997). The expressed gene set also included two isoforms of myofillin, a 20-kDa protein associated with myosin in the core of thick filaments that make up asynchronous insect flight muscle (Qiu *et al.* 2005). The transcription of genes coding for proteins associated with locomotory muscles (e.g. tropomyosin) has been described in long-winged forms of the cricket *Gryllus firmus* (Vellichirammal *et al.* 2014) as well as other migratory organisms (Postel *et al.* 2010).

Evidence for hormonal control of migratory flight activity

Juvenile hormone (JH) plays a central role in the trade-off between reproductive and migratory physiology in insects with lower JH titres favouring migration (Roff & Fairbairn 2007). The *Bombyx mori* homologue of the neuropeptide receptor A16 (*BNGR-A16*), downregulated in long-distance moths from China (FC = 2.67), is the receptor for the hormone allatotropin (AT) which stimulates juvenile hormone (JH) biosynthesis (Yamanaka *et al.* 2008). The relationship between AT mRNA expression and JH levels is tightly coupled and injection of AT induces resident behaviour in migrant insects (McNeil *et al.* 2005; Zhu *et al.* 2009; Jiang *et al.* 2011). The downregulation of the AT receptor, *BNGR-A16*, in long fliers from Dafeng suggests that the contrasting flight propensities between Dafeng and Anyang (Fig. 1B) are a reflection of true differences in migratory potential. The upregulation of an enzyme with JH degrading properties, epoxide hydrolase, supports evidence for the regulation of JH levels in insects from this population.

*Summary – a complex suite of genes underlies migratory flight activity in *H. armigera**

Migration biology is set to enter the genomic era as the number of sequenced genomes from nonmodel migratory species increases (Liedvogel *et al.* 2011). While there is evidence of selection acting on genes involved in a single trait associated with migration in insects (e.g. flight muscle proteins in the monarch butterfly (Zhan *et al.* 2014)), the migratory syndrome is polygenic and the consequence of a complex set of genes and the interactions between them (Roff & Fairbairn 2007). By

combining accurate phenotypic quantification of a migratory trait and next-generation sequencing, we hypothesize that the genes expressed in *H. armigera* displaying contrasting flight activities (and hence different migratory potentials) are strong candidates driving this essential life history characteristic. This complex suite of genes encode proteins for a broad range of physiological functions, and while many have clear direct associations with flight capacity, it is likely that the expression of a subset of our candidate genes is consistent with other aspects of migratory biology and additional pleiotropic effects. The functional role of each gene on migratory behaviour in this species and other noctuid moths needs further investigation. The activity of these proteins will undoubtedly differ between species and experimental approaches; however, they provide a valuable insight into the genetic adaptations that drive this fascinating phenomenon for future comparative studies in insects and other migratory taxa, including birds. It is worth noting that the propensity to engage in migratory flight is intrinsically linked to the physiological state of many insects with the PRP a key window within which migratory behaviour occurs (Dingle 2014). The duration of PRP varies in *H. armigera* reflecting differences in migratory potential between individuals or populations (Colvin & Gatehouse 1993). The transcription of genes associated with migratory flight will undoubtedly vary during the PRP (e.g. genes associated with juvenile hormone; Jiang *et al.* 2011). To reliably assess transcriptional differences between flight phenotypes, it is necessary to sample insects at a comparable age (1 day posteclosion in the present study); however, time-course analyses of migratory insects during the 'migratory window' may capture additional optimal gene expression profiles that contribute to an insect's full migratory capacity. Finally, facultative migrants, such as noctuid moths, must possess the genetic architecture to migrate and fly substantial distances within a single generation in response to adverse prevailing conditions. We propose that the ability to switch on and express many of the genes identified in this study allow *H. armigera* to power flight to high altitudes for extended periods of time and contribute to its status as a global pest.

Acknowledgements

We are grateful to the leaders of the *Helicoverpa* Genome Consortium at CSIRO and University of Melbourne for access to the *H. armigera* genome ahead of its publication. We would also like to thank Jennifer Swain for the maintenance of insect cultures and Suzanne Clark for statistical advice. Rothamsted Research is a national institute of bioscience strategically funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC).

References

- Anders S, McCarthy DJ, Chen Y *et al.* (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols*, **8**, 1765–1786.
- Arrese EL, Soulagès JL (2010) Insect fat body: energy, metabolism, and regulation. *Annual Review of Entomology*, **55**, 207–225.
- Brisson JA, Davis GK, Stern DL (2007) Common genome-wide patterns of transcript accumulation underlying the wing polyphenism and polymorphism in the pea aphid (*Acyrtosiphon pisum*). *Evolution & Development*, **9**, 338–346.
- Bustin SA, Benes V, Garston JA *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, **55**, 611–622.
- Candy DJ, Becker A, Wegener G (1997) Coordination and integration of metabolism in insect flight. *Comparative Biochemistry & Physiology B*, **117**, 497–512.
- Chapman JW, Reynolds DR, Hill JK, Sivell D, Smith AD, Woilwood IP (2008) A seasonal switch in compass orientation in a high-flying migratory moth. *Current Biology*, **18**, R908–R909.
- Chapman JW, Nesbit RL, Burgin LE *et al.* (2010) Flight orientation behaviors promote optimal migration trajectories in high-flying insects. *Science*, **327**, 682–685.
- Chapman JW, Bell JR, Burgin LE *et al.* (2012) Seasonal migration to high latitudes results in major reproductive benefits in an insect. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 14924–14929.
- Chapman JW, Reynolds DR, Wilson K (2015) Long-range seasonal migration in insects: mechanisms, evolutionary drivers and ecological consequences. *Ecology Letters*, **18**, 287–302.
- Colvin J, Gatehouse AG (1993) The reproduction flight syndrome and the inheritance of tethered flight activity in the cotton bollworm moth, *Heliothis armigera*. *Physiological Entomology*, **18**, 16–22.
- Dingle H (2014) *Migration: The Biology of Life on the Move*. Oxford University Press, Oxford.
- Dorhout DL, Sappington TW, Rice ME (2008) Evidence for obligate migratory flight behavior in young European corn borer (Lepidoptera: Crambidae) females. *Environmental Entomology*, **37**, 1280–1290.
- Feng HQ, Wu XF, Wu B, Wu KM (2009) Seasonal migration of *Helicoverpa armigera* (Lepidoptera: Noctuidae) over the Bohai Sea. *Journal of Economic Entomology*, **102**, 95–104.
- Garofalo S, Metsäranta M, Ellard J *et al.* (1993) Assembly of cartilage collagen fibrils is disrupted by overexpression of normal type-II collagen in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 3825–3829.
- Götz S, García-Gómez JM, Terol J *et al.* (2008) A high-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, **36**, 3420–3435.
- Gu SH, Sun L, Yang R-N *et al.* (2014) Molecular characterization and differential expression of olfactory genes in the antennae of the black cutworm moth *Agrotis ipsilon*. *PLoS ONE*, **9**, e103420.
- Hauenerland NH (1997) Transport and utilization in insect flight muscles. *Comparative Biochemistry & Physiology B*, **117**, 475–482.
- Jiang X, Luo L, Zhang L, Sappington TW, Hu Y (2011) Regulation of migration in *mythimna separata* (Walker) in China: a review integrating environmental, physiological, hormonal, genetic, and molecular factors. *Environmental Entomology*, **40**, 516–533.
- Jones S, Pfister-Genskow M, Cirelli C, Benca RM (2008) Changes in brain gene expression during migration in the white-crowned sparrow. *Brain Research Bulletin*, **76**, 536–544.
- Kent JW, Rankin MA (2001) Heritability and physiological correlates of migratory tendency in the grasshopper *Melanoplus sanguinipes*. *Physiological Entomology*, **26**, 371–380.
- Li WD, Nellaiappan K, Strassmaier T, Graham L, Thomas KM, Kagan HM (1997) Localization and activity of lysyl oxidase within nuclei of fibrogenic cells. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 12817–12822.
- Liedvogel M, Akesson S, Bensch S (2011) The genetics of migration on the move. *Trends in Ecology & Evolution*, **26**, 561–569.
- Liu ZF, McNeil JN, Wu KM (2011) Flight mill performance of the lacewing *Chrysoperla sinica* (Neuroptera: Chrysopidae) as a function of age, temperature, and relative humidity. *Journal of Economic Entomology*, **104**, 94–100.
- Magwere T, Pamplona R, Miwa S *et al.* (2006) Flight activity, mortality rates, and lipoxidative damage in *Drosophila*. *Journal of Gerontology Series A Biological Sciences & Medical Sciences*, **61**, 136–145.
- McKinney GJ, Hale MC, Goetz G, Gribskov M, Thrower FP, Nichols KM (2015) Ontogenetic changes in embryonic and brain gene expression in progeny produced from migratory and resident *Oncorhynchus mykiss*. *Molecular Ecology*, **24**, 1792–1809.
- McNeil JN, Maury M, Bernier-Cardou M, Cusson M (2005) *Manduca sexta* allatotropin and the in vitro biosynthesis of juvenile hormone by moth corpora allata: a comparison of *Pseudaletia unipuncta* females from two natural populations and two selected lines. *Journal of Insect Physiology*, **51**, 55–60.
- Mironidis GK, Stamopoulos DC, Savopoulou-Soultani M (2010) Overwintering survival and spring emergence of *Helicoverpa armigera* (Lepidoptera: Noctuidae) in northern Greece. *Environmental Entomology*, **39**, 1068–1084.
- Postel U, Thompson F, Barker G, Viney M, Morris S (2010) Migration-related changes in gene expression in leg muscle of the Christmas Island red crab *Gecarcoidea natalis*: seasonal preparation for long-distance walking. *Journal of Experimental Biology*, **213**, 1740–1750.
- Pulido F, Berthold P, van Noordwijk AJ (1996) Frequency of migrants and migratory activity are genetically correlated in a bird population: evolutionary implications. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 14642–14647.
- Qiu F, Brendel S, Cunha PM *et al.* (2005) Myofilin, a protein in the thick filaments of insect muscle. *Journal of Cell Science*, **118**, 1527–1536.
- R Core Development Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, ISBN 3-900051-07-0, Available from <http://www.R-project.org>.
- Roberts A, Pachter L (2012) Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods*, **10**, 71–73.

- Roff DA, Fairbairn DJ (2007) The evolution and genetics of migration in insects. *BioScience*, **57**, 155–164.
- Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C-T method. *Nature Protocols*, **3**, 1101–1108.
- Schumacher P, Weber DC, Hagger C, Dorn S (1997) Heritability of flight distance for *Cydia pomonella*. *Entomologia Experimentalis et Applicata*, **85**, 169–175.
- Scott KD, Wilkinson KS, Lawrence N *et al.* (2005) Gene-flow between populations of cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) is highly variable between years. *Bulletin of Entomological Research*, **95**, 381–392.
- Singh SP, Coronella JA, Beneš H, Cochrane BJ, Zimniak P (2001) Catalytic function of *Drosophila melanogaster* glutathione S-transferase DmGSTS1-1 (GST-2) in conjugation of lipid peroxidation end products. *European Journal of Biochemistry*, **268**, 2912–2923.
- Vanderzant ES, Fort SW, Richardson CD (1962) Rearing of bollworm on artificial diet. *Journal of Economic Entomology*, **55**, 140.
- Vellichirammal NN, Zera AJ, Schilder RJ, Wehrkamp C, Riethoven J-JM, Brisson JA (2014) De novo transcriptome assembly from fat body and flight muscles transcripts to identify morph-specific gene expression profiles in *Gryllus firmus*. *PLoS ONE*, **9**, e82129.
- Wanders RJA, Vreken P, Ferdinandusse S *et al.* (2001) Peroxisomal fatty acid alpha- and beta-oxidation in humans: enzymology, peroxisomal metabolite transporters and peroxisomal diseases. *Biochemical Society Transactions*, **29**, 250–267.
- Wang ZJ, Dong YC, Desneux N, Niu CY (2013) RNAi silencing of the HaHMG-CoA reductase gene inhibits oviposition in the *Helicoverpa armigera* cotton bollworm. *PLoS ONE*, **8**, e67732.
- Wang XH, Fang X, Yang P *et al.* (2014) The locust genome provides insight into swarm formation and long-distance flight. *Nature Communications*, **5**, 1–9.
- Weyl W, Wegener G (1996) Adenine nucleotide metabolism during anoxia and postanoxic recovery in insects. *Experientia*, **52**, 474–480.
- Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.
- Yamanaka N, Yamamoto S, Žitňan D *et al.* (2008) Neuropeptide receptor transcriptome reveals unidentified neuroendocrine pathways. *PLoS ONE*, **3**, e3048.
- Yan S, Ni H, Li H, Zhang J, Liu X, Zhang Q (2013) Molecular cloning, characterization, and mRNA expression of two cryptochrome genes in *Helicoverpa armigera* (Lepidoptera: Noctuidae). *Journal of Economic Entomology*, **106**, 450–462.
- Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, **13**, 134.
- Zhan S, Zhang W, Nitepöld K *et al.* (2014) The genetics of monarch butterfly migration and warning colouration. *Nature*, **514**, 317–321.
- Zhou JJ (2010) Odorant binding proteins in insects. *Vitamins and Hormones*, **83**, 241–272.
- Zhou XF, Faktor O, Applebaum SW, Coll M (2000) Population structure of the pestiferous moth *Helicoverpa armigera* in the Eastern Mediterranean using RAPD analysis. *Heredity*, **85**, 251–256.
- Zhu HS, Gegear RJ, Casselman A, Kanginakudru S, Reppert SM (2009) Defining behavioral and molecular differences between summer and migratory monarch butterflies. *BMC Biology*, **7**, 1437.

C.M.J., K.S.L., C.B. and J.W.C. conceived and designed the study. Y.Y., J.V. and G.K.M. performed field sampling and provided insect material. C.M.J. performed all tethered flight mill experiments and molecular laboratory work. A.P. and J.G.O. provided genomic resources. C.M.J. and A.P. conducted RNA-seq analyses. C.M.J., A.P., J.G.O., C.B. and J.W.C. wrote the manuscript.

Data accessibility

The flight mill data, raw RNA-seq read count data, output from edgeR and DESeq2, GO terms and raw qPCR data (Ct and validation analysis values) are deposited in the online Dryad Digital Repository (www-datadryad.org; doi:10.5061/dryad.bp7j7). RNA-seq fastq files and experimental information are available at ArrayExpress (Accession no. E-MTAB-3790). Gene expression data for candidate genes are provided as supporting information in file: Jones_et_al_DataS1.xlsx.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1. REML estimations of total distance flown by Chinese populations.

Figure S2. Flight performance of *H. armigera* from northern Greece.

Figure S3. Validation of RNA-seq by qPCR.

Table S1. Information on the origin and generation of each *H. armigera* collection used in either flight mill, RNA-seq and/or qPCR experiments.

Table S2. Output of explanatory variables from full REML model for total distance flown by Chinese *H. armigera*.

Table S3. Output of explanatory variables from full REML model investigating total distance flown by Dafeng *H. armigera* versus three other Chinese populations.

Table S4. Primer information for target and control genes used to validate and analyse expression in *H. armigera*.

Table S5. GO-term enrichment analysis of genes up-regulated in long-distance phenotypes of *H. armigera*.

Data S1. A list of differentially expressed genes, corresponding annotations and fold changes for the China and Greece flight mill experiments.