Received Date : 13-Nov-2015

Accepted Date : 13-Mar-2016

Article type : Research Papers

Rapid evolution and gene expression: A rapidly-evolving Mendelian trait that silences field crickets has widespread effects on mRNA and protein expression

Sonia Pascoal[1*], Xuan Liu[2], Tony Ly[3], Yongxiang Fang[2], Nichola Rockliffe[2], Steve Paterson[2], Sally L. Shirran[4], Catherine H. Botting[4], Nathan W. Bailey[1]

[1] Centre for Biological Diversity, University of St Andrews, St Andrews, KY16 9TH, U.K.

[2]Centre for Genomic Research, University of Liverpool, Liverpool, L69 7ZB, U.K.

[3]Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee, Dow St, Dundee, DD1 5EH, U.K.

[4]School of Chemistry, Biomedical Sciences Research Complex, BMS Annexe, University of St. Andrews, KY16 9ST, U.K.

*Corresponding authors: scmp@st-andrews.ac.uk; nwb3@st-andrews.ac.uk; Tel: 01334 463367; Fax: 01334 463366

Keywords: Rapid evolution, wing mutation, *Teleogryllus oceanicus*, transcriptome, proteome

Running title: Rapid evolution and gene expression in crickets

ABSTRACT

A major advance in modern evolutionary biology is the ability to start linking phenotypic evolution in the wild with genomic changes that underlie that evolution. We capitalised on a rapidly-evolving Hawaiian population of crickets (*Teleogryllus oceanicus*) to test hypotheses about the genomic consequences of a recent Mendelian mutation of large effect which disrupts the development of sound-producing structures on male forewings. The resulting silent phenotype, flatwing, persists because of natural selection imposed by an acoustically-orienting parasitoid, but it interferes with mate attraction. We examined gene expression differences in developing wing buds of wild-type and flatwing male crickets using RNA-seq and quantitative proteomics. Most differentially expressed (DE) transcripts were down-regulated in flatwing males (625 up vs. 1716 down), whereas up and down-regulated proteins were equally represented (30 up and 34 down). Differences between morphs were clearly not restricted to a single pathway, and we recovered annotations associated with a broad array of functions that would not be predicted *a priori*. Using a candidate gene detection test based on homology we identified 30% of putative *Drosophila* wing development genes in the cricket transcriptome, but only 10% were DE. In addition to wing related annotations, endocrine pathways and several biological processes such as reproduction, immunity and locomotion were DE in the mutant crickets at both biological

levels. Our results illuminate the breadth of genetic pathways that are potentially affected in the early stages of adaptation.

INTRODUCTION

A question that has dominated evolutionary research since the Modern Synthesis is how novel sequence variants arise within a genome and persist—or spread—under selection. Debate centres on the degree to which effects of mutations are limited (i.e. modular), and the mechanisms by which antagonistic fitness effects are counterbalanced (Stearns, 2010). Such factors can potentially affect whether mutations become established, and their rate of spread under selection. The development of high-throughput next generation sequencing technology has made it easier to inform this debate, even in non-model organisms, by dissecting the functional genomics of rapidly evolving mutations.

The rapidly evolving melanic form of the peppered moth *Biston betularia* provides a good example. Peppered moths have been a fixture of evolutionary research examining the action of natural selection for over a century and a half, but a description of the functional genetics of the melanic form of this species was only published in 2011 (van't Hof *et al.,* 2011; Cook & Saccheri, 2013). Prior to this work, it was not unreasonable to predict that the genetic mechanisms underlying development of melanic pigmentation might involve a component of the well-characterised insect melanin synthesis pathway. Surprisingly, an assay testing linkage between candidate melanisation genes and the melanic *carbonaria* phenotype in *B. betularia* revealed this was not the case (van't Hof & Saccheri, 2010). Instead, the *carbonaria* phenotype maps to a chromosomal region containing loci that contribute to wing patterning

in other lepidopterans (van't Hof & Saccheri, 2010; van't Hof *et al.,* 2011). Other examples of rapid evolution are well-known from field studies (Grant & Grant, 2002; Charlat *et al.,* 2007), but few contemporarily evolving traits in nature are well characterized on a genomic level (Stapley *et al.,* 2010). Here we capitalise on a rapidly spreading mutant phenotype in wild field crickets (*Teleogryllus oceanicus*) and use two complementary gene expression profiling approaches to test the specificity of expression changes associated with strong selection and rapid evolution.

In Hawaii, a mutant form of silent male crickets has recently arisen and rapidly spread in several populations, apparently in response to pressure from an acoustically-orienting parasitoid fly, *Ormia ochracea* (Zuk *et al.,* 2006). The mutation, flatwing, is sex-specific and affects male forewings. Females do not sing and their wings lack derived sound-producing structures such as the scraper, the file, and resonators. Mutant males do not develop these features or develop drastically reduced versions of them, and the differences between morphs are detectable in juvenile stages (Fig. 1). While silence protects males from parasitoid attack, it impedes their ability to attract females to reproduce. Flatwing males, however, appear to behave as “satellites” towards the remaining calling males (Zuk *et al.,* 2006). This behaviour may have pre-dated the origin and spread of the mutation (Tinghitella *et al.,* 2009), and along with the willingness of females to accept silent males for mating (Bailey *et al.,* 2008) facilitated the mutation’s rapid spread. Despite appearing only a decade ago, silent males now occur on two islands and have distinct forms that segregate as sex-linked, sex-limited Mendelian traits in each (Tinghitella, 2008; Pascoal *et al.,* 2014).

Despite a rudimentary understanding of the Mendelian genetics of flatwing, nothing is known about its molecular functional genetics. As was the case with the peppered moth, several pathways suggest themselves as obvious candidates underlying the developmental disruption causing flatwing. A substantial body of research on the developmental genetics of wing morphogenesis and venation patterns has been conducted using *Drosophila melanogaster* (Blair, 2007). Key regulatory genes in wing development and wing vein formation have been identified, for example *wingless*, *vestigial*, *rhomboid* and EGFR (epidermal growth factor receptor) (Baker, 1988; Sturtevant & Bier, 1995; Fuse *et al.,* 1996; Niwa *et al.,* 2010; Molnar & de Celis, 2013). These comparatively well-characterised genes and associated pathways provide candidates for investigating the genetic basis of altered wing venation patterns in male *T. oceanicus* crickets.

We took a multi-level approach to analyse transcriptomic and proteomic variation in the wing bud stage of wild-type *versus* mutant flatwing crickets. Despite the comparative rarity of quantitative proteomics approaches in evolutionary biology, contrasting and combining data from RNA and protein expression screens can increase the power to dissect and test hypotheses about the molecular machinery underlying adaptive phenotypes. Our goal was to qualitatively test the universality of effects on gene expression caused by the flatwing mutation(s) in developing wing bud tissue. To do this, we i) characterized the wing bud transcriptome and quantitatively compared gene expression profiles between wild-type and flatwing male wing buds, ii) similarly accessed the wing bud proteome and quantitatively compared protein expression patterns between wild-type and flatwing male wing buds iii) established patterns of co-expression between the datasets levels to gain focused information about the molecular basis of the flatwing phenotype and iv) used existing

information from *D. melanogaster* (Costello *et al.,* 2009) to test predictions about wing vein pathways implicated in the development of flatwing morphology in *T. oceanicus*. If relatively conserved venation patterns across insect wings are determined by the expression of conserved genes, we expected to be able to identify *Drosophila* candidates in our RNA-seq dataset on the basis of homology. Secondly, we predicted that those genes should be differentially expressed (DE) in developing flatwing males, as flatwing forewings resemble the un-differentiated forewings of females. Our aim was not to characterise the causative sequence mutation(s) underlying the phenotype. Rather, we tested the extent to which molecular pathways associated with other phenotypes might be disrupted by flatwing.

MATERIALS AND METHODS

Crickets

To examine wing bud differences between crickets carrying the flatwing mutation and crickets carrying normal X chromosomes (wild-type), lines fixed for each phenotype were produced for the Kauai population. To obtain homozygous lines for each morph, we performed thirty replicates of each of two types of crosses. The first was a cross between a Kauai female of unknown genotype with a wild type Kauai male (who was therefore know to carry a single normal X chromosome). The second was a cross between a Kauai female of unknown genotype and a flatwing Kauai male (who was therefore known to carry a single mutant X chromosome). We screened wing phenotypes in the resulting male offspring, which allowed us to identify 12 homozygous normal lines and 3 homozygous flatwing lines. We haphazardly selected 3 of the homozygous wild-type lines to retain for a total of 3 biological replicates of each genotype.

Test crickets were reared in a common garden environment following established protocols (e.g. Bailey & Macleod, 2014), in a growth chamber at ca. 25 ^{o}C with a 12:12 light:dark cycle. Crickets were reared in 16 L plastic containers, and cohorts developed together within these containers to prevent mixing between developmental stages. All boxes were kept in the same growth chamber, so all nymphs used in the experiment had experienced the same acoustic environment. Wing buds were collected from juvenile male crickets (about 2 months) at the stage when they first become everted and freely articulated from the thorax (Fig. 1). The number of instars varies in many insects, including grylline crickets (Esperk et al. 2007), and our interest was in patterns of gene expression in wing buds when they first appear externally. To ensure that we sampled crickets that were all at the same stage of wing bud development, i.e. when they were externally visible but not yet enlarged as in the penultimate instar, we selected crickets at the second-before-the-final moult. This stage is readily characterised by the features illustrated in Fig. 1. Excised wing buds were preserved in RNALater at -20 ^{o}C until processing.

RNA extractions, library preparation and sequencing

Total RNA was extracted from cricket wing buds using the TRIzol plus RNA purification kit (Life Technologies) with On-column PureLink DNase treatment during purification (Invitrogen). Twelve wing bud tissue samples were analysed (6 biological replicates for each morph, i.e. 2 replicates per line, with 3 crickets from the same line pooled per sample). Samples were subjected to an initial QC measuring RNA concentration with a Qubit RNA broad range kit. Sample integrity was also assayed using RNA pico chips on an Agilent Bioanalyser. One µg of each sample was RNA depleted with Ribo-Zero specific for human,

mouse and rat, using the manufacturer's protocol. Samples were heated with baits at 68 ^{o}C for 10 minutes and allowed to cool for 5 minutes. The hybridization mixture was mixed with prepared Ribo-Zero beads and left at room temperature for 5 minutes, then heated at 50 ^{o}C for 5 minutes and immediately placed on a magnet. The supernatant was collected and purified with RNA clean beads.

Purified RNA was checked for depletion and used directly in the ScriptSeq protocol (Epicentre), following the manufacturer's instructions. Samples were fragmented at 85 ^{o}C for 5 minutes and placed on ice, then converted to cDNA and purified with Ampure XP beads. The samples were PCR amplified for 14 cycles (barcodes were incorporated at this point) and checked using Qubit (Invitrogen) and Bioanalyser (Agilent). Samples were multiplexed based on concentration and average length. The quantity and quality of the final pool were assessed again and subsequently qPCR using the Illumina Library Quantification Kit from Kapa on a Roche Light Cycler LC480II, according to the manufacturer's instructions. The template DNA was denatured according to the protocol described in the Illumina cBot User's guide and loaded at 9 pM concentration. To improve sequencing quality control, 1% fragmented phage PhiX DNA was spiked-in to the libraries. The sequencing was carried out on an Illumina HiSeq 2000 with version 3 chemistry generating 2×100 bp paired end reads.

RNA-seq data analysis

Initial processing and quality assessment

An in-house pipeline was used for initial processing and quality assessment of the sequence data. Briefly, base calling and de-multiplexing of indexed reads was performed using CASAVA version 1.8.2 (Illumina) to produce samples in fastq format. The raw fastq files were trimmed

to remove Illumina adapter sequences using Cutadapt version 1.2.1 (Martin, 2011). The reads were further trimmed to remove low quality bases, using Sickle version 1.200 with a minimum window quality score of 20. After trimming, reads shorter than 10 bp were removed. The proportion of trimmed reads that were unpaired after trimming was generally low (<1%), indicating that the data are of good quality.

De novo assembly of transcripts, transcript quantification and annotation of DE sequences

Transcriptome assembly was carried out using Trinity (Grabherr *et al.,* 2011); read data were normalized and assembled using a k-mer size of 25bp. Prior to assembly, the effect of K-mer length (23 to 31 bp) was tested and digital normalization was applied to reduce redundancy in the dataset and reduce the memory requirements of the assembly. Further assembly quality was assessed using CEGMA (Parra *et al.,* 2007) testing for the presence and completeness of a set of 248 conserved eukaryotic genes (CEG).

RSEM (Li & Dewey, 2011) was used for quantifying transcript (genes and isoforms) abundances. In RSEM, trimmed illumina reads were mapped to the *de novo* transcriptome assembly using BOWTIE 2 (Langmead & Salzberg, 2012) and the mapping BAM files were then used to generate raw counts for differential expression analysis. EdgeR (Robinson *et al.,* 2010) was subsequently used for identifying genes and isoforms differentially expressed between wild-type *versus* flatwing, assuming a false discovery rate (FDR) threshold of 5% and a fold change threshold of at least 1.5. Trinotate (Haas *et al.,* 2013) (trinotate.sourceforge.net/), a Trinity tool, was used to annotate the transcriptome and the DE assembled sequences and Blast2GO online software (Conesa *et al.,* 2005)

(http://www.blast2go.com) was used for additional DE list GO organization, KEGG (Kyoto encyclopedia of genes and genomes) analysis and gene enrichment analysis using Fisher's exact test (FDR = 0.05).

Quantitative Proteomics

We confirmed that significant numbers of proteins could be identified from flight muscle tissue using mass spectrometry-based proteomics methods before embarking on a quantitative proteomics study; the methods and results are presented in the Supporting Information. We used extra samples of biological replicates from the RNA-seq experiment to perform an iTRAQ labelling experiment using nano-flow liquid chromatographic electrospray ionization tandem mass spectrometry (nLC-ESI MS/MS). These samples comprised wing buds collected from siblings of the juvenile males reared simultaneously in the same conditions/experiment as in the RNA-seq experiment. Six samples (3 wild-type and 3 flatwing biological replicates with 3 crickets pooled per sample) were analysed, each labelled with a different 8plex iTRAQ reagent (labels 113-118). Wing bud proteins were extracted by homogenizing the tissue in 4% SDS HEPES-EDTA lysis buffer containing protease and phosphatase inhibitors (Roche) at 95 ^{o}C (Ly *et al.,* 2014). To assess protein quality, a portion of the extracts were run on a NuPAGE® Novex® 4-12% Bis-Tris Protein Gel (Life Technologies) (Fig. S2) and protein concentration was measured using Qubit and Nanodrop methods. Subsequently, 100 µg of protein per sample was acetone precipitated and the resulting pellets were solubilised, digested with trypsin and the resultant peptides reduced and alkylated prior to iTRAQ labelling, following the manufacturer's protocol (ABSciex iTRAQ reagents). The combined iTRAQ labelled peptides were concentrated (SpeedVac, ThermoSavant) and resuspended in 1.4 mL load buffer (10 mM KH_2PO_4 pH 3.0 in 25 %

acetonitrile) and sonicated. The pH was assessed and if necessary adjusted to 3.0 with 0.5 M H_3PO_4.

The peptides were then separated by cation exchange chromatography on a PolySulfoethyl A column (PolyLC, Columbia, MD). The column was washed with 100% Buffer A_{scx} (10 mM KH_2PO_4, 20% acetonitrile (MeCN), pH 3.0) at 1 mL min^{-1} for 22 min allowing the absorbance on the UV chromatogram to return to baseline. A gradient of 0-50% B_{scx} (10 mM KH_2PO_4, 20% MeCN, 500 mM KCl, pH 3.0) was applied for 20 min, 50-100% B_{scx} for 3 min, followed by 100% B_{scx} for a further 3 min to wash the column, before reequilibration in 100% A_{scx} for another 11 min. 0.5 mL fractions were collected every 30 sec. The chromatogram was inspected and fractions pooled to give 7 fractions across the elution profile of similar peptide concentration (Fig. S3) which were concentrated (SpeedVac, ThermoSavant) prior to desalting. Fractions were resuspended in 0.1% TFA and desalted on C18 spin columns (PepClean C18 spin columns, Thermo Scientific) using the manufacturer's instructions, eluting in 2x20 µL 70% MeCN. The elution solvent was removed (SpeedVac) and the fractions resuspended in 20 µL loading buffer (98% H_2O, 2% MeCN, 0.05% TFA) prior to mass spectrometric analysis.

nLC-ESI-MSMS analysis

The peptides were then analysed by nLC-ESI MSMS using the instrument set up described in the Supporting Information. Half of each desalted fraction volume (10 uL) was loaded onto the nLC trap column and washed with loading buffer (98% H_2O, 2% MeCN, 0.05% TFA) for 20 min at 5 µL min^{-1}, the trap was then switched in line with the column and the peptides

eluted with a gradient of increasing MeCN as described in the Supporting Information and sprayed into the TripleTOF 5600 mass spectrometer and data collected as described in the Supporting Information except that the rolling collision energy used was adjusted to the preset iTRAQ settings, to give improved release of the reporter ions.

MS/MS data analysis

The MS/MS data was analysed using both the ProteinPilot search algorithm (ABSciex), using the predefined iTRAQ8plex settings, plus the Mascot algorithm (Matrix Science) against the NCBInr database Dec 2014. Species was restricted to metazoa, trypsin included as the cleavage enzyme, iTRAQ8plex as a fixed modification on peptide N-termini and lysines, methylthio as a fixed modification of cysteines and methionine oxidation and iTRAQ8plex on tyrosine as variable modifications. Additionally, searches were performed against a wing bud transcriptome database. For protein identification, at least one peptide above the identity (95% confidence) threshold was required. For protein quantification analysis, however, Mascot data was further filtered and only proteins with at least two peptides with ion scores above the 95% confidence threshold were considered. Final quantifications were performed manually using the Mascot results from searches against the filtered proteome data. Briefly, before protein quantification, data were normalized for uneven admixture in two consecutive steps. This normalization process assumes that most proteins are not DE and that true DE proteins constitute a minority of the sampled proteins. Firstly, the measured intensities for each sample (i.e., iTRAQ channel) were summed over all peptides. Each sum was then divided by the highest sum to generate normalization ratios that were then used to normalize each peptide intensity value by dividing each value by the corresponding ratio.

Secondly, a normalization factor by peptide was also calculated and proportions to the sum of each peptide across all samples (i.e., iTRAQ channels) were therefore produced. Fold changes and *p* values (obtained by Student's t-test) were then calculated using the normalized data for all peptides in a protein. To evaluate if proteins were differentially expressed, a *p*-value threshold of 0.05 and a fold change larger than 1.5 or smaller than 0.67 were applied.

The proteome and DE protein annotations, KEGG analysis and gene enrichment analysis were obtained using Blast2GO. Since the transcriptome assembly was used for protein identification searches, transcript identifications were directly assigned into the corresponding matching proteins. The comparison between RNA-seq and proteomics datasets was therefore evaluated to estimate patterns of expression in the transcriptome *versus* in the proteome.

Lastly, we implemented a candidate gene detection test to evaluate the hypothesis that coding genes implicated in *Drosophila* wing development and venation also underlie similar developmental processes in *T. oceanicus*, and that the flatwing mutation causes expression levels of those genes in male wing buds to be disrupted. To do this, we capitalised on a publically-available list of genes with documented involvement in *Drosophila* wing development (n = 296) (http://www.sdbonline.org/sites/fly/aimorph/wing.htm). We then searched for these genes in our *T. oceanicus* wing bud transcriptome, proteome and associated DE annotations, and then tested whether transcriptome and proteome

expression profiles were consistent with disruption of any of these during the development of flatwing morphology.

RESULTS

Wing bud transcriptome

Trimmed and normalized RNA-seq reads were assembled into 1,249,083 transcripts (639,586 "trinity genes" and 609,497 isoforms) with a mean sequence length of 1,212 bp and a size range of 201 to 19,098 bp. About 16% of the genes have multiple isoforms (Tables S1-S6). The CEGMA test for the presence and completeness of a set of 248 conserved eukaryotic genes (CEG) showed that the large majority of the genes (98%) are present, but a smaller proportion (79%) was complete, i.e. the full gene was contained within one assembled sequence. From Trinotate, BLASTx of transcripts against the UniProt/SwissProt database yielded 32,923 transcript annotations. From these, several known wing morphogenesis and wing venation regulatory genes such as *wingless*, *vestigial*, *snail*, *EGFR*, *rhomboid*, *hedgehog* and *notch* were identified and we therefore anticipated a good representation of these annotations in the DE analysis.

Gene expression analysis

2341 differentially expressed assembled sequences were identified and are therefore potentially involved in pathways leading to development of the flatwing phenotype (Fig. S1). Down-regulation was prevalent in flatwing males (Fig. 2A). Of the DE sequences, 1716 were

down-regulated whereas only 625 were up-regulated. Interestingly, PCA analysis and sample correlation heat maps of the full dataset and of the DE genes only revealed greater variation between wing bud transcriptomes of wild-type crickets than from mutant crickets. As expected, considering DE data alone maximizes the differences between the two morphs (Fig. 3).

The majority of the DE sequences are unknown but about 21% (502) had an available annotation and 222 had associated gene ontologies. Most genes that would be predicted *a priori* to be involved in development of the flatwing phenotype based on their known role in wing venation in other insects, such as *wingless*, *rhomboid* and *EGFR*, were not DE in this developmental stage of *T. oceanicus* flatwings. However, *Dystrophin* and *Protocadherin Fat 1*, involved in imaginal disc-derived wing vein morphogenesis/specification and establishment of imaginal disc-derived wing hair orientation, were both up-regulated in flatwing crickets. Also, the regulation of notch signalling was up-regulated and the hedgehog receptor activity was down-regulated (both play an important role in *Drosophila* wing development). Also of interest, several behavioural annotations were recovered. For example, the *protein white*, involved in male courtship behaviour was up-regulated in flatwings; *vacuolar sorting-associated protein* assigned to social behaviour was also up-regulated, and an uncharacterized protein linked to flight behaviour, sperm aster formation and cuticle pigmentation was down-regulated.

Focusing on GO terms only, biological processes such as response to stimulus, reproduction, immune system process and localization were DE between wild-type and flatwing crickets. Locomotion, rhythmic process (circadian) and signalling were consistently down-regulated

and biological adhesion was consistently up-regulated in flatwing males (Fig. 2B, Table S7). From KEGG analysis, enzymes present in steroid hormone biosynthesis, steroid degradation, drug metabolism by cytochrome P450 and metabolism of xenobiotics by cytochrome P450 (Table S8) were down-regulated in flatwings and gene enrichment analysis (Fisher's exact test) mainly showed a down-regulation of transport activity in flatwings (Table S9).

Wing bud proteome

When the MS/MS data from the iTRAQ labelled peptides was merged and analysed using the Mascot search algorithm against the NCBInr metazoan database, 113 protein families were identified in the wing bud samples. However, to improve protein annotation and decrease the number of unassigned peptides, a second search against the tissue-specific transcriptome database was performed. Individual ions scores > 47 indicated identity or extensive homology ($p<0.05$) and there were 1,032 peptide matches above identity threshold and 2,514 matches above homology threshold for 150,293 queries, which resulted in the identification of 332 protein families. After further high-stringency filtering, i.e., where proteins with at least 2 peptides above the 95% confidence level were retained, 156 proteins identifications remained for protein quantification analysis. From these, 64 proteins were differentially expressed with 30 proteins being up-regulated and 34 proteins being down-regulated in the flatwing crickets (Fig. 4A). The number of unassigned peptides, however, only marginally decreased from the first to the second search (from 147,637 using the NCBI database to 146,272 using the tissue-specific transcriptome database).

Most DE protein annotations were related to muscle and cuticle/chitin metabolic processes and structure and at the GO level; interesting GO annotations associated with DE proteins included signalling, response to stimulus, localization and immune system process (Fig. 4B). No functional enrichment was found but several KEGG pathways were present (Table S11), and amino sugar (a main component of chitin) metabolism was down-regulated.

Transcriptome and proteome correlation

The number of sequences present in both the wing bud RNA-seq transcriptome and in the proteome was limited. Nine DE assembled transcripts were identified in the wing bud proteome and from these, only 5 (representing 3 different cuticle proteins) were differentially expressed in both datasets. However, all of the latter showed the same expression pattern in the transcriptome and in the proteome. Also, for these, the magnitude of differential expression was more extreme at the transcriptome level than at the proteome level (Fig. 4C-D). However, it is well know that the iTRAQ quantitation method tends to suppress the degree of differential expression, due to the co-selection for fragmentation of ions not showing differential expression with those showing differential expression within the mass spectrometer (Shirran & Botting, 2010).

From the candidate gene detection test (Table 1), about 30% of the *Drosophila* wing development genes were identified in the cricket wing bud transcriptome, but of these, only 10% were DE (the transcription factor Myb, the secreted proteins Hedgehog and Vein, the receptors Dachsous, Fat, Fibroblast growth factor and Notch and other cytoplasmic proteins

Axin, Beta3 tubulin and Zipper). The proportion of genes by functional categories was similar between *T. oceanicus* and *D. melanogaster* (Fig. S4). At the proteome level, however, only 3 candidate annotations (Rho, Talin and Beta3 tubulin) were present and none was DE. These 3 annotations were also present in the wing bud transcriptome and Beta tubulin was additionally a DE transcript. Together, this strategy confirms that despite some of the candidate annotations being represented in the transcriptome (but not in the proteome), few known candidate wing development and wing venation genes appear to be differentially expressed between wild-type and flatwing crickets at the wing bud stage studied here (Table S12).

DISCUSSION

The rapidly proliferating flatwing mutation in *T. oceanicus* is associated with a broad spectrum of effects on mRNA and protein expression in developing male wing buds. These included: (i) signatures of flatwing-associated effects on several biological processes; (ii) evidence that endocrine disruption may contribute to the feminised wing phenotype; (iii) prevalent down-regulation of transcript expression in flatwing males' wing buds; (iv) candidate genes for *Drosophila* wing morphogenesis represented, but mostly not differentially-expressed, in the cricket transcriptome; (v) differences in patterns of gene expression at the transcriptome *versus* proteome level, but commonality of some functional pathways.

***Drosophila* candidate analysis and flatwing gene expression profiling**

We detected homologues for a considerable proportion of the *Drosophila* candidate gene set for wing and vein development in the *T. oceanicus* wing bud transcriptome, despite the evolutionary distance between these taxa. However, very few of these showed differential expression. The necessary genetic machinery for wing and vein development appears to be present and detectable, yet is not differentially expressed at this developmental stage/tissue. Another explanation is that genes directing highly specialised male forewing development in crickets are not well-characterised and are therefore less likely to be represented in our candidate gene set and differentially expressed between morphs. In contrast, genes we detected from the *Drosophila* candidate set may play a more conserved role in insect wing development, and therefore show patterns of regulatory canalisation in *T. oceanicus* males regardless of their morph. Further investigation using a more fine-grained tissue-specific and developmental time course should clarify this. Additionally, *doublesex*, involved in wing patterning in the butterfly *Papilo polytes,* was detectable in the wing bud transcriptome but was not DE in this *T. oceanicus* tissue.

Our results suggest an interesting pattern of lower transcriptional variation in flatwing crickets than in wild-type crickets, particularly when looking at the DE transcripts that maximize differences between morphs (Fig. 3). This might reflect lower genetic variation in genomic regions linked to the causal flatwing mutation(s) on the X chromosome after a selective hard sweep during the rapid spread of the mutation in the wild (Pardo-Diaz *et al.,* 2015). Owing to the apparent recentness of the event, such a reduction in genetic diversity in a large region of the genome might be expected to canalise expression levels of any genes with coding or regulatory regions residing in those regions. However, when examining

expression patterns across all transcripts recovered in the RNA-seq experiment, there was no apparent difference between normal and flatwing males (Figs. 3a&b). This supports the idea that no large-scale genomic features differ between the morphs, for example an entire X chromosome which would comprise approximately 20% of a male's genome (K. Klappert, pers. comm.).

Gene expression analysis GO terms such as response to stimulus, reproduction, immune system process and localization were differentially expressed between morphs and particularly, locomotion, circadian rhythm and signalling were consistently down-regulated in flatwings. Additionally, functional enrichment analysis mainly showed down-regulation of transport activity and KEGG analysis recovered potential endocrine disturbance and reduced metabolism of external substances. Curiously, from KEGG analysis of the DE proteins, the metabolism of terpenoids and polyketides that can be involved in steroid production were also down-regulated in flatwings and the production of the main component of chitin was similarly decreased. The prevalence of transcriptomic down-regulation suggests gene expression pathways are reduced or shut-down in flatwing males. The strong signature of change in endocrine pathways suggests a genetic mechanism underlying this rapidly evolving phenotype, given that flatwing males have female-like wings, though it is difficult to disentangle whether DE of endocrine pathways is a leading or following effect of the development of mutant wing morphology.

Some of the most interesting genomic functions we recovered were not related to wing morphogenesis, but instead had annotations and GO terms associated with behavioural or other physiological functions. With the caveat that GO annotations are unlikely to reflect the

full range of phenotypic effects of a transcript or pathway, plus the additional caution that functional information derived from e.g. *Drosophila melanogaster* studies will not necessarily reflect the function of orthologues in distantly-related organisms, it remains informative to examine enriched GO categories to develop hypotheses about the relationship between a trait under selection, flatwing, and other traits whose expression might impact the response to selection. For example, genes implicated in social and male courtship behaviour were DE in flatwing males. Flatwing males require greater investment in social behaviours, for example to encounter females using satellite tactics (Zuk *et al.,* 2006; Bailey *et al.,* 2010) and to mitigate costly agonistic encounters (Logue *et al.,* 2010). Additionally, a recent study found that flatwing males show increased locomotion in response to a lack of social cues in their environment (Balenger & Zuk, 2015) and locomotion related annotations were also DE in flatwing males in this experiment. Flatwing males also differ constitutively in the composition of their cuticular hydrocarbon (CHC) profiles, with mutant males showing a shift in abundance of long vs. short-chained hydrocarbons (Simmons et al. 2014).

Differences in reproductive phenotypes that have been documented in field and lab experiments with flatwing males (e.g. Zuk *et al.,* 2006; Bailey *et al.,* 2010) could be attributable to the effects of mating experience or plastic changes caused by the lack of perception of song in the environment of flatwings. However, the fact that in the present study, genes associated with reproductive, immunological and behavioural processes are differentially expressed in non-reproductive tissues, in developing juveniles that can neither reproduce nor perceive acoustic signals in their environment, implies either that there are numerous constitutive differences in gene expression arising from associated effects of the causal flatwing mutation(s) such as pleiotropy, rapid correlated responses to selection,

genomic 'hitchhiking', or extensive pleiotropy, or that these genes in fact have as-yet undescribed functions directly involved in the development of male-specific wing venation.

Combining and contrasting 'omic expression profiles

Transcriptome and proteome profiles of male wing buds provided valuable counterpoints to one another. While many DE transcripts between wild-type and mutant crickets were recovered, identification of DE proteins was limited by proteome coverage. Surprisingly, however, the relative proportion of DE proteins was an order of magnitude higher than the relative proportion of DE transcripts. This observed difference might reflect the canalization and general abundance of proteins that physically build the different wing phenotypes, for example chitin/structural proteins associated with variation in vein density, thickness and positioning between the wing morphs. In contrast, the transcriptomic machinery necessary to cause such proteomic differences might be more dynamic and stochastic, such that RNA expression differences ultimately become buffered by the processes translating genotypic variation into phenotypic variation.

Only 3 transcripts were DE in both the transcriptome and proteome datasets but for these, the trend of expression was the same in each. One explanation is that, at the wing bud stage, few DE genes are translated into DE proteins, perhaps due to post-transcriptional and translational modifications. However, when transcript and associated protein were both DE their expression patterns were consistent, thus these 3 proteins warrant further investigation. Another technical explanation is that proteome coverage was not as deep as our transcriptome coverage. Nevertheless, when comparing GO terms, both transcriptome

and proteome datasets showed differential expression of response to stimulus, localization, signalling and immune process, indicating that these biological processes might be particularly important for the formation of flatwing vs. normal wing venation during development. Also, KEGG endocrine pathways were down-regulated at both levels and chitin production processes were reduced in flatwings.

Conclusion

Understanding the full spectrum of gene expression changes during rapid, adaptive evolutionary events in the wild will be important for predicting the evolutionary pace and direction of organisms subject to novel selection pressures (Pardo-Diaz *et al.,* 2015). In *T. oceanicus*, flatwing segregates as a single-locus trait (Tinghitella, 2008; Pascoal *et al.,* 2014), suggesting relatively simple genetic control. Accumulating evidence suggests that flatwing's rapid response to selection on several Hawaiian islands may be linked to the expression of other traits, including plasticity in female responsiveness to acoustic signals (Bailey *et al*., 2008, Bailey & Zuk 2008, Tinghitella *et al*., 2009), male reproductive tactics and morphology (Bailey *et al*., 2010), immunity (Bailey *et al*., 2011), locomotion (Balenger & Zuk 2015), and cuticular hydrocarbon profiles (Simmons *et al*., 2014). The overall complexity of biological processes and molecular functions that we found to change either directly or indirectly as a result of the flatwing mutation provides strong counterevidence against the hypothesis that its molecular effects are restricted to only one or a handful of developmental modules. A peculiar feature of this finding is that we detected such changes in a tissue, wing bud, in which we would not have predicted *a priori* to observe differential expression of genes with effects on seemingly unrelated physiological processes or behaviours. This may reflect

inherent limitations of functional information: the genes mentioned above might in fact play pivotal roles in differentiating normal male wings from female wings in *T. oceanicus*, but naturally such information will not be available from annotations derived from other species. Our analysis of expression variation associated with flatwing is consistent with conceptions of "universal pleiotropy" ((Fisher, 1930; Wright, 1968) reviewed in (Stearns, 2010; Paaby & Rockman, 2013)), although the expression variation associated with flatwing need not be strictly caused by pleiotropic effects of the causative mutation(s). Ultimately, the response to selection of any mutation in any organism is determined by the complete sum of its fitness effects, but our ability to predict factors that constrain or accelerate that response will be sharpened with knowledge of the number and extent of such effects.

ACKNOWLEDGEMENTS:

We are thankful to Richard Gregory for assistance during initial processing and quality assessment of the RNA-seq data, Matt Fuszard for support uploading the *T. oceanicus* transcriptome databases into Mascot, Jose Peregrina for advice on NuPAGE gels run, Kirsten Klappert for estimating genome size of the crickets by flow cytometry and Michael G. Ritchie for providing feedback that improved the manuscript. This work was funded by Natural Environment Research Council grants (NE/I027800/1, NE/G014906/1, NE/L011255/1). RNA-seq and bioinformatics was supported by the NERC Biomolecular Analysis Facility at the University of Liverpool (NBAF717). The proteome analysis was supported by the Wellcome Trust [grant number 094476/Z/10/Z] which funded the purchase of the TripleTOF 5600 mass spectrometer at the St Andrews BSRC Mass Spectrometry and Proteomics Facility.

REFERENCES:

Bailey, N. & Macleod, E. 2014. Socially flexible female choice and premating isolation in field crickets (*Teleogryllus* spp.). *J. Evol. Biol.* **27**: 170-180.

Bailey, N.W., Gray, B., Zuk, M. 2010. Acoustic experience shapes alternative mating tactics and reproductive investment in male field crickets. *Curr. Biol.* **20**: 845-849.

Bailey, N.W., Gray, B., Zuk, M. 2011. Exposure to sexual signals during rearing increases immune defence in adult field crickets. *Biol. Letters.* **7**: 217-220.

Bailey, N.W., McNabb J.R., Zuk, M. 2008. Preexisting behavior facilitated the loss of a sexual signal in the field cricket *Teleogryllus oceanicus*. *Behav. Ecol.* **19**: 202-207.

Bailey, N.W. & Zuk, M. 2008. Acoustic experience shapes female mate choice in field crickets. *Proc. R. Soc. B.* **275**: 2645-2650.

Baker, N.E. 1988. Localization of transcripts from the *wingless* gene in whole *Drosophila* embryos. *Development* **103**: 289-298.

Balenger, S.L. & Zuk, M. 2015. Roaming Romeos: male crickets evolving in silence show increased locomotor behaviours. *Anim. Behav.* **101**: 213-219.

Blair, S.S. 2007. Wing vein patterning in *Drosophila* and the analysis of intercellular signaling. In: *Annual Review of Cell and Developmental Biology*, pp. 293-319.

Charlat, S., Hornett, E.A., Fullard, J.H.*, et al.* 2007. Extraordinary flux in sex ratio. *Science* **317**: 214-214.

Conesa, A., Gotz, S., Garcia-Gomez, J.M.*,* Terol, J., Talon, M., Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Cook, L.M. & Saccheri, I.J. 2013. The peppered moth and industrial melanism: evolution of a natural selection case study. *Heredity* **110**: 207-212.

Costello, J.C., Dalkilic, M.M., Beason, S.M.*, et al.* 2009. Gene networks in *Drosophila melanogaster*: integrating experimental data to predict gene function. *Genome Biol.* **10**.

Esperk, T., Tammaru, T., Nylin, S. 2007. Intraspecific variability in number of larval instars in insects. *J. Econ. Entomol.* **100**: 627-645.

Fisher, R.A. 1930. The Genetical Theory of Natural Selection. *Clarendon Press, Oxford*.

Fuse, N., Hirose, S., Hayashi, S. 1996. Determination of wing cell fate by the escargot and snail genes in *Drosophila*. *Development* **122**: 1059-1067.

Grabherr, M.G., Haas, B.J., Yassour, M.*,* Levin, J.Z., Thompson, D.A., Amit, I., *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**: 644-652.

Grant, P.R. & Grant, B.R. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**: 707-711.

Haas, B.J., Papanicolaou, A., Yassour, M.*,* Grabherr, M., Blood, P.D., Bowden, J., *et al.* 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**: 1494-1512.

Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* **9**: 357-359.

Li, B. & Dewey, C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics* **12**: 323.

Logue, D.M., Abiola, I.O., Rains, D., Bailey, N.W., Zuk, M., Cade, W.H. 2010. Does signalling mitigate the cost of agonistic interactions? A test in a cricket that has lost its song. *Proc. R. Soc. B.* **277**: 2571-2575.

Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M.J., *et al.* 2014. A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife* **3**: e01630

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

Molnar, C. & de Celis, J.F. 2013. Tay bridge is a negative regulator of EGFR signalling and interacts with Erk and Mkp3 in the *Drosophila melanogaster* wing. *Plos Genet.* **9**.

Niwa, N., Akimoto-Kato, A., Niimi, T., Tojo, K., Machida, R., Hayashi, S. 2010. Evolutionary origin of the insect wing via integration of two developmental modules. *Evol. Dev.* **12**: 168-176.

Paaby A.B. & Rockman, M.V. 2013. The many faces of pleiotropy. *Trends Genet.* **29**: 66-73.

Pardo-Diaz, C., Salazar, C., Jiggins, C.D. 2015. Towards the identification of the loci of adaptive evolution. *Methods Ecol. Evol.* **6**: 445-464.

Parra, G., Bradnam, K., Korf, I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**: 1061-1067.

Pascoal, S., Cezard, T., Eik-Nes, A.,Gharbi, K., Majewska, J., Payne, E., *et al.* 2014. Rapid convergent evolution in wild crickets. *Curr. Biol.* **24**: 1369-1374.

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

Shirran, S.L. & Botting, C.H. 2010. A comparison of the accuracy of iTRAQ quantification by nLC-ESI MSMS and nLC-MALDI MSMS methods. *J. Proteomics* **73**: 1391-1403.

Simmons, L.W., Thomas, M.L., Gray, B., Zuk, M. 2014. Replicated evolutionary divergence in the cuticular hydrocarbon profile of male crickets associated with the loss of song in the Hawaiian archipelago. *J. Evol. Biol.* **27**: 2249-2257.

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., *et al.* 2010. Adaptation genomics: the next generation. *Trends Ecol. Evol.* **25**: 705-712.

Stearns, F.W. 2010. One Hundred Years of Pleiotropy: A Retrospective. *Genetics* **186**: 767-773.

Sturtevant, M.A. & Bier, E. 1995. Analysis of the genetic hierarchy guiding wing vein development in *Drosophila*. *Development* **121**: 785-801.

Tinghitella, R.M. 2008. Rapid evolutionary change in a sexual signal: genetic control of the mutation 'flatwing' that renders male field crickets (*Teleogryllus oceanicus*) mute. *Heredity* **100**: 261-267.

Tinghitella, R.M, Wang, J.M., Zuk, M. 2009. Preexisting behavior renders a mutation adaptive: flexibility in male phonotaxis behavior and the loss of singing ability in the field cricket *Teleogryllus oceanicus*. *Behav. Ecol.* **20**: 722-728.

van't Hof, A.E., Edmonds, N., Dalikova, M., Marec, F., Saccheri, I.J. 2011. Industrial melanism in British peppered moths has a singular and recent mutational origin. *Science* **332**: 958-960.

van't Hof, A.E. & Saccheri, I.J. 2010. Industrial melanism in the peppered moth is not associated with genetic variation in canonical melanisation gene candidates. *Plos One* **5**, e10889.

Vizcaíno, J., Deutsch, E., Wang, R., Csordas, A., Reisinger, F., Rios, D., *et al.* 2014. ProteomeXchange provides globally co-ordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **30**: 223-226.

Wright, S. 1968. Evolution and the Genetics of Populations. Vol. 1. Genetics and Biometrical Foundations. *University of Chicago Press, Chicago*.

Zuk, M., Rotenberry, J.T., Tinghitella, R.M. 2006. Silent night: adaptive disappearance of a sexual signal in a parasitized population of field crickets. *Biol. Letters.* **2**: 521-524.

DATA ACCESSIBILITY:

RNA-seq trimmed data: NCBI Bioproject ID PRJNA283744.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaíno *et al.* 2014) via the PRIDE partner repository with the dataset identifier PXD002417 (gel chunks) and PXD002451 (iTRAQ).

AUTHOR CONTRIBUTIONS:

Conceived and designed the experiment: SP and NWB; Trancriptomics sample preparation and sequencing: SP and NR; Transcriptomics bioinformatics and statistical analysis: SP, XL, YF and SteveP; Proteomics sample preparation and MS/MS analysis: SP, CHB and SLS; Proteomics data analysis: SP, CHB and TL; Wrote the manuscript: SP and NWB.

TABLES:

Table 1. Candidate gene detection test. The number and percentage of *Drosophila* wing morphology gene annotations from the database (http://www.sdbonline.org/sites/fly/aimorph/wing.htm) found through homology searches to be present in the *T. oceanicus* wing bud transcriptome, proteome and associated DE lists.

	Drosophila	**Transcriptome**	**DE transcripts**	**Proteome**	**DE proteins**
Transcription factors	94	30	1	0	0
JAK-STAT pathway	3	1	0	0	0
EGF-R signalling	6	4	0	0	0
Secreted proteins	12	6	2	0	0
Receptors and cell surface	33	7	4	0	0
Neural differentiation	14	8	0	0	0
Other cytoplasmic proteins	69	23	3	3	0
Other proteins	65	13	0	0	0
All	**296**	**92**	**10**	**3**	**0**
% found in *Drosophila* gene set		31	3 -10*	1	0

*Percentage in relation to all the searched candidates (3%) or in relation to just the candidates present in the *T. oceanicus* transcriptome (10%).

FIGURES:

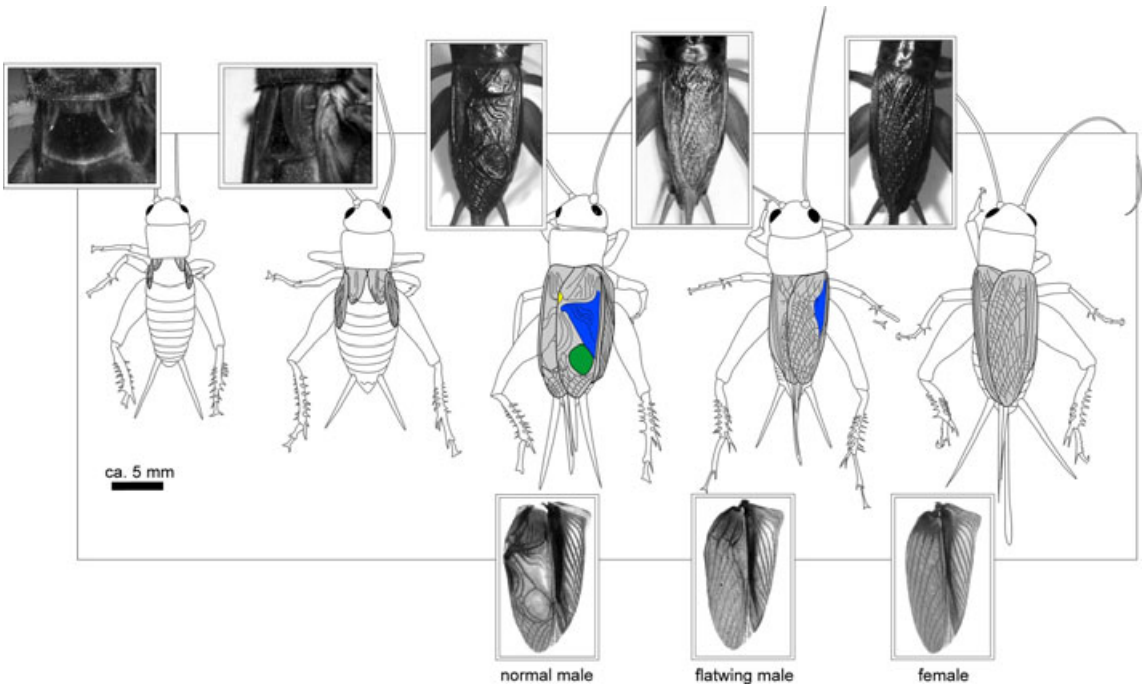

Figure 1. The final three developmental stages in *Teleogryllus oceanicus*. The left two figures highlight developing wing buds in males of the final two nymphal instars (shaded grey). The smaller, medial wing buds develop into forewings while the larger, lateral wing buds develop into hindwings. The middle figure depicts a normal male. Note the forewing lateral axial inversion between the penultimate and adult stages. Principal structures of the adult male forewing include the scraper (yellow), harp (blue) and mirror (green). The figure to the right of centre is an adult flatwing male, with the reduced harp highlighted, and the far right shows an adult female for comparison. The top row of inset photographs are magnified details of exemplar wings (not to scale). The bottom row of inset photographs are exemplars of mounted right forewings for each of the three adult types. Contrast, brightness and intensity were adjusted for clarity in CorelDRAW v.12. Photo credits: Nathan W. Bailey, David G. Forbes and Sonia Pascoal.

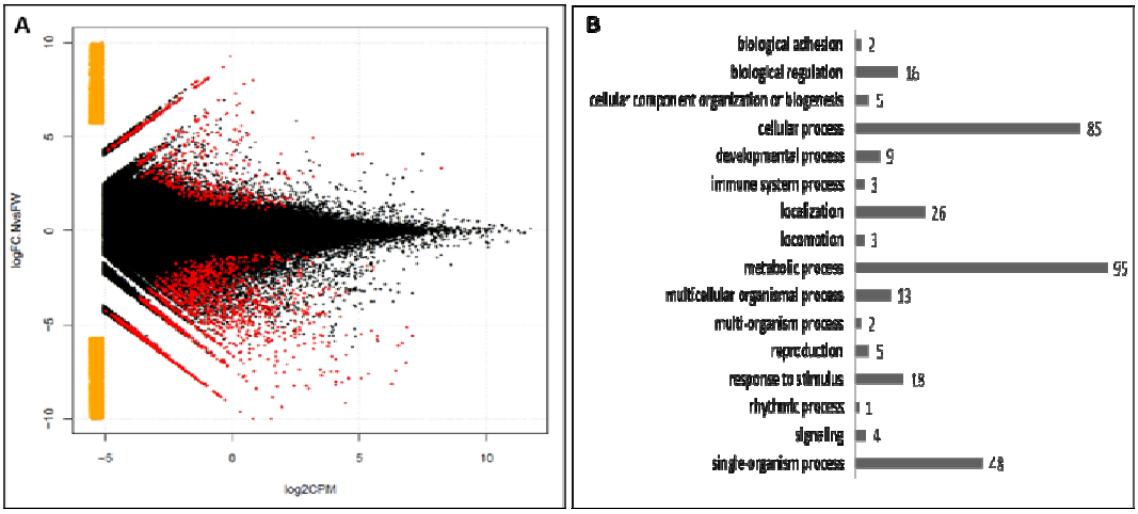

Figure 2. Wing bud RNA-seq gene expression differences between wild-type and flatwing males. A) MA plot. Red highlights the differentially expressed genes; FC: fold change; CPM: counts per million. For small count values, the corresponding log2(CPM) values will be negative and fall far away from zero. If the log2(CPM) values < -5, then they were forced to be -5 and corresponding points were coloured orange; B) Wing bud gene ontology terms: differentially expressed sequences for level 2 Biological Process. For level 3 Biological Process and Molecular Function GO terms see table S7.

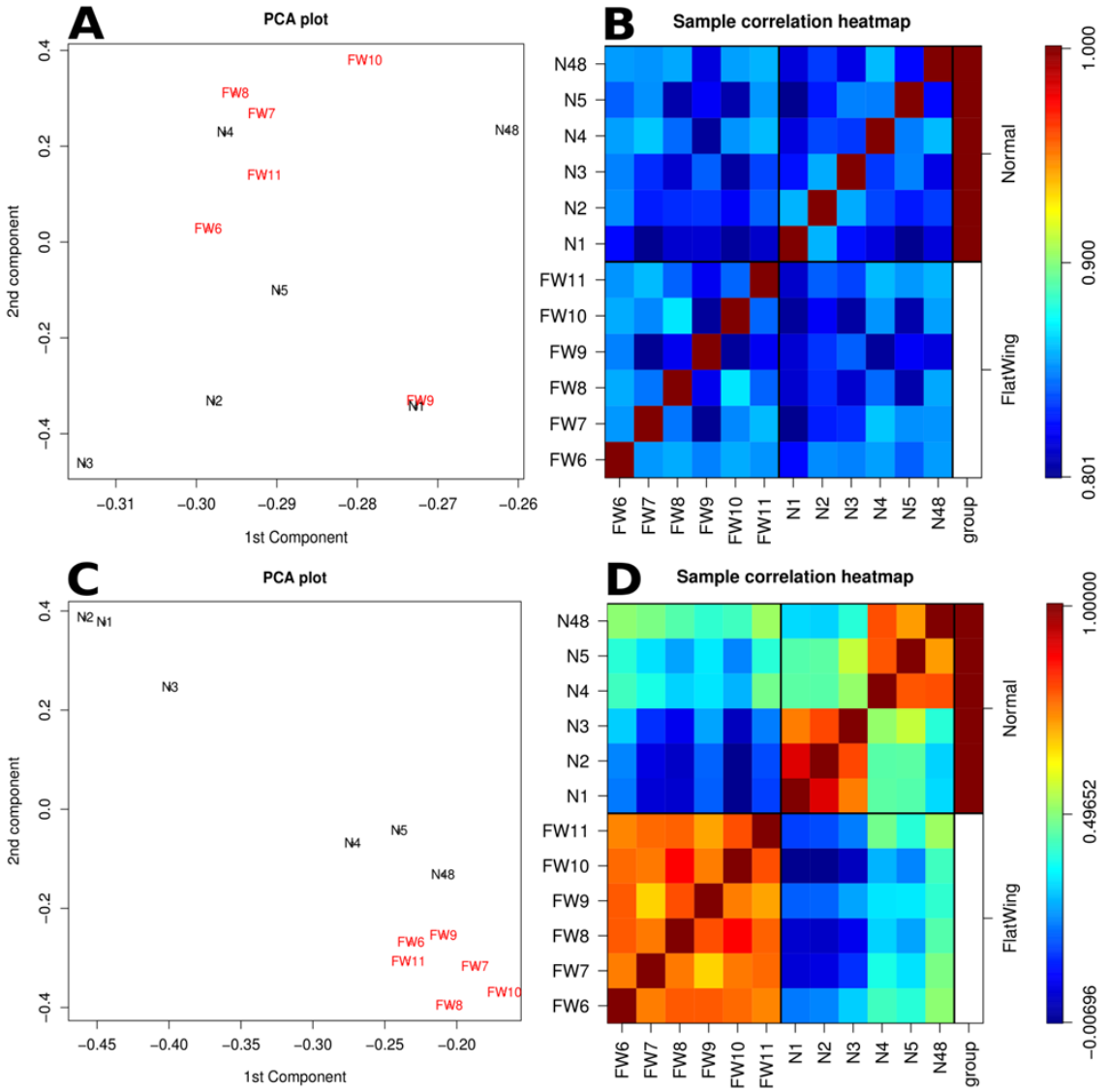

Figure 3. Wing bud gene expression analysis. A) Principal component analysis (PCA) for all assembled sequences; B) heat map for sample correlations using all assembled sequences; C) PCA for sequences DE between flatwing and normal males; D) heat map for sample correlations using only DE sequences. Labels refer to sample identity: N=normal, FW=flatwing, and numbers indicate biological replicate.

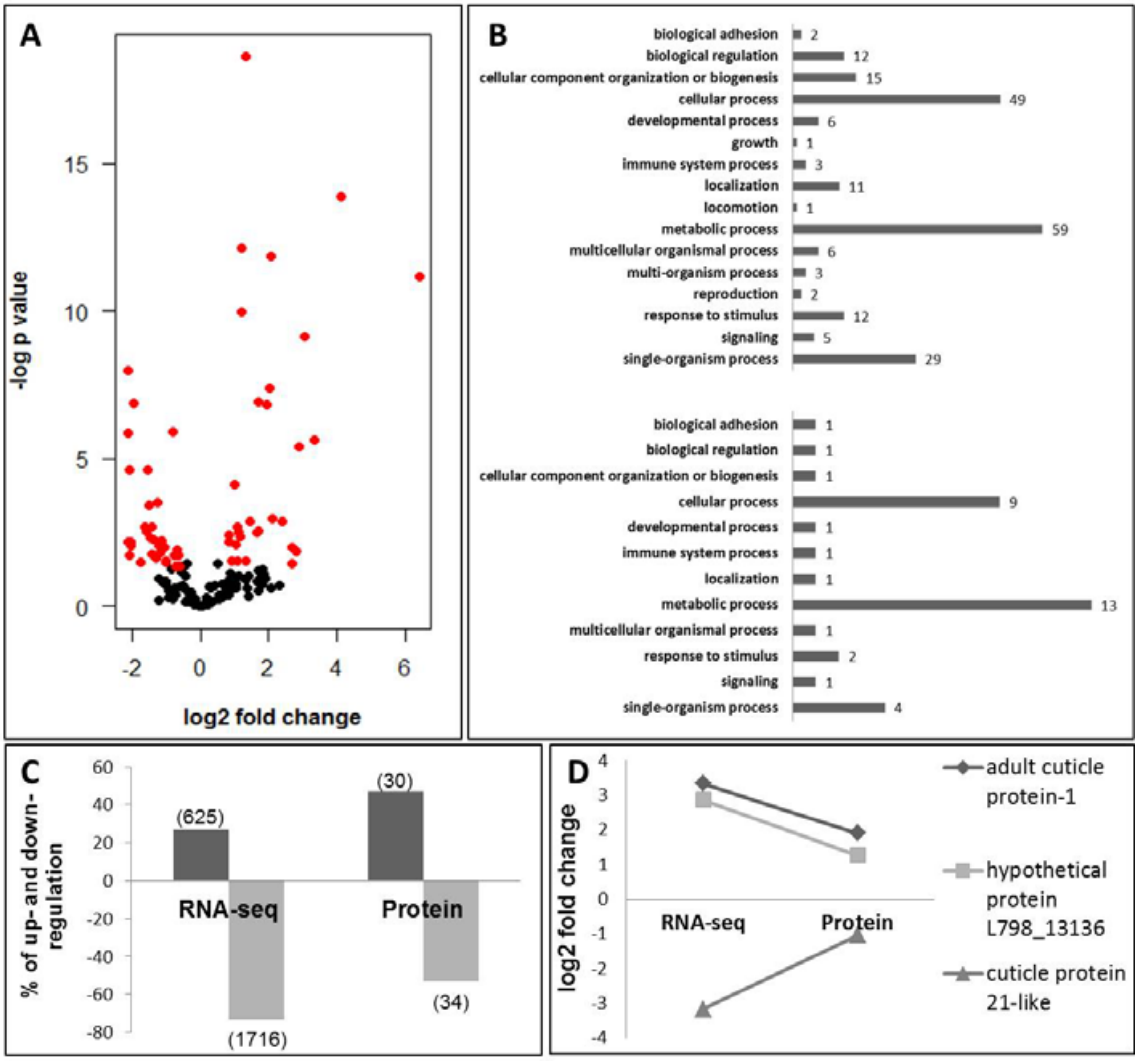

Figure 4. Wing bud proteomics and correlation with transcriptomics. A) Protein expression volcano plot. Differentially expressed proteins are highlighted in red; B) Proteome gene ontologies for level 2 Biological Process; top: all proteome (156 proteins identified and quantified), bottom: DE proteins; C) Percentage of up- *vs* down-regulated DE genes and proteins (number of DE genes and proteins indicated in parentheses); D) fold change comparison for the three differentially expressed transcripts and proteins common to both datasets.