Gene expression changes elicited by a parasitic B chromosome in the grasshopper *Eyprepocnemis plorans* are consistent with its phenotypic effects

Beatriz Navarro-Domínguez[1, 2], María Martín-Peciña[1], Francisco J. Ruiz-Ruano[1], Josefa Cabrero[1], José María Corral[3, 4], María Dolores López-León[1], Timothy F. Sharbel[3, 5] and Juan Pedro M. Camacho[1*]

[1] *Departamento de Genética, Facultad de Ciencias, Universidad de Granada, Avda. Fuentenueva s/n, 18071 Granada, Spain*

[2] Present address: *Department of Biology, University of Rochester, Rochester, NY 14627 U.S.A.*

[3] *Leibniz Institute for Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Corrensstrasse 3, D-06466 Stadt Seeland, Germany*

[4] Present address: *Friedrich-Alexander-Universität Erlangen-Nürnberg, Division of Biochemistry, Staudtstr. 5, D-91058 Erlangen, Germany*

[5] Present address: *Global Institute for Food Security (GIFS), 10 Gymnasium Place, University of Saskatchewan, Saskatoon, SK S7N 4J8 Canada*

[*]Correspondence

Short title: Gene expression changes elicited by a parasitic B chromosome

Keywords: B chromosome, gene expression, microarrays, qPCR, transcriptome

Abstract

Parasitism evokes adaptive physiological changes in the host, many of which take place through gene expression changes. This response can be more or less local, depending on the organ or tissue affected by the parasite, or else systemic when the parasite affects the entire host body. The most extreme of the latter cases is intragenomic parasitism, where the parasite is present in all host nuclei as any other genomic element. Here we show the molecular crosstalk between a parasitic chromosome (also named B chromosome) and the host genome, manifested through gene expression changes. The transcriptome analysis of 0B and 1B females of the grasshopper *Eyprepocnemis plorans*, validated by a microarray experiment performed on four B-lacking and five B-carrying females, revealed changes in gene expression for 188 unigenes being consistent in both experiments. Once discarded B-derived transcripts, there were 46 differentially expressed genes (30 up- and 16 down-regulated) related with the adaptation of the host genome to the presence of the parasitic chromosome. Interestingly, the functions of these genes could explain some of the most important effects of B chromosomes, such as nucleotypic effects derived from the additional DNA they represent, chemical defense and detoxification, protein modification and response to stress, ovary function and regulation of gene expression. Collectively, these changes uncover an intimate host-parasite interaction between A and B chromosomes during crucial steps of gene expression and protein function.

Introduction

Parasitism is very frequent in nature, and can be disguised in multiple forms and at several levels, from genes to species. As part of the molecular crosstalk between the different partners, parasites evoke gene expression changes in the host (Soumana et al. 2014). The most conspicuous parasites usually focus on a discrete organ or stage of host life cycle, and as such host response can be limited to the affected parts. However, intragenomic parasites are present in all host cell nuclei, hence host response needs to be systemic. Supernumerary (B)

chromosomes are a frequent genomic component of a wide variety of eukaryotes, behaving as parasitic elements whose spread in natural populations is based on transmissional advantage (drive), following the typical arms race dynamics of host-parasite symbioses (Camacho et al. 2000). As intragenomic parasites, B chromosomes can trigger gene expression changes at any stage of host ontogeny. On this basis, B chromosomes constitute an excellent model to shed light upon intragenomic adaptive changes in gene expression during parasitism.

Up to 12 years ago, the only DNA sequences known on B chromosomes were repetitive DNA sequences such as satellite DNA, ribosomal DNA and transposable elements (TEs) (Camacho 2005). However, from 2005 onwards, protein-coding genes or pseudogenes have been found in the B chromosomes of several species (Graphodatsky et al. 2005; Teruel et al. 2010; Yoshida et al. 2011; Martis et al. 2012; Trifonov et al. 2013; Banaei-Moghaddam et al. 2013; Valente et al. 2014; Huang et al. 2016; Carmello et al. 2017; Ma et al. 2017; Navarro-Dominguez et al. 2017a,b). The general belief of genic inactivity of B chromosomes (Camacho et al. 2000) has also changed during last years by the findings of transcription of B chromosome sequences (Leach et al. 2005; Van Vugt et al. 2005; Ruiz-Estévez et al. 2012; Carchilan et al. 2009; Zhou et al. 2012; Trifonov et al. 2013; Banaei-Moghaddam et al. 2013; 2015; Valente et al. 2014; Huang et al. 2016; Ma et al. 2017; Navarro-Dominguez et al. 2017a,b).

Gene activity of B chromosomes can potentially elicit a gene regulation response from the host genome (i.e. A chromosomes), and it has been investigated in the parasitic wasp *Nasonia vitripennis* (Akbari et al. 2013), maize (Huang et al. 2016) and rye (Ma et al. 2017) by means of comparative transcriptome analysis of B-carrying and B-lacking individuals. In *N. vitripennis*, transcriptome analysis in testes led to the identification of nine transcripts lacking homology to any known DNA sequence, which were expressed only in the B-carrying transcriptome and, by means of fluorescence *in situ* hybridization, Akbari et al. showed that at least three of them were physically located in the B chromosome (i.e. the Paternal Sex Ratio chromosome, PSR). Contrary to their expectations, they did not find any traces of gene expression changes associated to PSR chromosome presence in relation with the known effects of this B

chromosome, such as chromatin structure or condensation, transposable elements or small RNA regulation pathways (Akbari et al. 2013). In maize, B chromosome presence influences A-genome transcription, with 130 differentially expressed genes mainly involved in cell metabolism and nucleotide binding (Huang et al. 2016). Differential expression of genes related with metabolism and ATP syntesis has also been reported for B-carrying rye plants (Ma et al. 2017). These latter authors suggested that physiological effects, associated with the presence of Bs, may partly be explained by the activity of B-located genes or pseudogenes.

Almost all natural populations of the grasshopper *Eyprepocnemis plorans* in the circum-Mediterranean region carry B chromosomes (López-León et al. 2008). The presence of a same B variant (B1) in populations from Spain, Morocco, Tunisia and Sicily suggests a recent spread of B chromosomes into these areas (Cabrero et al. 2014). The high success of B chromosomes in this species results from their transmission advantage during female meiosis (Zurita et al. 1998) in spite of slight elimination during spermiogenesis (Cabrero et al. 2018). Although the presence of these B chromosomes does not influence body size (Camacho et al. 1980; Martín-Alganza et al. 1997), some endophenotypic effects have been found on the activity of the nucleolar organizer regions (NORs) (Cabrero et al. 1987; López-León et al. 1995; Teruel et al. 2007), chiasma frequency (Camacho et al. 2002), egg fertility and clutch size (Zurita et al. 1998; Bakkali et al. 2010), spermatid formation (Teruel et al. 2009) and heat shock protein 70 (Hsp70) level (Teruel et al. 2011).

B chromosomes in *E. plorans* are able to transcribe their ribosomal DNA and organize a nucleolus (Ruiz-Estévez et al. 2012), but this occurs only in a minority of males in most populations (Ruiz-Estévez et al. 2013) and the relative rRNA contribution of the B chromosome is insignificant compared to that of A chromosomes (Ruiz-Estévez et al. 2014), suggesting that B chromosomes in this species are highly repressed. However, our recent finding of ten protein-coding genes residing in the B chromosome of this species, six of which were actively expressed (Navarro-Dominguez et al. 2017a), indicates that B chromosomes are not so silenced as previously thought, and suggests the possibility that a transcriptomic crosstalk may be

operating between A and B chromosomes in B-carrying individuals.

To investigate this interesting possibility, we examine here changes in genome-wide gene expression patterns associated with the presence and absence of B-chromosomes. To do so we used two different techniques (transcriptome and microarrays analyses) on two types of samples (whole body and ovaries) from two populations harbouring two different B chromosome variants, namely B2 and B24, the latter being derived from the former (Henriques-Gil and Arana 1990).

Materials and methods

Materials

E. plorans individuals were collected in October 2012 at the Torrox (Málaga) and Salobreña (Granada) populations, harbouring the B24 and B2 variants, respectively All of them were collected the same day and were adults.

The number of B chromosomes was determined by C-banding of interphase hemolymph nuclei in females (Cabrero et al. 2006). B chromosome presence/absence was also confirmed by PCR amplification of the B-specific SCAR marker described in Muñoz-Pajares et al. (2011) on genomic DNA.

Complete bodies of two females from Torrox (0B and 1B) were used for the RNA Illumina sequencing experiment, both had been raised in our lab from individuals collected in October 2012. Nine females from Salobreña (four 0B and five 1B) were dissected the same day of field collection in order to obtain their ovaries which were flash frozen in liquid nitrogen and stored at -80ºC until RNA extraction for the microarray experiment.

Illumina sequencing

Figure 1 shows a summary of the molecular methods employed here. Total RNA was extracted from each of two females, one lacking B chromosomes (0B) and the other carrying 1B. Both libraries were prepared and sequenced by Macrogen (Inc) with 1 μg of each total RNA in a Illumina HiSeq2000 platform following manufacturer's standard protocol, each yielding about 5Gb of paired-end reads (2x101 nt). Illumina sequences are available in NCBI SRA database under accession numbers SRR2969416 (RNA_0B) and SRR2969417 (RNA_1B).

Transcriptome assembly, annotation and differential expression

De novo transcriptome assembly was carried out using Trinity software release 20131111 (Grabherr et al. 2011). Read preparation prior to assembly and downstream analysis of the *de novo* transcriptome were performed following the guidelines provided by Haas et al. (2013). We used Trimmomatic (Bolger et al. 2014) to remove adapters, low quality or N bases with quality lower than Q3 at the beginning and the end of the reads, nucleotides with an average quality lower than Q15 in a sliding window of four bases, and those reads which remained smaller than 36 bases long.

Prior to assembly, we normalized the libraries by kmer coverage in order to reduce redundant information from the deep sequencing reads, with the aim of making assembly easier, shortening computational time, increasing the chance for detection of rare transcripts, and avoiding bias derived from differences in expression between genes (Haas et al. 2013). For this purpose, we used the *normalize_by_kmer_coverage.pl* script provided by Trinity with the --JM 50G, --max_cov 30, --pairs_together and --PARALLEL_STATS parameters. This approach

extracts K-mers with the Jellyfish algorithm (Marçais and Kingsford 2011) so that each transcriptome read is probabilistically selected based on its median k-mer coverage value and the targeted maximum coverage value (Haas et al. 2013). Reads from the two libraries (0B and 1B) were assembled as a pooled data set, in order to assemble a reference *de novo* transcriptome. Assembled sequences being shorter than 200 base pairs were discarded.

Contamination in the *de novo* assembled transcriptome was examined using the standalone version of DeconSeq (Schmieder and Edwards 2011). This program aligns the sequences of our transcriptome to a database containing sequences from possible sources of contamination, using BWA (Li and Durbin 2009) and discarding matched sequences. We ran this software using the viruses, bacteria and human databases provided by the developers, plus two custom databases built from the genomes of *Saccharomyces cerevisiae* and *Arabidopsis thaliana*, which were downloaded from the NCBI FTP server (ftp://ftp.ncbi.nlm.nih.gov/genomes/). Subsequent analyses were carried out with the decontaminated assembly.

Functional annotation was done following the Trinotate (release 20140708) pipeline (Grabherr et al. 2011). Protein coding sequences (CDS) were predicted using TransDecoder (Haas et al. 2013), detecting open reading frames (ORFs) with 300 bp minimun length. Sequence homology search was performed with BLASTX of the transcripts and BLASTP (Altschul et al. 1990, 1997) of the predicted proteins against UniProtKB/Swiss-Prot and UniProt Reference Clusters (UniRef) databases (Uniprot Consortium 2014), using default settings. In addition, protein domains were analyzed with HMMER (Finn et al. 2011) and PFAM (Punta et al. 2011).

Gene function classifications were performed according to two standarized methods: Gene Ontology (GO) (Ashburner et al. 2000) and Eukaryotic Orthologous Groups (KOG) (Tatusov et al. 2003). GO assignments to predicted proteins were performed with Trinotate, and KOG classification was performed with the WebMGA software (Wu et al. 2011), searching the KOG database of NCBI.

In addition, we screened the *de novo* assembly transcriptome for known transposable elements (TEs), by means of RepeatMasker (Smit et al. 1996) in a database including TEs described in *Locusta migratoria* (data obtained from Repbase, Jurka et al. 2005).

Trinity output consists of a set of sequences (called "isoforms" in Trinity release 20131111, but we will call them simply "sequences" to avoid confusion) grouped into "clusters". In the case of protein-coding sequences, all sequences included in the same cluster and coding for homologous proteins will be considered as sequence variants of the same gene, i.e. a unigene. The sequences within unigenes being most similar to the homologous sequences in the annotation database were used as representative when needed. In case of sequences showing homology with transposable elements, we grouped those showing 80% or higher similarity using CD-HIT-EST (Fu et al. 2012), with default options except -c 0.8, in order to remove redundancy and assembly artifacts. Sequences of protein-coding genes putatively located in the B chromosome were individually inspected.

Differential expression analysis based on Illumina sequencing

For differential expression analysis, Illumina reads were mapped against the reference transcriptome in order to get an estimate of read abundance for each gene or isoform. Prior to mapping, libraries were trimmed with Trimmomatic (Bolger et al. 2014), in order to remove adapters, bases with quality lower than Q3 in the beginning and the end of the read, and bases with an average quality lower than Q15 in a sliding window of four bases. After trimming, reads that became smaller than 36 bases long were discarded. Mapping was performed using the Bowtie algorithm (Langmead et al. 2009) and read abundance per gene and sequence was estimated by RSEM (Li and Dewey 2011) and expressed in FPKM (Fragments Per Kilobase Per Million Fragments Mapped). Differential expression was analyzed using edgeR (Robinson et al. 2010). RSEM, Bowtie and edgeR were used as implemented in the Trinity pipeline (Haas et al.

2013).

Differential expression analysis based on microarrays

For microarray analysis, we used nine females (four 0B and five 1B) from Salobreña which were hybridized separately in three 3x1.4M custom NimbleGen microarrays. Given that, in this species, B chromosome drive takes place during female gametogenesis, any effect the B chromosome could have on the transcriptome would be more conspicuous and interesting in the ovaries. We extracted total RNA from the ovaries using the RNeasy Lipid Tissue Mini Kit (Qiagen). Samples were treated with DNAse I in a column as described above. Quality and absence of DNA contamination was verified in an Agilent 2100 Bioanalyzer using RNA Nanochips (Agilent Technologies). RNA quantity was determined in a Nanodrop ND-1000 spectrophotometer. 62.5 ng of RNA, which was the maximum quantity available from the least concentrated sample, was reamplified and retrotranscribed with a TransPlex Whole Transcriptome Amplification Kit (Sigma), following manufacturer's instructions and the recommendations provided in the NimbleGen Arrays User's Guide (Gene Expression Arrays v6.0).

NimbleGen Custom 3x1.4M Gene Expression arrays were designed using 434,838 *E. plorans* sequences, obtained by the assembly of the Illumina reads with TransABySS software (Robertson et al. 2010), which are available at Figshare (https://figshare.com/s/ddedc61cd7eefa716eec). Probe length was 60bp. Layout was outsourced to NimbleGen. Labelling, hybridization, washing and scanning of the arrays was performed following the instructions in the NimbleGen Arrays user's guide for gene expression arrays on an MS 200 microarray scanner.

To allocate expression values from the microarrays to the Trinity-assembled transcriptome, we searched for homology between the 434,838 TransABySS-assembled sequences used for

building the microarrays and the 73,889 sequences assembled by Trinity, by means of BLASTN (Altschul et al. 1990). Considering the number of sequences, each Trinity contig could show homology with 6 TransABySS contigs, on average. Therefore, for each Trinity sequence, the microarray expression values from up to 20 sequences showing high similarity (E-value<1e-100) were averaged.

RMA normalization of the microarray data was performed with the *oligo* R package (Carvalho and Irizarry 2010) and the differences in expression between the two groups were assessed by a *t*-test performed with the *limma* R package (Smyth 2005). Data analysis was performed following the steps described for NimbleGen arrays in the related vignette of the oligo package. Both packages are available in Bioconductor (Gentleman et al. 2004).

qPCR validation of microarray differential expression results

To increase the reliability of microarray results, eight genes (*AIG1*, *DCR1*, *MYSA*, *RIN2*, *NRX4*, *RDX*, *S39AE* and *SY65*) were selected for qPCR validation, using cDNA from the same 9 samples hybridized on the microarrays. These genes were chosen on the basis of being representative of the various functional groups that we described in the set of differentially expressed genes (DEGs), as well having diverse expression patterns and a range of expression fold change. Primer design and qPCR were carried out as described in Navarro-Dominguez et al. (2016, 2017 a, b), and are shown in Online Resource 1. qPCR was perfomed in a Chromo4 Real Time PCR thermocycler (BioRad), using SensiMix SYBR Kit (Bioline). Efficiency for each primer pair was calculated by a standard curve performed with serial 1:10 dilutions. Relative expression quantities were calculated using each gene's efficiency, referred to a calibrator sample and normalized by the geometric mean of two reference genes (actin and *RP49*), as described in Navarro-Dominguez et al. 2016.

Results

Illumina sequencing and de novo transcriptome assembly

Illumina sequencing generated 35,345,561 paired-end reads for the RNA_0B library and 27,247,068 for the RNA_1B library, implying about 0.74x and 0.64x coverage for gDNA, considering a haploid genome size of 10.525 Gb for the 0B individual and 11.885Gb for the 4B individual, according to Ruiz-Ruano et al. (2011).

After removing potential contaminants, the *de novo* assembled transcriptome included 73,889 sequences grouped into 45,555 unigenes. Sequence length ranged from 200 to 19,141 bp, with 788 bp median sequence length and N50 = 2,302. For the unigenes, median length was

485 bp and N50 was 1,702. A comparison of the assembly before and after decontamination demonstrated a low contamination level (~0.65%). A summary of Illumina sequencing results and assembly and decontamination statistics is shown in Table 1. The *de novo* assembled *E. plorans* transcriptome can be accessed in Figshare at https://dx.doi.org/10.6084/m9.figshare.3408580.v3

Functional annotation of *E. plorans* transcriptome

About 42% of the 45,555 unigenes showed significant similarity (Evalue < 10e-5) with proteins in Uniprot, Uniref90 or both. In most cases (~45%) a single potential coding sequence (CDS) was found, and a high percentage (~81%) of the sequences with a single CDS could be assigned to one or more Gene Ontology terms based on BLAST matches to sequences with known function. In other cases, more than one CDS (~27%) or no potential CDS (~28%) were found. This could be due to the expression of pseudogenes with fragmented CDS or to sequencing, assembling or ORF prediction artifacts, and this was taken into account for further analysis. The remaining 58% of the unigenes did not show significant homology with known coding proteins, 91% of them being apparently non protein-coding transcripts (i.e. lacking a CDS).

Summarizing, we found 18,999 unigenes with BLAST hits, 13,570 of which carried one or more CDSs, and 2,429 unigenes carrying a putative CDS but failing to show BLAST hits, whereas a high number of transcripts (24,127) did not bear any predicted CDS and neither show BLAST hits to known proteins (Figure 2A).

In addition, we screened the *de novo* assembled transcriptome for TEs using RepeatMasker. The number of unigenes showing homology with the TEs described in *Locusta migratoria* (data obtained from Repbase, Jurka et al.2005) was 5,555 (i.e. 12% of total unigenes), the most frequent being Mariner/Tc1 (923), Penelope (749) and RTE (638) (Figure 2B).

B chromosome presence triggers differential expression for 188 unigenes

According to edgeR results, we found that 24,462 sequences (grouped in 16,013 unigenes) showed significant differential expression between 0B and 1B females (p<0.05). We validated these transcriptome results with a microarray experiment performed on females from a different population (Salobreña) and using RNA from the ovary instead of the whole body. The *t*-test performed on microarray data resulted in 1,614 differentially expressed sequences (grouped in 1,202 unigenes) (P< 0.05) (see detailed results in Online Resource 2). Statistical corrections for multiple tests (e.g. Bonferroni) would erase significance for all but one contig showing homology with the retrotransposable element CR1 and three non-annotated sequences. Alternatively, we intersected the significant results obtained in the transcriptome and microarray experiments and considered as significant gene expression changes only those being present in both cases and in the same direction. This means that we actually tested differential expression in ovary. Since the B-carrying females used for the transcriptome and microarray experiments carried different B chromosome variants (B24 and B2, respectively), we can infer that the changes observed following our approach are common to both kinds of B chromosomes.

A comparative analysis of the identity of these sequences between the transcriptome and microarray samples revealed that 535 sequences showed significant differential expression, 258 of which showed opposite expression patterns, i.e. they were up-regulated according to one of the analyses and down-regulated according to the other. The remaining 277 sequences (grouped in 188 unigenes) showed matching expression patterns in both analyses, thus representing gene expression changes associated with the presence of B chromosomes taking place in both analyzed populations (Figure 3). Remarkably, 246 of these differentially expressed sequences (161 unigenes, i.e. 89%) were up-regulated in the presence of the B chromosome, whereas only 31 sequences (27 unigenes, i.e. 11%) were down-regulated (Table 2).

Among the 188 unigenes corresponding to the differentially expressed sequences, 95 unigenes (122 sequences) failed to show homology with any protein described in Uniprot or Uniref90 for any organism or repetitive element described for *L. migratoria* in RepBase. However, 42 out of the 93 remaining unigenes (66 sequences) showed homology with transposable elements, and the remaining 51 unigenes (89 sequences) showed homology with protein-coding genes described in Uniprot, Uniref90 or both (Table 2). This low proportion of annotations was undoubtedly due to the absence of a fully annotated genome in *E. plorans* or other grasshopper species, since the recently published genome of *L. migratoria* is still in the draft stage (Wang et al. 2014). Our subsequent interpretation of these results is thus conditioned by this partial annotation. Anyway, we will concentrate efforts on possible functional meaning.

Host genome adaptation for B chromosome presence stands on differential expression for 46 unigenes

Assuming that the five differentially expressed unigenes located in the B chromosome (i.e. *CIP2A*, *CKAP2*, *CAP-G*, *KIF20A* and *MYCB2*) showed up-regulation due to the transcription of B-located copies (Navarro-Dominguez et al. 2017a), we can delimit gene expression changes in the host genome, associated with the presence of the parasitic chromosome, to 46 unigenes (30 up- and 16 down-regulated), after discarding these five unigenes.We then analyzed the GO terms associated with these 46 unigenes and interpreted them as gene expression changes that the parasitic chromosome triggers on the host genome. They were related with post-transcriptional gene expression regulation, endoplasmic reticulum unfolded protein response and endoplasmic reticulum stress, post-transcriptional gene silencing via small RNA, histone-methyltransferase activity, DNA conformation change, protein kinase activity and regulation of cell death (see Online Resource 3). Remarkably, some of these terms define expected functional consequences of the crossfire between the genome and the parasitic chromosome (see below).

Most these functions were also apparent in terms of the KOG classification (Figure 4). Surprisingly, no gene expression changes in the host genome were related with cell cycle and cell division (KOG class D), in contrast with some of the B-located genes (Navarro-Domínguez et al. 2017a).

Drammatic up-regulation of TEs in the B-carrying transcriptome

About 22% of the 188 unigenes showing differential expression in B-carrying ovaries were annotated as TEs (Table 2). Among them we found *Mariner/Tc1* (12), *RTE* (7), *SINE* (4), Unknown (3), *Daphne* (3), *Penelope* (3), *DNA* (2), *Gypsy* (2), *CR1* (1), *hAT* (1), *L2* (1), *R1* (1), *Sola* (1) and *Vingi* (1). Remarkably, 40 out of these 42 unigenes showed up-regulation, and only two (one *Daphne* and one *Mariner/Tc1* sequences) showed down-regulation.

qPCR validation

To validate the microarrays results, eight DEGs (*AIG1*, *DCR1*, *MYSA*, *RIN2*, *NRX4*, *RDX*, *S39AE* and *SY65*) were selected for quantitative PCR validation. All of the selected genes but *RIN2* showed concordant expression patterns between the microarray and qPCR results. Despite this single discordant value, there was a highly significant positive linear correlation (R^2=0.6732; p=0.0151) among the computational fold-change calculated from the microarray data and the experimental fold-change measured by qPCR data (Figure 5), giving strong support to microarray results. *RIN2* exhibits low levels of change in the microarray data, which is known to be an important factor in the lack of concurrence between methods (Morey et al. 2006).

Discussion

B chromosomes are intragenomic parasites being in intimate contact with host genes within the same nuclei, for which reason they have the chance to elicit a true transcriptomic arms race with A chromosomes. Our present results show that this is the case. Recent transcriptomic analyses in interspecific parasitisms have unveiled gene expression changes mostly focused on parasite attack and host resistance (for instance, see Nishimura et al. 2015 and references therein). The transcriptomic signature of an intragenomic parasitism, shown here for B-carrying ovaries, shows that the presence of a parasitic chromosome evokes a response in the host genome implying gene expression changes associated with the presence of additional DNA (nucleotypic effects), sensing cellular stress, detoxifying and immune defense, ovary function and regulation of gene expression (Figure 6). These changes can be viewed as a manifestation of a molecular arms race between the parasitic and host counterparts of a same genome, which parallels population dynamics pathways (Camacho et al. 1997).

The presence of additional DNA in the form of B chromosomes can passively derive in a series of sequence-independent physiological changes known as nucleotypic effects (Bennet 1971). In several cases, B chromosomes have been shown to increase cell volume and slow cell division (for review, see Jones and Rees 1982). As Gregory (2000) remarked, "the addition of only one or two B chromosomes appears not to have a noticeable effect on cell size (John and Jones 1970), perhaps indicating that cells are able to compensate for the effects of these selfish elements". In *E. plorans*, cell or nucleus size has not been measured, but body size, which is frequently positively correlated with cell size (Gregory 2002), has been measured twice (Camacho et al. 1980; Martín-Alganza et al. 1997) and no association with the number of B chromosomes was found. It is thus conceivable that some of the observed gene expression changes are actually revealing how B-carrying cells cope with the presence of this extra DNA in terms of signal and molecule transport, intercellular communication and trafficking, membranes, ion transport, signal transduction mechanisms, intracellular trafficking, secretion, and vesicular transport.

For instance, the observed down-regulation of *PI3K2* in B-carrying *E. plorans* females (Online Resource 3) could be a response to maintain cell size in spite of the extra DNA added by the B chromosome, as this gene is involved in the regulation of cell morphogenesis by promoting cell growth (Leevers et al. 1996; Weinkove et al. 1999), and also in the regulation of exaggerated trait growth in insects (Lavine et al. 2015). It is also conceivable that the presence of the extra B-DNA elicits some changes in how cells communicate themselves, and this is reinforced by the observed expression changes in genes with functions related with membranes and ion transport, such as *S39AE* and *IL16* (Online Resource 3), which play a role in parasitic infections (Pappu et al. 2011), in strong parallelism with other types of parasitism (for instance, see Nishimura et al. 2015 and references therein).

It was highly remarkable the abundance of gene expression changes dealing with chemical defense and detoxification observed in presence of the B chromosome in *E. plorans* (Online Resource 3). Metazoan genomes contain many genes involved in responses to environmental

stressors. Chemical defense genes include cytochromes *P450* and other oxidases, various conjugating enzymes, ATP-dependent efflux transporters, oxidative detoxification proteins, and transcription factors that regulate these genes which, as a whole, account for more than 400 genes in the sea urchin genome (Goldstone et al. 2006). P450 up-regulation in B-carrying *E. plorans* females could thus be a response of the host genome against parasitic chromosomes. However, the down-regulation of a related gene, Cytochrome P450 6A1 (*CP6A1*) (Online Resource 3) makes it difficult to assess the net effect of both changes. Anyway, most of the gene expression changes dealing with chemical defense were up-regulations (Online Resource 3) and could thus constitute part of the resistance response to the parasitic chromosome.

It has been suggested that gene expression in host–parasite interactions is likely to evolve toward greater immunological surveillance and reduced parasite conspicuousness (Nuismer and Otto 2005), in a Red Queen dynamics (Barribeau et al. 2014). Remarkably, some of the gene expression changes found in B-carrying ovaries can be viewed in this way, as they affect genes involved in protein ubiquitination and degradation by the proteasome (Online Resource 3). Although we cannot know the precise mechanism of action of these gene expression changes, it is tempting to speculate that these changes regarding ubiquitinization and endoplasmic reticulum stress might constitute a host genome response to the presence of topologically anomalous polypeptides derived from B-gene or pseudogene transcripts (see Navarro-Dominguez et al. 2017a,b). The down-regulation of a histone methyltransferase (*SE1BA*) (Online Resource 3) might constitute a host genome attempt to reduce the presence of B chromosome transcripts.

Several of the observed DEGs are involved in lipid transport and metabolism, with conceivable consequences on egg fertility which, in *E. plorans*, decreases in presence of B chromosomes (Zurita et al. 1998, Muñoz et al. 1998). Six of these genes were up-regulated in B-carrying females (Online Resource 3). However, five other genes, also involved in lipid metabolism and transport, regulation of oocyte development and contraction of ovarian muscle, were down-regulated (Online Resource 3). Taken together, these gene expression changes

might reflect the harmful effects of the parasitic chromosome on yolk production in the eggs and molecule transport between cells, but more detailed analysis of these gene expression networks is needed to unveil possible causal relationships between B presence and the decrease in egg fertility in *E. plorans* females.

Four other DGEs related with muscle function were up-regulated in B-carrying females (Online Resource 3). Interestingly, *TITIN* and *MYSA* also play a role during mitosis (Machado and Andrew 2000; Fabian et al. 2007; Fabian and Forer 2007). Collectively, these four gene expression changes could be a consequence of higher effort by B-carrying females in muscle function, probably related with ovary function, and/or higher costs of building mitotic spindles in ovaries due to the higher number of chromosomes to move apart during cell division.

Among the most interesting gene expression changes observed in ovaries carrying the parasitic chromosome were those associated with the regulation of gene expression. Two DEGs were involved in the biogenesis of the 60S ribosomal subunit, one was up-regulated and the other down-regulated in B-carrying females (Online Resource 3). In addition, three up-regulated DEGs might reflect the extra transcription effort due to the higher gene expression observed in B-carrying females, with double number of up- than down-regulations for protein-coding genes (Online Resource 3) and extraordinary up-regulation for many TEs (see Table 1). Interestingly, B-carrying females showed down-regulation of Dicer 1 (*DCR1*), an RNA polymerase III being essential for RNA interference (RNAi) and microRNA (miRNA) gene silencing. Remarkably, the down-regulation of Dicer 1 in B-carrying females of *E. plorans* was associated with an explosive up-regulation for many transposable elements, in high consistency with the role of RNA interference as an important defence against viruses and transposable elements (Obbard et al. 2009, Fablet 2014). Recently, it has been shown that the hpRNA/RNAi pathway plays an important role in the suppression of the intragenomic conflict caused by the selfish sex-ratio distorters in *Drosophila* (Lin et al. 2018). The RNAi pathway could be thus implied in the neutralization of selfish DNA in general, since previous studies have also discussed the importance of this pathway in other B chromosome systems. For instance, Ma et al. (2017)

showed that A- and B-encoded Argonaute (AGO4B) protein variants possess RNA slicer *in vitro* activity, thus demonstrating unambiguously the presence of a functional AGO4B gene on rye B chromosomes. As Argonaute and Dicer 1 are two of the principal RNA silencing mechanisms, this parallelism between rye and *E. plorans* opens the door to new exciting research lines on possible relationships between B chromosomes and the RNA silencing machinery, in line with recent suggestions by Ramos et al. (2016) and Valente et al. (2017).

Our present results are logically limited by the small sample size of RNAseq analysis (see Liu et al. 2013, Schurch et al. 2016). This likely hindered the detection of an important fraction of all gene expression changes triggered by B chromosome presence. Nevertheless, our results are valuable since they show that some effects of B chromosomes, previously reported in this species (see above), can result from changes at transcriptional level. This opens new lines of thought about the molecular crosstalk between B chromosomes and the host genome by converting gene expression in a new arena for this kind of host-parasite arms race.

Taken together, the gene expression changes observed here in B-carrying ovaries of *E. plorans* appear to constitute a logical response of the host genome (summarized in Figure 6) to counteract gene expression of B chromosome genes and pseudogenes reported by Navarro-Dominguez et al. (2017a,b). It is remarkable that some of the active genes in the B chromosome (e.g. *CIP2A* and *KIF20A*), but almost none of the observed gene expression changes in the host genome, code for cell division regulation. In contrast, most changes in host genome gene expression had to do with chemical defense and detoxification, protein modification and response to stress, ovary function and regulation of gene expression, suggesting an intimate host-parasite crosstalk during crucial steps of gene expression and protein function. As a whole, our present results illuminate a broad spectrum of future molecular research on this evolutionarily interesting intragenomic parasitism, and the common molecular arena for the two counterparts makes this a good model for studies on gene regulation. Finally, some of the enriched gene ontology categories found in B-carrying ovaries of the grasshopper *E. plorans*, reported here (e.g. microtubule movement, cell division, cell metabolism, gene silencing and

protein stabilization), were also found for rye B chromosomes (Ma et al. 2017). This interesting coincidence illuminates some adaptive responses of the host genome to the presence of B chromosomes, at transcriptional level. However, the transcription of genes or pseudogenes residing in B chromosomes (Ma et al. 2017; Navarro-Dominguez et al. 2017a,b) makes it difficult to disentangle the conundrum posed by hundreds of DEGs whose A or B origin can only be distinguished through B-specific sequence changes. In *E. plorans*, we did not have this information, for which reason we discarded the ten genes residing in B chromosomes, shown by Navarro-Dominguez et al. (2017a), and observed that B chromosome presence elicits transcriptional changes explaining many of their previously reported effects at cell and physiological levels.

Integrity of research All experiments comply with the current Spanish laws.

Compliance with ethical standards

Conflict of interest   The authors declare that they have no conflict of interest

Human and animal rights and informed consent   All applicable international, national, and/or institutional guidelines for the care and use of animals were followed

References

Akbari OS, Antoshechkin I, Hay BA, Ferree PM (2013) Transcriptome profiling of *Nasonia vitripennis* testis reveals novel transcripts expressed from the selfish B chromosome, paternal sex ratio. G3-Genes Genom Genet 3:1597-1605. https://doi.org/10.1534/g3.113.007583

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

Ashburner M, Ball CA, Blake JA et al (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25:25-29. https://doi.org/10.1038/75556

Bakkali M, Manrique-Poyato MI, López-León MD, Perfectti F, Cabrero J, Camacho JPM (2010) Effects of B chromosomes on egg fertility and clutch size in the grasshopper *Eyprepocnemis plorans*. J Orthoptera Res 19:197-203. https://doi.org/10.1665/034.019.0205

Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, Houben A (2013) Formation and expression of pseudogenes on the B chromosome of rye. Plant Cell 25:2536-2544. https://doi.org/10.1105/tpc.113.111856

Banaei-Moghaddam AM, Martis MM, Macas J et al (2015) Genes on B chromosomes: Old questions revisited with new tools. Biochim Biophys Acta 1849:64–70. https://doi.org/10.1016/j.bbagrm.2014.11.007

Barribeau SM, Sadd BM, du Plessis L, Schmid-Hempel P (2014) Gene expression differences

underlying genotype-by-genotype specificity in a host–parasite system. P Natl Acad Sci USA 111:3496-3501. https://doi.org/10.1073/pnas.1318628111

Bennett MD (1971) The duration of meiosis. P Roy Soc Lond B Bio 178:277-299. https://doi.org/10.1098/rspb.1971.0066

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Cabrero J, Alché JD, Camacho JPM (1987) Effects of B chromosomes on the activity of nucleolar organizer regions in the grasshopper *Eyprepocnemis plorans*: activation of a latent nucleolar organizer region on a B chromosome fused to an autosome. Genome 29:116-121. https://doi.org/10.1139/g87-020

Cabrero J, López-León MD, Ruíz-Estévez M et al (2014). B1 was the ancestor B chromosome variant in the western Mediterranean area in the grasshopper *Eyprepocnemis plorans*. Cytogenet Genome Res 142:54-58. https://doi.org/10.1159/000356052

Cabrero J, Manrique-Poyato MI, Camacho JPM (2006) Detection of B chromosomes in interphase hemolymph nuclei from living specimens of the grasshopper *Eyprepocnemis plorans*. Cytogenet Genome Res 114:66-69. https://doi.org/10.1159/000091930

Cabrero J, Martín-Peciña M, Ruiz-Ruano FJ, Gómez R, Camacho JPM (2018) Post-meiotic B chromosome expulsion, during spermiogenesis, in two grasshopper species. Chromosoma (in press) https://doi.org/10.1007/s00412-017-0627-8

Camacho JPM (2005) B chromosomes. In: Gregory TR (ed) The evolution of the genome. Elsevier, San Diego, pp 223-286

Camacho JPM, Bakkali M, Corral JM et al (2002) Host recombination is dependent on the degree of parasitism. P Roy Soc Lond B Bio 269:2173-2177. https://doi.org/10.1098/rspb.2002.2135

Camacho JPM, Shaw MW, López-León MD, Pardo MC, Cabrero J (1997) Population dynamics

of a selfish B chromosome neutralized by the standard genome in the grasshopper

Eyprepocnemis plorans. Am Nat 149:1030–1050. https://doi.org/10.1086/286037

Camacho JPM, Carballo AR, Cabrero J (1980) The B-chromosome system of the grasshopper

Eyprepocnemis plorans subsp. *plorans* (Charpentier). Chromosoma 80:163-176.

https://doi.org/10.1007/BF00286298

Camacho JPM, Sharbel TF, Beukeboom LW (2000) B-chromosome evolution. Philos T Roy

Soc B 355:163-178. https://doi.org/10.1098/rstb.2000.0556

Carchilan M, Kumke K, Mikolajewski S, Houben A (2009) Rye B chromosomes are weakly

transcribed and might alter the transcriptional activity of A chromosome sequences.

Chromosoma 118:607-616. https://doi.org/10.1007/s00412-009-0222-8

Carmello BO, Coan RLB, Cardoso AL et al (2017) The hnRNP Q-like gene is retroinserted into

the B chromosomes of the cichlid fish *Astatotilapia latifasciata*. Chromosome Res

25:277–290. https://doi.org/10.1007/s10577-017-9561-0

Carvalho BS, Irizarry RA (2010) A framework for oligonucleotide microarray preprocessing.

Bioinformatics 26:2363-2367. https://doi.org/10.1093/bioinformatics/btq431

Fabian L, Xia X, Venkitaramani DV, Johansen KM, Johansen J, Andrew DJ, Forer A (2007)

Titin in insect spermatocyte spindle fibers associates with microtubules, actin, myosin

and the matrix proteins skeletor, megator and chromator. J Cell Sci 120:2190-2204.

https://doi.org/10.1242/jcs.03465

Fabian L, Forer A (2007) Possible roles of actin and myosin during anaphase chromosome

movements in locust spermatocytes. Protoplasma 231:201-213.

https://doi.org/10.1007/s00709-007-0262-y

Fablet M (2014) Host control of insect endogenous retroviruses: Small RNA silencing and

immune response. Viruses 6:4447-4464. https://doi.org/10.3390/v6114447

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity

searching. Nucleic Acids Res 39: W29-W37. https://doi.org/10.1093/nar/gkr367

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation

sequencing data. Bioinformatics 28:3150-3152.

https://doi.org/10.1093/bioinformatics/bts565

Gentleman RC, Carey VJ, Bates DM et al (2004) Bioconductor: open software development for

computational biology and bioinformatics. Genome Biol 5:R80.

https://doi.org/10.1186/gb-2004-5-10-r80

Goldstone JV, Hamdoun A, Cole BJ et al (2006) The chemical defensome: environmental

sensing and response genes in the Strongylocentrotus purpuratus genome. Dev Biol 300:

366–384. https://doi.org/10.1016/j.ydbio.2006.08.066

Grabherr MG, Haas BJ, Yassour M et al (2011). Full-length transcriptome assembly from RNA-

Seq data without a reference genome. Nat Biotechnol 29:644-652.

https://doi.org/10.1038/nbt.1883

Graphodatsky AS, Kukekova AV, Yudkin DV et al (2005) The proto-oncogene C-KIT maps to

canid B-chromosomes. Chromosome Res 13:113-122. https://doi.org/10.1007/s10577-

005-7474-9

Gregory TR (2000) Nucleotypic effects without nuclei: genome size and erythrocyte size in

mammals. Genome 43:895-901.

Gregory TR (2002) A bird's-eye view of the C-value enigma: genome size, cell size, and

metabolic rate in the class aves. Evolution 56:121-130. https://doi.org/10.1554/0014-

3820(2002)056[0121:ABSEVO]2.0.CO;2

Haas BJ, Papanicolaou A, Yassour M 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. https://doi.org/10.1038/nprot.2013.084

Henriques-Gil N, Arana P (1990) Origin and substitution of B chromosomes in the grasshopper *Eyprepocnemis plorans*. Evolution 44:747-753. https://doi.org/10.1111/j.1558-5646.1990.tb05952.x

Huang W, Du Y, Zhao X, Jin W (2016) B chromosome contains active genes and impacts the transcription of A chromosomes in maize (*Zea mays* L.). BMC Plant Biol 16:88. http://doi.org/10.1186/s12870-016-0775-7

John PCL, Jones RN (1970) Molecular heterogeneity of soluble proteins and histones in relationship to the presence of B-chromosomes in rye. Exp Cell Res 63:271-276. https://doi.org/10.1016/0014-4827(70)90213-2

Jones RN, Rees H (1982) B Chromosomes. London: Academic Press.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462-467. https://doi.org/10.1159/000084979

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lavine L, Gotoh H, Brent CS, Dworkin I, Emlen DJ (2015) Exaggerated trait growth in insects. Annu Rev Entomol 60:453-472. https://doi.org/10.1146/annurev-ento-010814-021045

Leach CR, Houben A, Field B, Pistrick K, Demidov D, Timmis JN (2005) Molecular evidence for transcription of genes on a B chromosome in *Crepis capillaris*. Genetics 171:269-278. https://doi.org/10.1534/genetics.105.043273

Leevers SJ, Weinkove D, MacDougall LK, Hafen E, Waterfield MD (1996) The Drosophila

phosphoinositide 3-kinase Dp110 promotes cell growth. EMBO J 15:6584-6594.

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or

without a reference genome. BMC Bioinformatics 12:323. https://doi.org/10.1186/1471-

2105-12-323

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform.

Bioinformatics 25:1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Lin CJ, Hu F, Dubruille R et al (2018) The hpRNA/RNAi pathway is essential to resolve

intragenomic conflict in the *Drosophila* male germline. Dev Cell 46: 316-326.

https://doi.org/10.1016/j.devcel.2018.07.004

Liu Y, Zhou J, White KP (2013) RNA-seq differential expression studies: more sequence or

more replication? Bioinformatics 30.3: 301-304.

https://doi.org/10.1093/bioinformatics/btt688

López-León MD, Cabrero J, Camacho JPM (1995) Changes in DNA methylation during

development in the B chromosome NOR of the grasshopper *Eyprepocnemis plorans*.

Heredity 74:296-302. https://doi.org/10.1038/hdy.1995.43

López-León MD, Cabrero J, Dzyubenko VV, Bugrov AG, Karamysheva TV, Rubtsov NB,

Camacho JPM (2008) Differences in ribosomal DNA distribution on A and B

chromosomes between eastern and western populations of the grasshopper

Eyprepocnemis plorans plorans. Cytogenet Genome Res 121:260-265.

https://doi.org/10.1159/000138894

Ma W, Gabriel TS, Martis MM et al (2017) Rye B chromosomes encode a functional

Argonaute- like protein with in vitro slicer activities similar to its A chromosome

paralog. New Phytol 213:916-928. https://doi.org/10.1111/nph.14110

Machado C, Andrew DJ (2000) D-Titin A giant protein with dual roles in chromosomes and

muscles. J Cell Biol 151:639-652. http://doi.org/10.1083/jcb.151.3.639

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of

occurrences of k-mers. Bioinformatics 27:764-770.

https://doi.org/10.1093/bioinformatics/btr011

Martín-Alganza A, Cabrero J, López-León MD, Perfectti F, Camacho JPM (1997)

Supernumerary heterochromatin does not affect several morphological and physiological

traits in the grasshopper *Eyprepocnemis plorans*. *Hereditas*, *126*(2), 187-189.

https://doi.org/10.1111/j.1601-5223.1997.00187.x

Martis MM, Klemme S, Banaei-Moghaddam AM, Blattner et al (2012) Selfish supernumerary

chromosome reveals its origin as a mosaic of host genome and organellar sequences. Proc

Natl Acad Sci USA 109:13343-13346. https://doi.org/10.1073/pnas.1204237109

Morey JS, Ryan JC, Dolah FM (2006) Microarray validation: factors influencing correlation

between oligonucleotide microarrays and real-time PCR. Bio Proced Online 8:175.

Muñoz-Pajares AJ, Martínez-Rodríguez L, Teruel M, Cabrero J, Camacho JPM, Perfectti F

(2011) A single, recent origin of the accessory B chromosome of the grasshopper

Eyprepocnemis plorans. Genetics 187:853-863.

https://doi.org/10.1534/genetics.110.122713

Muñoz E, Perfectti F, Martín-Alganza A, Camacho JPM (1998) Parallel effects of a B

chromosome and a mite that decrease female fitness in the grasshopper *Eyprepocnemis

plorans*. Proc Roy Soc Lond B 265:1903-1909. https://doi.org/10.1098/rspb.1998.0518

Navarro-Domínguez B, Cabrero J, Camacho JPM, López-León M. D. (2016) B-chromosome

effects on Hsp70 gene expression does not occur at transcriptional level in the

grasshopper *Eyprepocnemis plorans*. Mol Genet Genomics 291:1909-1917.

Navarro-Domínguez B, Ruiz-Ruano FJ, Cabrero J, Corral JM, López-León MD, Sharbel TF,

Camacho JPM (2017a) Protein-coding genes in B chromosomes of the grasshopper

Eyprepocnemis plorans. Sci Rep-UK 7:45200. https://doi.org/10.1038/srep45200

Navarro-Domínguez B, Ruiz-Ruano FJ, Camacho JPM, Cabrero J, López-León MD (2017b)

Transcription of a B chromosome CAP-G pseudogene does not influence normal

condensin complex genes in a grasshopper. Sci Rep-UK 7:17650.

https://doi.org/10.1038/s41598-017-15894-5

Nishimura M, Tanaka S, Ihara F et al (2015) Transcriptome and histopathological changes in

mouse brain infected with *Neospora caninum*. Sci Rep-UK 5:7936.

https://doi.org/10.1038/srep07936

Nuismer SL, Otto SP (2005) Host–Parasite Interactions and the Evolution of Gene Expression.

PLoS Biol 3:e203. https://doi.org/10.1371/journal.pbio.0030203

Obbard DJ, Gordon KH, Buck AH, Jiggins FM (2009) The evolution of RNAi as a defence

against viruses and transposable elements. Philos T Roy Soc B 364:99-115.

https://doi.org/10.1098/rstb.2008.0168

Pappu R, Ramirez-Carrozzi V, Sambandam A (2011) The interleukin-17 cytokine family:

critical players in host defence and inflammatory diseases. Immunology 134:8–16.

https://doi.org/10.1111/j.1365-2567.2011.03465.x

Punta M, Coggill PC, Eberhardt RY et al (2011) The Pfam protein families database. Nucleic

Acids Res 40(Database issue): D290–D301. https://doi.org/10.1093/nar/gkr1065

Ramos E, Cardoso AL, Brown J et al (2017) The repetitive DNA element BncDNA, enriched in

the B chromosome of the cichlid fish *Astatotilapia latifasciata*, transcribes a potentially

noncoding RNA. Chromosoma 126:313–323. https://doi.org/10.1007/s00412-016-0601-x

Robertson G, Schein J, Chiu R et al (2010) *De novo* assembly and analysis of RNA-seq data.

Nat Methods 7:909-912. https://doi.org/10.1038/nmeth.1517

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential

expression analysis of digital gene expression data. Bioinformatics 26:139-140.

https://doi.org/10.1093/bioinformatics/btp616

Ruiz-Estévez M, Badisco L, Broeck JV, Perfectti F, López-León MD, Cabrero J, Camacho JPM

(2014) B chromosomes showing active ribosomal RNA genes contribute insignificant

amounts of rRNA in the grasshopper *Eyprepocnemis plorans*. Mol Genet Genomics

289:1209-1216. https://doi.org/10.1007/s00438-014-0880-y

Ruiz-Estévez M, López-León MD, Cabrero J, Camacho JPM (2012) B-chromosome ribosomal

DNA is functional in the grasshopper *Eyprepocnemis plorans*. PLoS One, 7:e36600.

https://doi.org/10.1371/journal.pone.0036600

Ruiz-Estévez M, López-León MD, Cabrero J, Camacho JPM (2013) Ribosomal DNA is active

in different B chromosome variants of the grasshopper *Eyprepocnemis plorans*. Genetica

141:337-345. https://doi.org/10.1007/s10709-013-9733-6.

Ruiz-Ruano FJ, Ruiz-Estévez M, Rodríguez-Pérez J, López-Pino JL, Cabrero J, Camacho JPM

(2011) DNA amount of X and B chromosomes in the grasshoppers *Eyprepocnemis*

plorans and *Locusta migratoria*. Cytogenet Genome Res 134:120-126.

https://doi.org/10.1159/000324690

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets.

Bioinformatics 27:863-864. https://doi.org/10.1093/bioinformatics/btr026

Schurch NJ, Schofield P, Gierliński M et al (2016). How many biological replicates are needed

in an RNA-seq experiment and which differential expression tool should you use? RNA

22:839-51. 10.1261/rna.053959.115

Smit AFA, Hubley R, Green P (1996-2010) RepeatMasker Open-3.0.

http://www.repeatmasker.org

Smyth GK (2005) limma: Linear Models for Microarray Data. In: Gentleman R., Carey V.J.,

Huber W., Irizarry R.A., Dudoit S. (eds) Bioinformatics and Computational Biology

Solutions Using R and Bioconductor (pp. 397-420). Statistics for Biology and Health. Springer, New York, NY. https://doi.org/10.1007/0-387-29362-0_23

Soumana IH, Tchicaya B, Simo G, Geiger A (2014) Comparative gene expression of *Wigglesworthia* inhabiting non-infected and *Trypanosoma brucei gambiense*-infected *Glossina palpalis gambiensis* flies. Front Microbiol 5:620. https://doi.org/10.3389/fmicb.2014.00620

Tatusov RL, Fedorova ND, Jackson JD et al (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41. https://doi.org/10.1186/1471-2105-4-41

Teruel M, Cabrero J, Perfectti F, Camacho JPM (2007) Nucleolus size variation during meiosis and NOR activity of a B chromosome in the grasshopper *Eyprepocnemis plorans*. Chromosome Res 15:755-765. https://doi.org/10.1007/s10577-007-1159-5

Teruel M, Cabrero J, Perfectti F, Camacho JPM (2010) B chromosome ancestry revealed by histone genes in the migratory locust. Chromosoma 119:217-225. https://doi.org/10.1007/s00412-009-0251-3

Teruel M, Cabrero J, Perfectti F, Alché JD, Camacho JPM (2009) Abnormal spermatid formation in the presence of the parasitic B_{24} chromosome in the grasshopper *Eyprepocnemis plorans*. Sex Dev 3:284-289. https://doi.org/10.1159/000253307

Teruel M, Sørensen JG, Loeschcke V, Cabrero J, Perfectti F, Camacho JPM (2011) Level of heat shock proteins decreases in individuals carrying B-chromosomes in the grasshopper *Eyprepocnemis plorans*. Cytogenet Genome Res 132:94-99. https://doi.org/10.1159/000319621

Trifonov VA, Dementyeva PV, Larkin DM et al (2013) Transcription of a protein-coding gene on B chromosomes of the Siberian roe deer (*Capreolus pygargus*). BMC Biol 11:90. https://doi.org/10.1186/1741-7007-11-90

UniProt Consortium (2014) Activities at the universal protein resource (UniProt). Nucleic Acids

Res 42 (Database issue):D191-D198. https://doi.org/10.1093/nar/gkt1140

Valente GT, Conte MA, Fantinatti BE et al (2014) Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. Mol Biol Evol 31:2061-2072. https://doi.org/10.1093/molbev/msu148

Valente GT, Nakajima RT, Fantinatti BEA et al (2017) B chromosomes: from cytogenetics to systems biology. Chromosoma 126:73–81. https://doi.org/10.1007/s00412-016-0613-6

Van Vugt JJ, de Nooijer S, Stouthamer R, de Jong H (2005) NOR activity and repeat sequences of the paternal sex ratio chromosome of the parasitoid wasp *Trichogramma kaykai*. Chromosoma 114:410-419. https://doi.org/10.1007/s00412-005-0026-4

Wang X, Fang X, Yang P et al (2014) The locust genome provides insight into swarm formation and long-distance flight. Nat Commun 5:2957. https://doi.org/10.1038/ncomms3957

Weinkove D, Neufeld TP, Twardzik T, Waterfield MD, Leevers SJ (1999) Regulation of imaginal disc cell size, cell number and organ size by *Drosophila* class I_A phosphoinositide 3-kinase and its adaptor. Curr Biol 9:1019-1029. https://doi.org/10.1016/S0960-9822(99)80450-3

Wu S, Zhu Z, Fu L, Niu B, Li W (2011) WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics 12:444. https://doi.org/10.1186/1471-2164-12-444

Yoshida K, Terai Y, Mizoiri S et al (2011) B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. PLoS Genet 7:e1002203. https://doi.org/10.1371/journal.pgen.1002203

Zhou Q, Zhu HM, Huang et al (2012) Deciphering neo-sex and B chromosome evolution by the

draft genome of *Drosophila albomicans*. BMC Genomics 13:109.

https://doi.org/10.1186/1471-2164-13-109

Zurita S, Cabrero J, López-León MD, Camacho JPM (1998) Polymorphism regeneration for a

neutralized selfish B chromosome. Evolution 52:274-277. https://doi.org/10.1111/j.1558-

5646.1998.tb05163.x

Figure Legends

Fig 1 Methods workflow for the present research

Fig. 2 Overview of *E. plorans* transcriptome annotation. (A) Results of BLAST searches in UniProtKB/Swiss-Prot and UniProt Reference Clusters (UniRef) databases. (B) Number of Trinity unigenes annotated by RepeatMasker as different families of Transposable Elements in the *E. plorans* transcriptome

Fig. 3 Differential expression analysis of B-carrying and B-lacking samples. Volcano plots (Fold Change vs. Statistical Significance) for transcriptome (A) and microarray (B) data. MA plots (Fold Change vs. Mean Expression) for transcriptome (C) and microarray (D) data in *E. plorans*. Statistical significance value of differential expression obtained in the microarray (x-axis) and in the transcriptome (y-axis) analysis (E). Fold change of differential expression obtained in the microarray (x-axis) and in the transcriptome (y-axis) analysis (F). Coloured dots represent the values of the sequences that were differentially expressed (p<0.05) in coincident expression patterns in both analysis. Up-regulations are labelled in green and down-regulations in orange. Note the higher number of green dots, indicating that B chromosome presence causes a general increase in gene expression

Fig. 4 Number of differentially expressed unigenes, in the host genome, within each KOG class, defined as: A: RNA processing and modification, B: Chromatin Structure and dynamics, J: Translation, ribosomal structure and biogenesis, K: Transcription, M: Cell wall/membrane/envelope biogenesis, O: Post-translational modification, protein turnover, chaperone functions, T: Signal Transduction Mechanisms, U : Intracellular trafficking and

secretion, W: Extracellular structures, Z: Cytoskeleton, C: Energy production and conversion, G: Carbohydrate metabolism and transport, I: Lipid transport and metabolism, P: Inorganic ion transport and metabolism, Q: Secondary Structure, R: General Functional Prediction only, S: Function Unknown. (See also Online Resource 3)

Fig. 5 qPCR validation results for 8 selected DEGs. The X-axis represents the $\log_2FC$ obtained by qPCR and the Y-axis represents the $\log_2FC$ values derived from the microarray analysis for each gene (labelled dots). Note that the $\log_2FC$ shows the same trend in 7 out of the 8 genes, meaning that they show concordant expression pattern in both techniques. The blue line marks the regression line and the shadowed area delimits the 95% confidence interval.
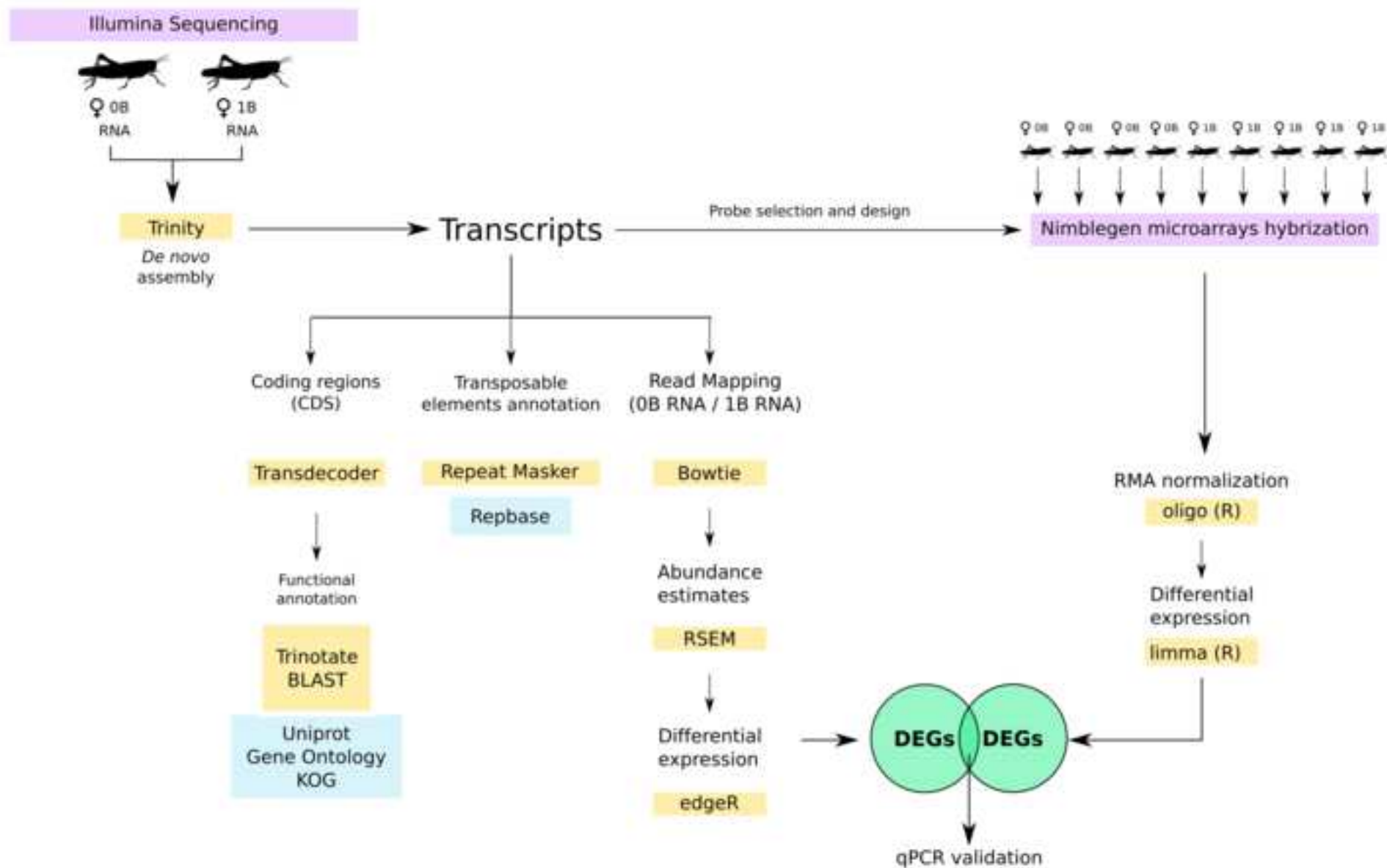
Fig. 6 Graphical summary of gene expression changes observed in B-carrying ovaries

Table 1

Table 1: Summarized statistics from Trinity assembly before and after removing potential contaminations with DeconSeq

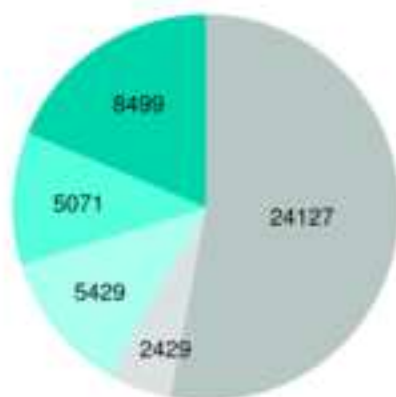| Item	Type	Before DeconSeq	After DeconSeq
Number of sequences	Transcripts	74378	73889
	Genes (Isoform clusters)	45633	45555
N50	Transcripts	2330	2302
	Longest isoform per gene	1703	1702
Max. length	Transcripts	21010	19141
	Longest isoform per gene	21010	19141
Min. length	Transcripts	201	201
	Longest isoform per gene	201	201
Average length	Transcripts	1325.27	1314.8
	Longest isoform per gene	914.75	914.65
Median length	Transcripts	793	788
	Longest isoform per gene	485	485
Total assembled bases	Transcripts	98571226	97149070
	Longest isoform per gene	41742580	41666905
Percent GC		40.37	40.24

Table 2

Table 2. Summary of annotation for the differentially expressed
unigenes. TE= Transposable element

Item	Annotated CDS	Non-annotated	TEs	Total
Up-regulated	35	86	40	161
Down-regulated	16	9	2	27
Total	51	95	42	188

Figure 1

Figure 2

Figure 3   Click here to access/download;Figure;Fig.3.tiff ⬇

Figure 4　　Click here to access/download;Figure;Fig.4.tiff ⬇

Figure 5

$R^2=0.6732$
$p=0.0151$

Figure 6

Click here to access/download;Figure;Fig.6.tif ±

Nucleotypic effects

Host defense mechanisms

Fertility and ovarian function

Regulation of gene expression