The B chromosome of *Pseudococcus viburni*: a selfish chromosome that exploits whole-genome meiotic drive

Isabelle M. Veal^{1*}, Andrés G. de la Folia^{2*}, Kamil S. Jaron², Andrew J. Mongue², Francisco J. Ruiz-Ruano^{3,4},

Scott E.J Barlow², Ross Nelson², & Laura Ross²⁺

¹ *Department of Biological Sciences, University of Illinois at Chicago, USA*

² *School of Biological Sciences, Institute of Evolutionary Biology, The University of Edinburgh, Edinburgh, EH9 3FL, UK.*

³ *School of Biological Sciences, University of East Anglia, NR4 7TJ, UK*

⁴ *Department of Organismal Biology, Uppsala University, Sweden*

* These authors contributed equally

+ Correspondence contact: laura.ross@ed.ac.uk

Keywords: B chromosomes, paternal genome elimination, genomic conflict, mealybugs

Abstract

Meiosis, the key process underlying sexual reproduction, is generally a fair process: each chromosome has a 50% chance of being included into each gamete. However in some organisms meiosis has become highly aberrant with some chromosomes having a higher chance of making it into gametes than others. Yet why and how such systems evolve remains unclear. Here we study the unusual reproductive genetics of mealybugs, in which only maternal-origin chromosomes are included into the gametes during male meiosis, while paternally-derived chromosomes degrade. This “whole genome meiotic drive” occurs in all males and is evolutionarily conserved. However one species - the obscure mealybug *Pseudococcus viburni* - has a segregating B chromosome that increases in frequency by escaping paternal genome elimination. Here we present whole-genome and gene expression data from laboratory lines with and without B chromosomes. These data allow us to identify B-linked sequences including >70 protein-coding genes as well as a B-specific satellite repeat that makes up a significant proportion of the chromosome. We also used these data to investigate the evolutionary origin of the B chromosome. The few paralogs between the B and the core genome are distributed throughout the genome, showing that it is unlikely that the B originated through a simple duplication of one of the autosomes. We also find that while many of the B-linked genes do not have paralogs within the *P.viburni* genome, but they do show orthology with genes in other hemipteran insects suggesting that the B might have originated from fission of one of the autosomes, possibly followed by further translocations of individual genes. Finally in order to understand the mechanisms by which the B is able to escape elimination when paternally-derived we generated gene expression data for males and females with and without B chromosomes. We find that at the developmental stage when meiosis is taking place only a small number of B-linked genes show significant expression. Only one gene was significantly over-expressed during male meiosis, which is when the drive occurs: an acetyltransferase involved in H3K56Ac, which has a putative role in meiosis and is therefore a promising candidate for further studies. Together, these results form a promising foundation for studying the mechanisms of meiotic drive in a system that is uniquely suited for this approach.

Introduction

Multicellular organisms can be regarded as nested hierarchies of cooperating entities: genes within chromosomes, chromosomes within cells, and cells within individuals. But under the surface there is a hidden world of evolutionary conflict. The integrity of higher levels is challenged by lower-level entities trying to improve their representation in future generations (Hurst 1992; Szathmáry and Smith 1995; Burt and Trivers 2006). Intragenomic conflict – which arises when different entities within a genome have antagonistic evolutionary agendas (Gardner and Úbeda 2017) – is especially pronounced during reproduction, since only limited numbers of genes are transmitted to the next generation. Meiotic drive – whereby a selfish genetic element hijacks gametogenesis to increase its transmission to future generations – is a key mechanism for intragenomic conflict (Lyttle 1993; Lindholm et al. 2016; Lenormand et al. 2016). Normally, during meiosis, each gene has a 50% chance of being passed on to the gametes. Meiotic drivers however can bias this fair gamble, increasing their transmission to the offspring at the expense of other genes. Which part of the genome “drives” can vary from just a single gene, to individual chromosomes, and even in some cases entire haploid genomes (reviewed in Burt and Trivers 2006; Gardner and Ross 2014). However, despite an abundance of drive systems across the tree of life, we know very little about their origin, evolution, or the mechanisms that these drivers use to increase their transmission.

One of the most widespread classes of meiotic drivers are supernumerary B chromosomes. They are found across fungi, plants and animals and estimated to be present in at least 10% of eukaryotic species (Camacho 2005; Ahmad and Martins 2019). B chromosomes are genomic parasites that exist within cells of some individuals of a species, sometimes in multiple copies, in addition to autosomes and sex chromosomes – the A chromosomes (Camacho et al. 2000; Werren 2011; Dalla Benetta et al. 2019a); B chromosomes are non-essential for the host. Yet they are maintained in populations, often in spite of neutral or even deleterious fitness effects, through meiotic drive mechanisms (Jones 1991; Houben 2017). These mechanisms are diverse and with few exceptions remain poorly understood. Many Bs appear to exploit existing asymmetries in meiosis: for example, securing their inclusion into oocyte rather than polar bodies during oogenesis in animals (e.g. in a locust, Pardo et al., 1994) or into the generative nucleus rather than the tube nucleus during pollen formation in plants (e.g. in rye, Banaei-Moghaddam et al., 2012). In other cases, the drive mechanism exploits asymmetries in transmission of chromosomes from parents to offspring:

In the haplodiploid wasp *Nasonia vitripennis*, haploid males normally do not pass genetic material on to male offspring. However a sperm-derived B chromosome in this species drives by eliminating all other paternally-inherited autosomes from fertilised female embryos during embryogenesis. This turns these embryos into haploidised males which will incorporate the B chromosome into their sperm (Reed and Werren 1995; Swim et al. 2012; Dalla Benetta et al. 2019b).

However, probably the most extreme example of B chromosomes taking advantage of asymmetries in reproduction is found in a small plant feeding hemipteran insect, the obscure mealybug *Pseudococcus viburni* (Hemiptera: Pseudococcidae). Here the B chromosome hijacks their peculiar reproductive system that involves non-Mendelian chromosome transmission to increase its own transmission: Mealybugs reproduce via paternal genome elimination (PGE), in which maternally-inherited chromosomes drive at the expense of paternally-inherited chromosomes during male meiosis, which are eliminated from sperm (Bongiorni et al., 2004; Brown and Nur, 1964; Dallai, 2012; Nur, 1990). The B chromosome in *P. viburni* drives by avoiding elimination when paternally-derived and is included in all sperm (Nur 1962; Nur 1966; Nur 1966; Nur 1969). So in this species B chromosomes increase their transmission by avoiding meiotic drive (PGE constitutes whole-genome post-meiotic drive), rather than inducing it themselves. As a result, studying the drive mechanisms of this B chromosome not only provides potential insights into the way a single-chromosome reproductive parasite can increase its transmission, but might also help unravel the mechanisms that govern paternal genome elimination, a type of reproduction that is poorly understood despite its widespread distribution and repeated evolution (seven origins and 10,000s species) across Arthropods (Gardner and Ross 2014; de la Filia et al. 2015).

Cytogenetic studies have provided a detailed description of the behaviour of the B chromosome in *P.viburni* (summarized in Figure 1), while those in a number of other mealybug species give further insights into the behaviour of maternal and paternal chromosomes. In males of all mealybugs without B chromosomes, under PGE all paternally-inherited chromosomes are eliminated during spermatogenesis (Nur 1980 and citations therein). In addition, in most male tissues paternal chromosomes are heavily condensed through heterochromatinization and are transcriptionally suppressed (Brown and Nur 1964; Bongiorni et al. 2001; de la Filia et al. 2020, Figure 1). During male meiosis, chromosomes segregate based on their parental origin during anaphase II, which involves a monopolar spindle that only interacts with the euchromatic maternal set while the

heterochromatic paternal complement lags behind (Bongiorni et al. 2004). Only spermatids containing maternal chromosomes mature into spermatozoa, while those containing heterochromatic paternal chromosomes degrade (Hughes-Schrader 1948; Nur 1962; Bongiorni et al. 2004; Bongiorni et al. 2009). B chromosomes found in populations of *P. viburni* appear to exploit this unusual inheritance pattern as an accumulation mechanism. Regardless of its parental origin, the B shows the same heterochromatic state as the paternal chromosomes in somatic tissues of males (Figure 1&2). However, the B becomes fully decondensed at the beginning of male meiosis, thus mimicking the maternal chromatin state, and migrates with the maternal chromosomes even when inherited from the father (Nur, 1962 and Figure 1&2). By doing so it avoids elimination by always securing inclusion into the two maturing spermatocytes (instead of only when maternally-derived), rather than into the other two degrading nuclei. Via this accumulation mechanism, the B can reach a transmission rate as high as >0.9 through males in spite of clear fitness consequences, such as delayed developmental time and sperm depletion (Nur, 1969; Nur, 1966; Nur, 1966). The B chromosome is also transmitted through females although it does not show any drive, in fact transmission through female meiosis is slightly <0.5 (Nur, 1962, Figure 1). The ability of the paternal B chromosome to increase its transmission in males also varies among different “host” genotypes and studies have shown that it is the maternal rather than the paternal genome that determines this variability in transmission rate (Nur and Brett 1985; Nur and Brett 1988).

Figure 1. A schematic depiction of the behaviour of the B chromosome during male and female meiosis and gametogenesis in *Pseudococcus viburni*. The top panel shows female meiosis, the bottom two panels show male meiosis with two scenarios where the B is either paternally-derived (top) or maternally-derived (bottom). Throughout the chromosome sequence is shown for $N=2$ autosomes (maternally-derived autosomes in pink, paternally-derived autosomes in blue) and one unpaired B chromosome (in yellow). Please note that mealybugs have an inverted meiotic sequence and have holocentric chromosomes without a localized centromere. During interphase, prophase each chromosome consists of two loosely connected chromatids that segregate at meiosis I into two daughter cells. For simplicity only one of these daughter cells is depicted in meiosis II and as a result only two rather than four gametes are shown. In male meiosis paternally-derived chromosomes form a condensed heterogametic body in interphase, but separate at prophase and during the two meiotic divisions, however throughout the entire sequence paternal chromosomes appear more condensed than maternal chromosomes. Maternally-derived B chromosomes behave like the other maternally-derived chromosomes and are always included into the maturing sperm nuclei. Paternally-derived Bs are less condensed than the paternal A chromosomes during meiosis and tend to segregate with the less condensed maternal chromosomes into the maturing sperm nuclei, rather than with the paternally-derived chromosomes which segregate into the degrading nuclei.

Based on these cytogenetic studies we hypothesize that the B chromosomes in *P.viburni* escape elimination by changing their chromatin structure thereby enhancing their

transmission. However up until now we do not have any genetic data that might help us test this hypothesis and help to better understand the molecular underpinnings of the Bs drive. Here we redress this lack of data by sequencing and comparing genomes of individuals with and without B chromosomes. We combine these data with a large gene expression study comparing the expression of B-linked genes during male and female meiosis and gametogenesis. We expect genes involved in drive to be upregulated in males as the B does not show drive in females. The few similar studies in other species with B chromosomes to date show that most B chromosomes do contain actively transcribed genes, but are often gene-poor relative to autosomes and highly enriched for repetitive sequences (as reviewed by Ahmad and Martins 2019). In a few cases, these studies were actually able to identify B-linked genes responsible for the drive, such as the gene *haplodizer* in the wasp *N. vitripennis* (Dalla Benetta et al. 2020). We present a highly contiguous *de novo* draft genome for an inbred *P. viburni* line which contains a single B chromosome. By comparing this reference genome to genome data from lines without Bs and with 2 B copies we were able to identify and assemble 1.37Mb of B-linked sequence containing 78 protein coding genes. We found that while most of these genes showed low levels of expression, we identified a number of promising B-linked candidate genes which have annotated functions in meiosis and histone modification and are expressed during gametogenesis. Most notable is the histone acetyltransferase Rtt109/CBP which has a putative role in meiosis and is upregulated in males compared to females. Finally, we investigate the relationship of the B chromosome to the rest of the core genome and find that autosomal paralogs of B-linked genes are equally distributed across the genome. This suggests that the B is not derived from one particular autosome.

Results

B chromosomes in P. viburni populations

We collected *P. viburni* from glasshouses at the Royal Botanical Gardens in Edinburgh, Scotland and established isofemale lines. To identify B- and B+ *P. viburni* lines for DNA and RNA-seq (Supplementary Table S1), we stained and counted chromosomes in embryos. We found that one of our *P. viburni* populations, RBGE25, did not carry a B chromosome (Fig. 1A(i)) and established two B- experimental lines: PV21 and PV23. We then selected two other populations that carried either one B, RBGE31 (Fig. 1A(ii); from which we established an experimental line, PV13) or on average two Bs, RBGE29 (Fig. 1A(iii), experimental line PV04).

B chromosome behaviour during meiosis

We used cytological preparations of second-instar male dissected testes to observe how the B chromosome behaves during meiosis and spermiogenesis and compare this to the earlier studies which were done in a different population. At the beginning of meiosis, the maternal genome (MG) starts condensing to form metaphase chromosomes, but since the paternal genome (PG) is already in the heterochromatin state, the paternal chromosomes start to individualize before distinct maternal chromosomes become apparent (Fig. 1B(i)). In B- populations, we observed $n=5$ paternal chromosomes. In B+ populations, during the same early meiosis stage we could observe $n=5+1/2Bs$ paternal chromosome in the heterochromatic side of the nucleus (Fig. 1B(iii) carrying 2Bs). Male meiosis in mealybugs results in a quadrinucleate cell (four nuclei within a shared cytoplasm). This cell contains two opposite nuclei containing the heterochromatic PG and two opposite nuclei with the euchromatic MG. In B- populations, we could clearly see the $n=5$ PG chromosomes (Fig. 1B(ii)), while in the B+ populations, the B chromosome(s) either segregates with the MG or the PG (Fig. 1B(iv, v)). However in both 1B and 2B populations we saw that generally the nuclei containing the MG carried 1-2 Bs at the end of meiosis, observed as condensed dots within the nuclei (not shown, but see 1B(v)), while segregation with the PG genome was rare. After meiosis the quadrinucleate cell elongates, resulting in the elongation of the MG nuclei only (Fig. 1B (vi and vii)). The PG remains condensed and is eventually eliminated. So it appears that even paternally-derived B chromosomes that colocalize with the PG at the beginning of meiosis segregate with the MG by the beginning of spermiogenesis, when the PG is eliminated. As such, B chromosomes escape PGE despite their paternal origin.

Figure 2. Cytological preparations of screened *Pseudococcus viburni* populations used for sequencing. (A) Embryonic cells with (i) only A chromosomes $2n=10$, from PV23. (ii) A chromosomes $2n=10 + 1B$ chromosome, from PV13, (iii) A chromosomes $2n=10 + 2B$ chromosomes, from PV04. (B) Testes cells of second-instar males with cells at the beginning of meiosis, where the paternal chromosomes are showing either $n=5$ from PV23 (i) or $n=5+2B$ from PV04 (ii); and cells at the end of meiosis, where the future gametes do not carry B in 0B lines (PV23) form (iii) or where the B chromosome either segregates with the PG (iv) or the MG (v) from PV13. During spermiogenesis, the PG becomes heterochromatinised again, while the MG decondenses and the nucleus starts elongation to form the future gametes (vi and vii), from PV13. All images are at the same scale (x40 or x20), and only show DNA, cells stained with DAPI

Genome assembly

We generated a *P. viburni* genome assembly from pooled females of the B-carrying experimental line PV13. We assembled 28.6Gb of PacBio subreads and used two Illumina short read libraries generated from three females from the same line to polish the assembly (see Materials and Methods and Supplementary Table S1 for a summary of our sequencing efforts). Our genome assembly, after removing contigs derived from the bacterial endosymbionts housed by the species (International Aphid Genomics Consortium 2010;

Chen et al. 2016), comprises 435.6Mb in 2,392 scaffolds with a GC content of 33.7% and a N50 of ~0.87Mb. These and other assembly metrics are shown in Table 1.

Table 1. Genome assembly and annotation statistics

Feature	Statistic
Genome span (Mb)	435.36
GC fraction	0.337
Number of scaffolds	2,392
N50	874,986
Num N50	152
Longest scaffold	4,205,925
Average length of scaffolds	182,007
BUSCO Insecta	95.2% (S: 91.7%, D: 3.5%)
BUSCO Hemiptera	92.6% (S:89.3%,D:3.3%)
Protein coding genes	23,629
mRNAs	25,677
Repeats (%)	49.7%

Our genome assembly shows high completeness, as revealed by the presence of >90% single copy orthologs from the insect and hemipteran BUSCO gene sets. Half of the genome (216.2Mb) consists of repetitive sequences, higher than the 38-45% range estimated in *Phenacoccus solenopsis* (Li et al., 2020) and *Pseudococcus longispinus* (Garber et al. 2021 publicly available at https://ensembl.mealybug.org/Pseudococcus_longispinus_v1) and hemipterans such as *Acyrtosiphon pisum* or *Bemisia tabaci* (International Aphid Genomics Consortium, 2010; Chen et al., 2016). The vast majority of repeats (96.6%) are unclassified. We predicted a set of 23,629 protein coding genes, which is twice the number found in the most complete mealybug genome available to date, *P. solenopsis*, but very similar to the gene estimate in the more closely related *P. longispinus*. We annotated 13,515 genes using homology searches and functional domain prediction. The BUSCO scores for our annotated gene set were 95.1% (Insecta) and 93.1% (Hemiptera), which were extremely close to our assembly

benchmarking and therefore indicative of a successful gene annotation. Moreover, less than 5% of BUSCOs were duplicated, suggesting our higher gene count is probably not the result of systematic assembly errors.

Identifying B-derived scaffolds in the reference genome

We reasoned that B-derived Illumina reads might either be unique to B chromosomes or shared between A and B chromosomes, as is common in supernumerary chromosomes (Ahmad and Martins 2019). Therefore, we combined several lines of evidence to assign scaffolds to autosomal (A) and B chromosomes using the short read libraries generated for the two B+ (PV13 and PV04) and B- lines (PV21 and PV23) (see Materials and Methods for full details).

First, we mapped the short reads to our reference assembly and identified scaffolds with higher coverage in B+ lines compared to B- lines using two coverage thresholds: a strict threshold of $\log_2(\text{coverage ratio}) \geq 2$ (145 scaffolds) and a more relaxed threshold of $\log_2 \geq 0.58$, corresponding to a 3:2 ratio that could be expected for scaffolds shared between autosomes and a single copy of the B chromosome (a further 136 scaffolds) (Fig. 2A, Supplementary Figure S1). Second, we assembled the short reads for each library into reference-independent assemblies and aligned them to our reference assembly to identify scaffolds in the reference genome present in the PV13 and PV04 Illumina assemblies but not in PV21 or PV23 (107 scaffolds) (Fig. 2B, Supplementary Table S2). Third, we used a kmer-based approach to identify overrepresented kmers in the B+ libraries using the 3:2 rule. We recovered 553,816,347 27mers with at least 3 counts in all libraries, of which 49.78M (~9%) were assigned as putative B kmers. We mapped both sets of B and non-B kmers to the PacBio assembly and identified 58 scaffolds with at least as many putative B as non-B kmer mappings (Fig. 2C). We further reasoned that B-linked sequences should be present in higher coverage in PV04 (the 2B line) than in PV13 (the 1B line), although we did not incorporate this as a formalised rule in our decision process since B number can vary within lines and we did not directly karyotype the sequenced individuals.

Combining these different criteria, we were able to assign B status to reference scaffolds with varying degrees of confidence (B1 = high confidence, B2 and B3 = lower confidence), with the highest confidence set, B1, comprising 103 scaffolds and 1.37Mb of reference sequence (Fig. 2D, Table 2, Supplementary File S1.0). Approximately half of the

B1 scaffolds are larger than 10Kb, with the largest scaffold spanning 190Kb (Supplementary Figure S2).

We used PCR to validate these results by targeting the largest B1 scaffold, scaffold_552. We designed three sets of primers (Supplementary Table S3) and genotyped females from the four experimental lines and another mealybug species, *Planococcus citri*, as a negative control. As expected, only PV13 and PV04 individuals showed amplification (Supplementary Figure S7).

Table 2. Assignments of A and B scaffolds in the *P. viburni* reference assembly.

Set*	N scaffolds	Mb	Mean PV04/PV13 cov ratio	Predicted genes (of which annotated)
A	2084	428.2	1.01 (0.64 SD)	23143 (13309, 58%)
B set 1	103	1.37	2.99 (0.84 SD)	78 (32, 41%)
B set 2	106	3.16	1.81 (0.90 SD)	297 (112, 37%)
B set 3	99	2.58	0.51 (0.28 SD)	111 (62, 56%)

* Assignment strategy as follows: **B set 1 (highest confidence)**, $\log_2$ read coverage ratio between all B+/B- pairs ≥ 2 and coverage in PV04 > coverage in PV13 and 1-1 mapping to a contig in the B+ Illumina assemblies only or $\log_2$ putative A/B kmer ratio ≥ 0 . **B set 2**, cov in PV04 > cov in PV13 and at least one of the following: $\log_2$ cov ratio between all B+/B- pairs ≥ 0.58 , 1-1 mapping in B+ assemblies only, $\log_2$ A/B kmer ratio ≥ 0 . **B set 3 (lowest confidence)**, at least one of the following: $\log_2$ cov ratio between all B+/B- pairs ≥ 0.58 , 1-1 mapping in B+ assemblies only, $\log_2$ A/B kmer ratio ≥ 0 .

Figure 3. B-linked scaffold assignment. (A) Assignment based on read coverage differences between B+ and B- lines. (B) Assignment based on mappings to individual Illumina assemblies of the individual lines. Counts of scaffolds with 1-to-1 mappings are shown on the X axis. (C) Assignment based on alignment of kmers overrepresented in B+ lines. (D) Scatterplot of scaffold read coverage in the two B+ lines, PV04 (2B) and PV13 (1B). B-assigned scaffolds are represented in shades of blue according to assignment confidence. A chromosomes are shown in light gray.

Composition of B-linked sequences

B chromosomes are typically rich in repetitive sequences. Accordingly, we found a higher repetitive component in the highest confidence B1 scaffolds, 71.9%, compared to the whole genome (see S1.15 and 1.16 in the supplement file), with the vast majority of these repeats unclassified. However it is likely that this substantially underestimates the true repeat content of the B as the assembled scaffolds might not capture highly repetitive regions. Therefore we also analysed the repeat content of the raw reads from the B- and B+ libraries, following a pipeline in Ruiz-Ruano et al. (2016). Across all libraries we identified 19

satellite repeats (Figure 4, Repeat Analysis Supplement), four of which (with repeat-length ranging from 30-414) show a strong over-abundance in B+ libraries. One 172bp B-linked repeat is particularly abundant, comprising 3.9% of all reads in the 2B library and 1.3% in the 1B library. Divergence patterns between copies of this repeat suggest that much of its abundance is due to recent extensive amplification (Figure 4).

Figure 4. Satellite repeats. Repeat landscapes for the assembly and each separate library (average across two 0Bs, 1B and 2B). Each plot shows the abundance for each of the 19 satellite repeats at different divergence intervals (Kimura substitution levels). The satellites are numbered in decreasing order of 0B abundance and with the monomer size at the end of the name. The 1B and 2B libraries are enriched for the PviSat05-1280 and in particular the PviSat16-172 satellites. PviSat16-172 shows very low divergence between repeats indicating recent amplification.

We identified 78 protein-coding genes (81 transcripts) located on 26 of the 103 B1 scaffolds (Supplementary file S1.1). Most of the genes are located on the largest scaffold (scaffold_552). Additionally, 408 genes are located on the lower-confidence B2 and B3 scaffolds, of which 174 are annotated (Supplementary file S1.3). We tested whether the B chromosome shows a clear homology to the rest of the genome using reciprocal blast of all the B-linked scaffolds against the core genes annotated in the assembly, and furthermore against all annotated genes of closely related species *P. solenopsis* (Li et al. 2020) and *P. longispinus* (Garber et al. 2021). Of all 486 B-linked genes we identified 56 that carry a paralog in the core genome (Supplementary Table S6). Furthermore for 43 and 11 B-linked genes we identified orthologs in *P. longispinus* and *P. solenopsis* respectively (Supplementary Table S6). In the case of *P. solenopsis* we were able to identify the orthologs located on all 5 chromosomes. The low numbers of identified paralogs and orthologs, together with scattered hits across the *P. solenopsis* suggest the B chromosome is not

derived from a core genome chromosome. Additionally, we found no support that the B chromosome is homologous to a *Phenococcus* core chromosome using nucleotide-alignment to *P. solenopsis* genome (Li et al. 2020). We found none of the scaffolds in the most stringent putative B chromosome set (B1) showed conserved homology with any of the five chromosomes. Even for relaxed criteria (B2, B3) there was almost no detectable synteny (Supplementary Table S6). This is unlikely a result of methodological approach as we were able to find homologous chromosomes of 80% of the *P. viburni* core genome scaffolds (see Supplementary Table S6). To test for a more recent origin, we also compared *P. viburni* to the more closely related congener, *Pseudococcus longispinus*. More putative B linked scaffolds have syntenic alignments with *P. longispinus* than with *P. solenopsis*, but even between the congeners there is little evidence of long tracts of conserved sequence between the *P. viburni* B and the A chromosomes of other mealybug species (Supplementary Table S7, Figure S5).

Furthermore, for 17 B-linked genes we identified a homolog in at least two out of three core genomes. In all 17 cases these genes showed patterns consistent with intraspecific origins. Divergences in all cases followed phylogenetic relationships of the three species - the highest similarity to *P. viburni* core genome, followed by similarity to *P. longispinus* and lowest similarity to *P. solenopsis*. Although this pattern of nucleotide similarity is consistent with intraspecific origin, we found a homologous sequence only for a relatively small fraction of B-linked genes - 93 out of a total 486 predicted genes (those in the B1, B2 and B3 criteria). To further investigate possible other sources we searched for homologies against NR, NCBI, and UniProt databases using the same strategy used by blobtools2 (<https://github.com/blobtoolkit/blobtools2>; Laetsch and Blaxter, 2017). We found a significant hit for additional 104 out of 394 genes of unknown origin. Majority of homologs were found in Arthropods, and most of them specifically in Hemiptera (Supplementary Table S5). One notable exception is scaffold_360, which carries a set of 18 viral genes as well as six arthropod genes and two paralogs to the core genome.

Putative functions of B-linked genes

In our genome annotation, thirty-two B1 genes appear annotated either with protein family domains and/or significant homology to protein databases. Among the functions associated with these genes, we found processes such as microtubule motor activity, ion transport, transmembrane transport, histone modifications, transcription regulation, DNA

integration and transposition activity (Supplementary file S1.1). Additionally, 408 genes are located on the lower-confidence B2 and B3 scaffolds, of which 174 are annotated (Supplementary file S1.3).

Our RNA-seq data of males and females from the four experimental lines revealed, however, that most B1 genes are not expressed in second instars. Only 13 B1 genes are expressed in at least one of the sample groups (TPM>1), of which three have expression levels about 10 TPM (Supplementary file S1.4). When averaging across all B+ and B- samples, we found that B1 genes show significantly lower expression than B2 and B3 genes in both cases (Supplementary Figure S3).

Sex-biased gene expression

Before exploring whether the presence of B chromosomes can differentially affect males and females, particularly during meiosis, we compared the gene expression profiles of our male and female RNA-seq samples in order to identify sex-biased expression and confirm we captured the right developmental stage. The 2nd instar stage is critical as it corresponds to the onset of reproductive tissue maturation, but also to the establishment of diverging developmental pathways: at later larval stages and adulthood, mealybugs exhibit striking sexual dimorphism (Vea et al. 2016; Bain et al. 2021)). Therefore, our RNA-seq data gives us the opportunity to directly examine sex-specific differences in gene expression at late 2nd instar stage (Fig. 3A).

Thus, we conducted a differential expression analysis between all male and female RNA-seq samples in our study. Out of a total of 15,673 genes with enough expression in at least one sex to be included in the analysis (Materials and methods), we identified 3,830 genes (24.4%) with sex-biased expression ($FDR < 0.05$, $|\log FC| > 1$) (Fig. 3B, Supplementary file S1.5). Of these, 2,186 are overexpressed in females and 1,644 in males, of which 57 and 484, respectively, can be considered strongly sex-biased ($\log FC > 5$). We conducted a GO enrichment analysis of sex-biased genes in both sexes ($p < 0.01$, Fig. 3C) and found that female-biased genes are enriched in metabolic processes such as lipid and carbohydrate metabolism (Supplementary file S1.6), while male-biased genes are enriched in protein modification, membrane organisation, lipid translocation and microtubule organisation (Supplementary file S1.7).

Among the male-biased genes, a few are involved in endocrine pathways linked to insect metamorphosis. Particularly, g2298 codes for JHAMT protein, the enzyme involved in the last step of juvenile hormone (JH) synthesis. This transcript was previously found to have

diverging expression profiles between males and females at the onset of metamorphosis in mealybugs (Vea et al. 2016). However, we only found a single strongly male-biased gene that is directly related to reproduction: g17352 (protamine-like sperm chromatin condensation).

Fig 5. Sex-biased expression in 2nd instars of *P. viburni*. (A) Histogram of sex-specificity (SPM, a proportion of expression ranging from 0 (in males only) to 1 (in females only) (Xiao et al. 2010; Kryuchkova-Mostacci and Robinson-Rechavi 2017). (B) Scatterplot of expression fold-changes in genes and average levels of expression. Significantly differentially expressed genes are shown in red (female-biased) and blue (male-biased). Strongly sex-biased genes ($|FC| > 5$) are shown in darker shades. (C) Significantly enriched GO terms related to biological processes in females (red circles) and males (blue squares). 803 female-biased genes and 653 male-biased genes had GO terms associated with them. The background population consisted of 6,968 genes included in the differential expression analysis with GO terms.

Differential gene expression between B-carrying and non B samples

We then examined differential expression between B-carrying and B- sample groups with a focus on B+ males, due to the distinct behaviour and meiotic drive of the B chromosome in males. To do so, we performed a differential expression analysis for 18,066 genes which included four contrasts: B-carrying males versus the other three sample groups (B- males, B+ females and B- females) and, additionally, B+ females versus B- females (18,066 genes with significantly differential expression set to $|\log_2FC| > 0.58$ and FDR < 0.01; Table 3; Supplementary Figure S4; Supplementary file S1.8).

We found 432 differentially expressed genes between B+ males and B- males, 232 of which are annotated (Fig 4A; Supplementary file S1.9). Among the 286 genes that are overexpressed in B+ males, we found seven located on B1 scaffolds. These genes include five annotated genes: a histone acetyltransferase Rtt109/CBP (g13953), two fatty acid synthases (g1208, g5582) and two contiguous genes with predicted AAA-ATPase-like domains (g9061, g9062). As expected, we did not find any B1-linked genes among the 146 genes that are overexpressed in B- samples. A GO term enrichment analysis revealed that the differentially expressed genes between B+ and B- males are enriched in four biological processes: DNA integration (GO:0015074), transposition, DNA-mediated (GO:0006313), proteasome assembly (GO:0043248) and nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184) (122 genes with associated GO terms out of 7,375 annotated genes in the background population, Supplementary file S1.10).

The results in females were similar, with 233 genes overexpressed in B+ (7 on B1 scaffolds, including g1208, g5582, g9061 and g9062) and 171 in B- (none on B1) (Fig 4B; Supplementary file S1.11). The differentially expressed genes are enriched in DNA metabolic process (GO:0006259) and telomere maintenance (GO:0000723) (119 genes with associated GO terms out of 7,375 annotated genes in the background population, Supplementary file S1.12). Overall, we found substantial overlap in the annotated differentially expressed genes between B+ males and B- males and B+ females and B- females: 94 genes are upregulated in B+ males and females (Supplementary file S1.13), while 34 genes are downregulated compared to B- males and females (Supplementary file S1.14). We did not find any annotated genes with opposite differential expression patterns in males and females: i.e. upregulated in B+ males and downregulated in B- females, or vice versa.

We then focused on the genes that are differentially expressed in B+ males compared to the other three sample groups (Figs. 4C-D, Table 4). Among the 88 genes that are overexpressed in B+ males, we found a single B1 gene: the histone acetyltransferase g13953. Additionally, we found two annotated genes located on low-confidence B-linked scaffolds: g5312 (B2), a transposase IS4; and g5232 (B3), which contains a N-terminal domain of TAR DNA-Binding Protein 43. The remaining annotated genes are located on A scaffolds and are therefore autosomal. Nine genes are underexpressed in B+ males, including five genes on A scaffolds and a catalase gene, g5248, located on a B3 scaffold.

In conclusion, the presence of B chromosomes triggers similar transcriptional changes in both males and females, both originating from the B chromosome itself and

autosomal genes. Interestingly, the single B gene that is uniquely overexpressed in B+ males is a histone acetyltransferase involved in H3K56Ac, which has a role in meiosis.

Table 3. Differentially expressed genes between contrasts ($FDR < 0.01$, $\log_2FC > 0.58$).

Genes	B+♂ vs B-♂♀	B+♀ vs B-♀	B+♂ vs B+♀	B+♂ vs B-♀
Not DE	17634	17662	12403	12283
Overexpressed in focal group	286	233	2660	2815
Underexpressed in focal group	146	171	3003	2968

Fig 6. Differential gene expression analysis depending on presence or absence of B chromosomes in

2nd instar mealybugs. (A) Scatterplot of expression fold-changes in genes and average levels of expression between B+ and B- males. Upwards and downwards arrowheads represent genes that are significantly overexpressed and underexpressed, respectively, in B+ males. Genes located on B1 scaffolds are shown in dark blue. (B) Scatterplot of expression fold-changes in genes and average levels of expression between B+ and B- females. (C) Venn diagram showing overlaps in differentially expressed genes between B+ males and the other three groups: B- males, B+ females and B- females. Counts of overexpressed genes in B+ males shown above; underexpressed genes below. (D) Number of differentially expressed genes between B+ males and the other three sample groups grouped by location on the genome assembly. Number of annotated genes in each group are also shown in brackets.

Table 4. Annotated differentially expressed genes in B-carrying males with respect to the other three groups (B- males, B+ females and B- females), with best BLAST/DIAMOND hits ($E < 1e-25$) to UniProtKB/SwissProt and reference proteome databases. Genes located on candidate B scaffolds are highlighted in bold. Descriptions are inferred from Pfam/InterPro domains.

Gene	Scaffold	UniProt	RefProt	Description
Overexpressed in B+ males				
g1592	A (824)	-	-	Peptidase aspartic, putative
g1752	A (1732)	-	-	gag-polypeptide of LTR copia-type; Zinc finger, CCHC-type
g2486	A (49)	-	D6WTN9	Endonuclease
g2571	A (90)	-	E2A388	DDE superfamily endonuclease
g2891	A (89)	-	-	7TM chemoreceptor
g3441	A (10)	FACR1_DROME	J9K7D4	Male sterility, NAD-binding; fatty acyl-CoA reductase, C-terminal
g4226	A (28)	POLX_TOBAC	A0A0J7K6U3	Gag-polypeptide of LTR copia-type; RNA-dependent DNA polymerase); integrase core domain
g4426	A (94)	-	A0A2P8XIJ0	Transposase IS4
g5232	B3 (2160)	-	-	TAR DNA-binding protein 43, N-terminal
g5312	B2 (1578)	-	-	Transposase IS4
g5653	A (1013)	POLX_TOBAC	A0A182GA29	Gag-polypeptide of LTR copia-type; RNA-dependent DNA polymerase); integrase core domain
g5671	A (555)	-	A0A2P8ZLW3	Transposase IS4
g5914	A (139)	RENT1_DROME	T1HV29	Domain of unknown function (DUF5599); DNA2/NAM7 helicase (AAA domain); RNA helicase (UPF1, UPF2-interacting domain)
g6828	A (81)	-	A0A1U8N8I4	Domain of unknown function (DUF4371); hAT family C-terminal dimerisation region
g7375	A (774)	-	-	7TM olfactory receptor
g8319	A (120)	-	-	Myb/SANT-like DNA-binding domain
g9804	A (47)	TADBP_XENTR	T1HYV6	RNA recognition motif domain
g11244	A (4)	-	-	Myb/SANT-like DNA-binding domain
g11727	A (168)	-	X1WKK7	Transposase IS4

g12210	A (449)	-	Q8A7P0	Predicted AAA-ATPase
g13119	A (13)	-	D6WTN9	Endonuclease
g13773	A (1833)	-	K7JKD4	Hemolysin E
g13953	B1 (2040)	EP300_HUMAN	A0A3B3RK67	Histone acetyltransferase Rtt109/CBP
g14433	A (344)	-	A0A2P8XL08	FLYWCH zinc finger domain; MULE transposase domain
g14861	A (33)	CMTR2_MOUSE	A0A2P8ZED2	FtsJ-like methyltransferase
g15418	A (133)	-	-	Alpha carbonic anhydrase
g17869	A (146)	C4G15_DROME	J9K6X1	Cytochrome P450
g17908	A (983)	CP4G1_DROME	J9JU30	Cytochrome P450
g18982	A (163)	-	A0A482WV26	Uncharacterised
g19949	A (236)	-	A0A2P8XIJ0	Transposase IS4
g20017	A (894)	TCPA_MACFA	E2AJN5	Chaperonin Cpn60/TCP-1 family
g20018	A (894)	TCPA_DROME	A0A2J7QHGO	Chaperonin Cpn60/TCP-1 family
g20623	A (215)	PPIA_RABIT	A0A482XF32	Cyclophilin-type peptidyl-prolyl cis-trans isomerase
g20946	A (91)	-	A0A2P8Z519	FLYWCH zinc finger domain; MULE transposase domain
g21055	A (1221)	-	-	Gag-polypeptide of LTR copia-type
g22528	A (551)	-	-	Family of unknown function (DUF5641); peptidase aspartic, putative
g22599	A (6)	-	D6WTN9	Integrase zinc-binding domain
g22863	A (37)	SPOPB_XENLA	A0A1I8Q4G3	BTB/POZ domain
g23373	A (470)	-	-	Major facilitator, sugar transporter-like
g23380	A (1806)	DYH3_HUMAN	A0A482WUC2	Dynein heavy chain (axonemal)
g23478	A (48)	-	A0A0V0XSF6	Family of unknown function (DUF5641); Pao retrotransposon peptidase
Underexpressed in B+ males				
g1484	A (20)	GGH_HUMAN	A0A0L0CCZ4	Peptidase C26
g4173	A (28)	MRP1_CHICK	A0A139WC08	ABC transporter
g5248	B3 (814)	CATA_ASCSU	A0A2J7REJ5	Catalase
g17216	A (57)	VGLU1_XENLA	K7J9C8	Major facilitator superfamily
g19773	A (60)	GRM1_HUMAN	B0WMA9	Receptor, ligand binding region; GPCR, family 3, nine cysteines domain; GPCR family 3, C-terminal
g23148	A (64)	AMPN_RABIT	A0A195F0B1	Peptidase M1, membrane alanine aminopeptidase; ERAP1-like C-terminal domain

B chromosome transmission and prevalence in two greenhouse populations

The PCR primers that we designed to validate the B status of scaffold_552 also allowed us to easily detect B chromosome presence within individuals. We used primer sets 1 and 2 to survey two glasshouse populations in Edinburgh for the presence, frequency and transmission rate of the B chromosome. The first population at the Royal Botanical Gardens Edinburgh (RBGE), where we sampled two adjacent glasshouses, was the same population where we first discovered the B chromosome and from which we collected specimens that founded the laboratory populations used for this study. The second is an unrelated population at Butterfly World Edinburgh. We identified the prevalence and transmission in each population by taking females into the lab and genotyping 6-14 of their offspring. We found that the B chromosome was present in all glasshouses in both populations, but not fixed: we detected B chromosomes in at least one offspring of all females collected from the first RBGE glass house, but only in 3 out of 5 in the other RBGE glasshouse and in 13 out of 17 in the Butterfly world glasshouse. On average, 69% of offspring within each family carried a B chromosome (Supplementary figure S6), suggesting that the transmission rate is not 100% which likely contributes to the B chromosome prevalence we observed.

Discussion

Here we present the first genomic analysis of an unusual B chromosome found in a species of mealybug. This chromosome increases its transmission through males by escaping the fate that awaits other paternally-derived chromosomes under paternal genome elimination: being eliminated during meiosis. Using *de novo* generated genome data from individuals with different numbers of Bs (0-2) we were able to identify 1.37Mb of B-linked sequences (up to 7.11Mb when using more lenient criteria). These data allowed us to study both the origin and evolution of the B, as well as identify putative B-linked genes that might be involved in the drive.

B-linked gene identity, expression, and mechanisms of drive

From the genomic analysis of the B chromosomes in *P. viburni* we confidently identified 78 B-linked genes. We then studied the expression of these genes in 2nd/3rd instar males and females as this is the development stage when meiosis takes place. We identified 13 B-linked genes that were expressed during this life stage. As the B drives during male--but not female--meiosis, we expected that any B-linked gene involved in drive to be predominantly expressed in males at this stage. We identified just one B-linked gene, g13953, that was significantly overexpressed in males compared to females and encodes a histone acetylation protein. The protein is required for the histone modification H3K56Ac which is associated with actively transcribed genes (Skalska et al. 2015) and has a role in meiosis (Wang et al. 2017). This gene is a promising candidate for its involvement in B chromosome drive in males, as the chromatin state of the B chromosome appears to play a key role in its ability to drive (Nur and Brett 1985; Nur and Brett 1988), so we expect any genes involved in chromatin remodeling could play a role. Apart from the H3K56 acetylation gene, g20108 is a gene with no known function in insects; however, it contains an integrase core domain and putative chromo domain. Even though this gene shows expression in both sexes in our RNA-seq analysis, a separate qPCR analysis shows an upregulation in male pupae compared to female last nymphal instar and earlier male stages (see Fig S8). There are also two B-linked kinesin-like protein genes, g16323 and g16324, that could be of interest. Kinesins are proteins that move along microtubule filaments and support cellular functions including meiosis. Despite the intriguing functional annotation, however, they show very low expression at this developmental stage.

In comparison with B chromosomes in other species, the B in *P. viburni* is remarkably gene rich. But, we find that most of these genes have either no or very low expression in the stage that we investigated. There are a number of possible explanations for this. First of all, it is possible that these genes are expressed in other life stages which we did not examine. Secondly, it is possible that their expression is tissue-specific and therefore challenging to detect in whole-body RNA-seq data. Finally, it is likely that the expression of B-linked genes is suppressed, especially in males when paternally-derived (which will be by far their most common state as transmission is male-biased). Throughout development in both sexes, the B appears highly condensed (heterochromatic) in comparison to the autosomes (Nur, 1966). In males, they are not only condensed but also part of the heterochromatic body at the periphery of the nucleus which contains the paternally-derived chromosomes. Recent work in another mealybug species (*P. citri*) shows that gene expression of the condensed paternal chromosomes is low but not absent (de la Filia, 2021). So even if Bs are heavily condensed, they might still show some gene expression. However, to better understand patterns of expression of B-linked genes, we will need to obtain expression data from different tissues and life stages throughout development. Combining this with population genetic analyses will also provide insights in the patterns of selection experienced by B-linked genes.

The B chromosome in *P. viburni* is not alone in its apparent ability to drive by manipulation of the chromatin state of itself and the other chromosomes. In the haplodiploid wasp *Nasonia vitripennis*, a B chromosome called *PSR* prevents the sperm chromatin from dissolving after fertilization, which leads to the elimination of all sperm-derived chromosomes with exception of *PSR* (Nur et al. 1988; Swim et al. 2012; Dalla Benetta et al. 2020). The resulting embryo develops as a haploid male rather than a diploid female, giving the *PSR* chromosome an advantage as it drives in males. The mechanism by which *PSR* accomplishes its drive through chromatin manipulation is fairly well understood: a B-linked gene called *haplodizer* appears to manipulate the histone code on the autosomes (Dalla Benetta et al. 2020). In particular, *PSR* induces an enrichment of H3K9Me3 and H3K20Me1, histone modifications that are associated with transcriptional suppression. Interestingly, the same histone modifications are associated with paternally-derived chromosomes in mealybugs and may play a key role in the condensation level of paternal chromosomes (Bongiorni et al. 2009; Bain 2019; Li et al. 2020). Furthermore, the degree of condensation could affect the probability that paternal chromosomes are eliminated during meiosis, with those with a more diffuse appearance more likely to pair with the maternal chromosomes and being included into functional sperm (Brown and Nur 1964; Nur and Brett 1985; Nur and Brett 1988). So a possible mechanism for drive in *P. viburni* would be that, similar to

N.vitripennis, a B-linked gene directly manipulates the chromatin state. The two genes we have identified that have putative roles in regulating chromatin structure are therefore promising candidates for further exploration.

The mechanism by which the B chromosome drives is not likely dependent on its gene content only, but also on its structural properties, in particular its repeat content. Many B chromosomes are enriched for different types of repetitive elements. In particular satellite repeats - often unique to the B - make up a large proportion of the B chromosome in many species. For example, *PSR* in *N. vitripennis* has a repeat content of 89,8%, >70% of which consists of just 4 satellite repeats (Li et al. 2017). While the B chromosome in the migratory locust contains almost 95% repeats, with over 50% of the chromosome consisting of just a single satellite (Palacios-Gimenez et al. 2020). Although the role of B-specific repeats is still poorly understood, B-specific repeats may influence the chromatin state of the B and possibly also of the autosomes, if the B acts as a sink for chromatin proteins (Eickbush et al. 1992). Translocation experiments in *P. viberni* support this possibility, as even small fragments of the B translocated onto the autosomes decondense (and therefore presumably change their chromatin structure) during meiosis to similar degree as the entire B chromosome (Nur & Brett, 1988). This suggests it is unlikely that a single gene or region might be sufficient for controlling the Bs behaviour. In this study, we identify a number of B-specific satellite repeats that make up a significant fraction of the chromosome. These would be an interesting target for future studies into the role repeats play in the Bs ability to drive.

Evolutionary origin of the B chromosome

Apart from determining the genetic composition of the B chromosome in *P. viberni*, we also investigated its evolutionary origin. Despite their frequent and widespread evolution, the origins of B chromosomes often remain obscure. In general, B chromosomes are assumed to have evolved from the core genome (Camacho et al. 2000). However, as they are generally rich in repetitive sequences and significantly diverged, it can be hard to pinpoint from which part of the genome they originate. There is also a suggestion that at least in some species, the B might have a hybrid origin (McAllister and Werren 1997). In *P. viberni*, it is unlikely that the B originates from pericentromeric regions of the core genome, as mealybug chromosomes are holocentric and lack a localised centromere (Bongiorni et al. 2004). We find that the B contains a large number of repetitive elements and is gene-rich. Yet, only a small number of B genes have paralogs on the core genome and several of the repeats are unique to the B. We also find that genes shared between the B and A

chromosomes show substantial divergence. The same is true for homologous non-coding regions shared between the A and B chromosomes. However, when we compare divergence of B-linked sequences between the core genome of *P. viburni* and two other mealybug species, we find that they are most closely related to *P. viburni*, suggesting a within-species origin. The B chromosome does not show clear homology to any particular chromosome in the core genome and many of the genes and repeats are unique to the B chromosomes. However, many of the genes were found to be homologous to other hemipteran genes. Hence the genes unique to B chromosomes likely represent translocations from the core genome onto the B. Alternatively the B could be derived from a small portion of one of the autosomes that disassociated, something that is possible in species with holocentric chromosomes such as *P. viburni* where small chromosomal fragments can fatefully segregate.

B chromosome frequency and transmission

We designed B-specific primers that allowed us to study the frequency of the B chromosome in two glass house populations in Edinburgh, UK. We find a high prevalence of the B chromosome in both populations; however, the B is not fixed and seems to be at about 70% frequency. B chromosome frequencies are likely determined by a number of factors. The rate at which the B drives through males is doubtless important, but so is the transmission rate through females. Previous studies suggest that the B shows a lower rate of transmission through female meiosis than expected under Mendelian inheritance (U. Nur 1962). In addition, the B may affect the fitness of its carriers; there is evidence in *P. viburni* that the B chromosome reduces male fitness and fertility (Nur, 1966). Finally, it is possible that autosomal suppressors are segregating in populations. B drive suppression mechanisms have been documented in other species, such as the post-meiotic ejection of Bs in grasshoppers (Cabrero et al. 2017) (Cabrero 2017). Nur & Brett (1987, 1988) found that the transmission rates of the B vary and are determined by the maternal genotype of a male. This suggests that there might be resistance alleles segregating in the host population. Comparing both the B chromosome chromatin landscape and autosomal gene expression in these mealybug genotypes will be another promising avenue for studying the mechanism of B drive.

Conclusion

In this study, we identify B-linked sequences from the mealybug *Pseudococcus viburni*. This chromosome increases its transmission through males by exploiting the unusual reproductive system of mealybugs: in males all paternally-derived chromosomes are silenced (heterochromatinised) early in development and subsequently eliminated during spermatogenesis. The B chromosome is able to avoid paternal genome elimination by pairing with maternally-derived chromosomes during meiosis and ensuring its inclusion into cells destined to form mature sperm. It appears to do so by changing its chromatin state to match that of the euchromatic maternal chromosomes. We were able to identify a large number of B-linked genes that could play a role in the chromosome's ability to drive. In particular, two genes overexpressed in males and involved in histone modification and chromatin structure are promising candidates for future work. We also identified B-specific satellite repeats that might play a role in the Bs ability to drive, either by influencing its chromatin structure, or by being the source of long or small non-coding RNA. Further studies into these B-linked genes and regions will help better understand not only the mechanisms by which the B chromosome accumulates, but also improve our understanding of the way paternally-inherited chromosomes are eliminated during male meiosis. This would provide insights into paternal genome elimination, a peculiar but widespread reproductive system that has evolved repeatedly across arthropods.

Materials and methods

Mealybug populations and cultures

The obscure (or glasshouse) mealybug *Pseudococcus viburni* (Hemiptera: Pseudococcidae) is a small phloem-feeding insect with a global distribution. It is able to feed on a large number of plant species and is a common pest across many regions in glasshouses, fruit orchards and vineyards. We collected egg-laying *P. viburni* females from a variety of host plants in the glasshouses of the Royal Botanic Gardens of Edinburgh (UK) and transferred them to sprouted potatoes. These experimental populations were kept at 25°C degrees (16h-light/8h-dark photoperiod without humidity control) in tupperware boxes modified to have a large hole on the lid and sealed with a screen printing mesh, with a tissue paper at the bottom of the box to absorb excess moisture.

Species identification was confirmed using multiplex PCR on extracted DNA from single individuals per collected population, using primers previously published (Daane et al. 2011). gDNA was extract using the prepGEM Insect kit (ZyGEM, New Zealand) following manufacturer instructions, and multiplex PCR was carried out using the Qiagen Multiplex PCR Kit (Qiagen), following reaction specifications in (Daane et al. 2011). 4ul of PCR product was visualised by electrophoresis on a 2% agarose gel stained with GelRed Nucleic Acid Gel Stain (Biotium).

We established the experimental lines used in this study in plastic fly medium bottles containing a sprouted potato by transferring single egg-laying females and letting the offspring grow, reach adult stage and mate. Several lines were established from the base population and these were regularly screened for B chromosomes using karyotyping. Information on the final lines used for this study can be found in Supplementary Table S1. Successive generations were established by repeating the process once egg-laying female offspring could be spotted. This resulted in a number of highly inbred iso-female lines.

B chromosome screening

To screen for B chromosomes, we prepared cytological slides at embryonic stage (1-2 days after egg laying) and during male second instar and prepupal stages. For embryos, laid eggs were transferred onto a coverslip with 20uL of 45% acetic acid, and squashed using a microscope slide. The coverslip was then snapped after emerging the microscope slide in liquid nitrogen, and about 40 uL of 1:3 acetic acid:ethanol was applied

on the tissue and air dried. For male juvenile stages, whole males were fixed in Bradley Carnoy between 24-72 hours at 4°C. Testes were then dissected on a coverslip in 20uL of 45% acetic acid. A slide was then applied on top of the coverslip and turned coverslip face up for squashing between a piece of filter paper. The coverslip was then snapped by dipping the slide into liquid nitrogen. 40uL of Ethanol:Acid Acetic 3:1 was applied on the sample and the slide was left to dry. DNA was visualised with Vectashield Antifade Mounting Medium with Dapi (Vector Laboratories), and images were taken on a Leica DM2500 fluorescence microscope, using the Leica application suite software. Link to detailed protocol: dx.doi.org/10.17504/protocols.io.mnec5be.

Sequencing of B and non-B experimental lines

DNA-seq. We transferred 1-3 egg-laying females from two B-carrying experimental lines (PV04 and PV13) and two non-B lines (PV21 and PV23) from the cultures to individual bottles and let them lay eggs for 24h. The females were then flash frozen and stored at -80C. After karyotyping of eggs to confirm expected presence/absence of B chromosomes, we extracted genomic DNA from a pool of three females from the PV13 line and single females from the other three lines using a combination of two DNA extraction kits. The full protocol is available online (dx.doi.org/10.17504/protocols.io.pvgdn3w). The amount of extracted DNA was measured using the Qubit dsDNA Assay Kit (ThermoFisher). PV04, PV13, PV21 and PV23 TruSeq DNA Nano libraries (Illumina) with 350bp insert sizes and an additional 550bp library for PV13 were generated by Edinburgh Genomics (UK) and sequenced on a Novaseq S1 flowcell to yield 203-402M 150bp paired-end reads per line. Additionally, a PacBio SMRTbell 20Kb library (Pacific Biosciences) was generated from a pool of 60 adult PV13 females and sequenced on a SMRT Cell 1M v3 on the PacBio Sequel platform, generating 28.6Gb of data (number of subreads = 1.7M, subread N50 = 14.9Kb). All reads were subjected to filtering for endosymbionts that are not part of this study and deposited to ENA under accession (PRJEB47083).

RNA-seq. We isolated, sexed, and flash-froze pools of 30-60 individuals from PV04, PV13, PV21, and an additional experimental line, PV15, from the same original non-B population. All specimens were late 2nd or early 3rd instars. This is the stage when males spin a cocoon and undergo a process similar to metamorphosis and therefore the earliest stage at which specimens can be sexed reliably. We extracted total RNA from 3-4 biological replicates per sex and line with a PureLink RNA purification kit (Thermo Fisher Scientific, USA) and validated the samples with the Bioanalyzer RNA 6000 Nano kit (Agilent). 23

TruSeq stranded mRNA-seq libraries (Illumina) were generated by Edinburgh Genomics and sequenced on the NovaSeq platform, yielding 86-136M 50bp paired-end reads per sample.

Genome assembly and annotation

We assembled the PacBio subreads with wtdbg2/Redbean v2.5 (Ruan and Li 2020) using the default settings for PacBio Sequel technology. To polish the raw assembly, we ran three rounds of HyPo v1.0.3 (Kundu et al. 2019) using the PacBio reads and Illumina short reads from the 350 and 550bp PV13 libraries (see below for read processing). During polishing, we used minimap2 v2.17 (Li 2018) to map long and short reads to the successive assemblies with the --secondary=no setting.

After polishing, we performed a first round of scaffolding using transcriptomic evidence with SCUBAT2 (<https://github.com/GDKO/SCUBAT2>), using a *de novo* transcriptome assembly assembled from all male and female RNA-seq samples using Trinity-v2.8.5 (Grabherr et al. 2011). We only used transcripts with TPM>2 in all 23 RNA-seq samples and a maximum intron size of 50,000. We then ran a second round of scaffolding with BESST v2.2.8 (Sahlin et al. 2014) and the Illumina short reads. Then, we followed the blobtools v1.1.1 pipeline (Laetsch and Blaxter 2017b) to remove contaminants from the assembly: briefly, contigs/scaffolds with homology hits to Proteobacteria and increased genomic coverage, indicating a bacterial endosymbiont origin, and contigs without taxonomic hits and coverage <2). Throughout the process, we used BUSCO v4.0.6 (Simão et al. 2015) with the Hemiptera and Insecta gene sets to assess completeness of the successive iterations of the assembly.

To annotate the genome, we first ran RepeatMasker v4.1.0 (Smit et al. 2019) to soft mask repetitive regions in the assembly using a repeat library combining a *de novo* library built with RepeatModeler v2.0 and the hemipteran RepBase database. To predict genes, we followed the BRAKER2 v2.1.4 (Hoff et al., 2015) pipeline using our 23 RNA-seq samples, which we mapped to the assembly using STAR v2.7.4a (Dobin et al. 2013) in the single pass mode. We annotated the predicted transcripts with homology-based searches against the UniProtKB/Swiss-Prot (BLASTp v2.9.0, E-value = 1e-25) and reference proteome databases (diamond v0.9.29, E-value = 1e-25) and running InterProScan (Quevillon et al. 2005). to assign Pfam and InterPro domains and GO terms.

Identification of B chromosome scaffolds in the assembly

We took three complementary approaches to define a set of B-linked scaffolds in the assembly: genomic coverage, assembly mapping, and kmer coverage.

To identify candidate B regions based on differential genomic coverage, we mapped the Illumina short reads generated for the four experimental lines to our reference genome, which should contain the B. We used FastQC v0.11.5 (<https://qubeshub.org/resources/fastq>) for read quality control and fastp v0.20 (Chen et al. 2018) for quality and adapter trimming with the following settings: `--detect_adapter_for_pe --cut_by_quality5 --cut_by_quality3 --cut_window_size 4 --cut_mean_quality 20 --trim_poly_g`. We mapped the reads from each line to the reference genome using bwa mem v0.7.17-r1188 (Li 2013) with default settings (mapping rates 94.8-97.8%). For PV13, we first mapped both 350bp and 550bp libraries separately and then merged the resulting BAM files. We then used samtools v1.9 (Li et al. 2009) to filter out secondary mappings and bamtools (Barnett et al. 2011) to remove reads with edit distance >0 (NM tag) in order to keep only primary mappings with no mismatches. We collected numbers of reads mapping to each scaffold using samtools idxstats and estimated scaffold coverage differences between pairs of lines after normalising by the median coverage difference across all scaffolds. We defined two thresholds to identify putative B-linked scaffolds: a 'relaxed' threshold of $\log_2(B^+ \text{ line}/B^- \text{ line}) > 0.58$ in all 4 pairs (which considers a hypothetical B region of autosomal origin that is present three times in an individual carrying one B chromosome and twice in a non-B individual) and a more conservative threshold of $\log_2 > 2$ (the 'strict' criterion).

To identify B scaffolds via assembly mapping, we assembled the trimmed short read libraries separately into assemblies for each of the four lines with SPAdes v3.11.1 (Bankevich et al. 2012) with the following settings: `--only-assembler -k 21,33,55,77,99,127`. Then, we mapped the four assemblies to the reference genome using the nucmer aligner from MUMmer v4.0.0beta2 (Marçais et al. 2018) and used the dnadiff wrapper to extract 1-to-1 alignments between the reference genome and the assemblies and determine which reference scaffolds map only to B+ lines.

To identify B scaffolds using kmer coverage, we loosely followed the strategy developed by Hodson et al. (2021). This method uses differences in k-mer frequencies between sequencing libraries to identify overrepresented (and thus putatively B-linked) kmers in B+ libraries independently of genome assembly. We used KMC3 (Kokot et al. 2017) to extract 27-mers with at least 3 counts in all libraries and their coverages from the raw reads (after removing contaminant reads using the blobtools pipeline; see above). We used

total kmer coverage estimates from each sequencing library (as determined with GenomeScope v1, <http://qb.cshl.edu/genomescope/info.php>) to account for differences in sequencing depth between lines and normalise coverages of individual kmers. Since B chromosome kmers may be unique to Bs or shared between A and B, we used the relaxed threshold of $\log_2(B+/B- \text{ lines}) > 0.58$ (see above) to assign putative B origin to kmers. Considering only kmers with normalised coverage > 5 in at least one line, we classified 49.7M kmers as putatively B-linked, while 504M kmers were considered autosomal. We then mapped these kmers to the assembled scaffolds in the reference genome using bwa mem with the following settings: -k 27 -T 27. Scaffolds with as many putative B as autosomal kmer mappings ($\log_2 \geq 0$) were considered B candidates.

Gene expression quantification and differential expression analyses

We used FastQC and fastp for raw RNA-seq data processing and quantified gene expression levels in all 23 samples using RSEM v1.3.0 (Li and Dewey 2011) with the STAR aligner (Dobin et al. 2013).

We evaluated differential gene expression using the edgeR/limma workflow described in (Law et al. 2016). For the sex-biased differential expression analysis, we only included sex as a factor. Briefly, we removed lowly expressed genes (minimum read count = 10), converted raw data to $\log_2$ (CPMs) and normalised expression distributions with the TMM method using edgeR (Robinson et al., 2010). We then validated all our RNA-seq samples using density plots, $\log_2$ (CPM) boxplots, MDS plots and correlation heatmaps. We transformed the data using the voom function and built a linear model in limma with empirical Bayes moderation (Ritchie et al., 2015). We required a threshold of $\log(\text{fold change}) > 1$ (twice expressed in one sex compared to the other) for differentially expressed genes (with an $\text{FDR} < 0.05$). We conducted the differential expression analysis between B+ and B- samples identically, with sex and B presence as factors, but with the minimum read count and $\log(\text{fold change})$ lowered, respectively, to 5 and 0.58 (50% higher expression) to account for the overall low expression of B1 genes (Supplementary Figure S3). We conducted gene ontology (GO) enrichment analyses for differentially expressed using the hypergeometric test implemented in the GOSTats R package (Falcon and Gentleman 2007) with a p-value cutoff of 0.01.

PCR amplification of B-linked primers

We used Primer3 (<https://primer3.ut.ee>) to design PCR primers for the longest B-linked scaffold, scaffold_552 (Supplementary Table S3). For a single individual, PCR amplification was performed in a 25 µl reaction volume containing 1 µL of prepGEM-extracted gDNA product, 12.5 µl DreamTaq Green PCR Master Mix (ThermoFisher), 0.25 µl forward primer and 0.25 µl reverse primer. PCR reactions were performed under the following conditions: initial denaturation at 95°C for 1 min, 30 cycles of denaturation at 95°C for 30s, annealing at 55°C for 30s and extension at 72°C for 1m and a final extension step at 72°C for 5 minutes. PCR products were placed under electrophoresis at 100V in a 2% agarose gel with 1% TAE buffer for 40 minutes with a 1kb DNA ladder.

Acknowledgements

We would like to thank Patrick Ferree for his comments, help and support designing this study, glasshouse staff from the Royal Botanical Gardens and Butterfly House Edinburgh for allowing us to collect specimens. The work was funded by an Marie Skłodowska-Curie Individual Fellowship to IV, a Dorothy Hodgkins Fellowship to LR while KJ and AM were funded by an ERC starting grant.

References

- Ahmad SF, Martins C. 2019. The modern view of b chromosomes under the impact of high scale omics analyses. *Cells* 8:156.
- Bain SA. 2019. Epigenetic mechanisms underlying paternal genome elimination. <http://hdl.handle.net/1842/35675>
- Bain SA, Marshall H, Filia AG de la, Laetsch DR, Husnik F, Ross L. 2021. Sex-specific expression and DNA methylation in a species with extreme sexual dimorphism and paternal genome elimination. *Mol. Ecol.* <https://doi.org/10.1111/mec.15842>
- Banaei-Moghaddam AM, Schubert V, Kumke K, Weiß O, Klemme S, Nagaki K, Macas J, González-Sánchez M, Heredia V, Gómez-Revilla D. 2012. Nondisjunction in favor of a chromosome: the mechanism of rye B chromosome drive during pollen mitosis. *Plant Cell* 24:4124–4134.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455–477.
- Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. BamTools: a C++

- API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27:1691–1692.
- Bongiorni S, Fiorenzo P, Pippoletti D, Prantera G. 2004. Inverted meiosis and meiotic drive in mealybugs. *Chromosoma* 112:331–341.
- Bongiorni S, Mazzuoli M, Masci S, Prantera G. 2001. Facultative heterochromatinization in parahaploid male mealybugs: involvement of a heterochromatin-associated protein. *Development* 128:3809–3817.
- Bongiorni S, Pugnali M, Volpi S, Bizzaro D, Singh PB, Prantera G. 2009. Epigenetic marks for chromosome imprinting during spermatogenesis in coccids. *Chromosoma* 118:501–512.
- Brown SW, Nur U. 1964. Heterochromatic Chromosomes in Coccids. *Science* 145:130–136.
- Burt A, Trivers RL. 2006. *Genes in Conflict*. Cambridge: Harvard University Press
- Cabrero J, Martín-Peciña M, Ruiz-Ruano FJ, Gómez R, Camacho JPM. 2017. Post-meiotic B chromosome expulsion, during spermiogenesis, in two grasshopper species. *Chromosoma* 126:633–644.
- Camacho JPM. 2005. B chromosomes. In: *The evolution of the genome*. Elsevier. p. 223–286.
- Camacho JPM, Sharbel TF, Beukeboom LW. 2000. B-chromosome evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 355:163–178.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *bioRxiv*:274100–274100.
- Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, Stensmyr MC, Zheng Y, Liu W, Sun H. 2016. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol.* 14:1–15.
- Daane KM, Middleton MC, Sforza R, Cooper ML, Walton VM, Walsh DB, Zaviezo T, Almeida RP. 2011. Development of a multiplex PCR for identification of vineyard mealybugs. *Environ. Entomol.* 40:1595–1603.
- Dalla Benetta E, Akbari OS, Ferree PM. 2019a. Sequence expression of supernumerary B chromosomes: Function or fluff? *Genes* 10:123.
- Dalla Benetta E, Akbari OS, Ferree PM. 2019b. Sequence expression of supernumerary B chromosomes: Function or fluff? *Genes* 10:123.
- Dalla Benetta E, Antoshechkin I, Yang T, Nguyen HQM, Ferree PM, Akbari OS. 2020. Genome elimination mediated by gene expression from a selfish chromosome. *Sci. Adv.* 6:eaaz9808.
- Dallai R. 2012. Baccio baccetti and the comparative spermatology of insects. *Redia* 95.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
- Eickbush DG, Eickbush TH, Werren JH. 1992. Molecular characterization of repetitive DNA sequences from a B chromosome. *Chromosoma* 101:575–583.
- Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* 23:257–258.
- de la Filia AG, Bain SA, Ross L. 2015. Haplodiploidy and the reproductive ecology of Arthropods. *Curr. Opin. Insect Sci.* 9:36–43.
- de la Filia AG, Mongue AJ, Dorrens J, Lemon H, Laetsch DR, Ross L. 2020. Males that silence their father's genes: genomic imprinting of a complete haploid genome. *bioRxiv*:2020.04.27.063396.
- Garber AI, Kupper M, Laetsch DR, Weldon SR, Ladinsky MS, Bjorkman PJ, McCutcheon JP. 2021. The evolution of interdependence in a four-way mealybug symbiosis. *Genome Biol. Evol.*:evab123.
- Gardner A, Ross L. 2014. Mating ecology explains patterns of genome elimination. Van Baalen M, editor. *Ecol. Lett.* 17:1602–1612.
- Gardner A, Úbeda F. 2017. The meaning of intragenomic conflict. *Nat. Ecol. Evol.*

- 1:1807–1807.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29:644.
- Hodson CN, Jaron KS, Gerbi S, Ross L. 2021. Evolution of gene-rich germline restricted chromosomes in black-winged fungus gnats through introgression (Diptera: Sciaridae). *bioRxiv*:2021.02.08.430288.
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2015. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32:767–769.
- Houben A. 2017. B chromosomes—a matter of chromosome drive. *Front. Plant Sci.* 8:210.
- Hughes-Schrader S. 1948. Cytology of coccids (Coccoidea-Homoptera). *Adv. Genet.* 35:127–203.
- Hurst LD. 1992. Intragenomic conflict as an evolutionary force. *Proc. Biol. Sci.* 248:135–140.
- International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLOS Biol* 8:e1000313.
- Jones RN. 1991. B-chromosome drive. *Am. Nat.* 137:430–442.
- Kokot M, Długosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* 33:2759–2761.
- Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.* 18:205–214.
- Kundu R, Casey J, Sung W-K. 2019. HyPo: Super Fast & Accurate Polisher for Long Read Genome Assemblies. *bioRxiv*.
- Laetsch DR, Blaxter ML. 2017a. BlobTools: Interrogation of genome assemblies. *F1000Research* 6:1287.
- Laetsch DR, Blaxter ML. 2017b. BlobTools: Interrogation of genome assemblies. *F1000Research* 6:1287.
- Law CW, Alhamdoosh M, Su S, Dong X, Tian L, Smyth GK, Ritchie ME. 2016. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* 5.
- Lenormand T, Engelstädter J, Johnston SE, Wijnker E, Haag CR. 2016. Evolutionary mysteries in meiosis. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20160001–20160001.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323–323.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Prepr. ArXiv13033997*.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li M, Tong H, Wang S, Ye W, Li Zicheng, Omar MAA, Ao Y, Ding S, Li Zihao, Wang Y, et al. 2020. A chromosome-level genome assembly provides new insights into paternal genome elimination in the cotton mealybug *Phenacoccus solenopsis*. *Mol. Ecol. Resour.* [Internet] n/a. Available from: <https://doi.org/10.1111/1755-0998.13232>
- Li Y, Jing XA, Aldrich JC, Clifford C, Chen J, Akbari OS, Ferree PM. 2017. Unique sequence organization and small RNA expression of a “selfish” B chromosome. *Chromosoma* 126:753–768.
- Lindholm AK, Dyer KA, Firman RC, Fishman L, Forstmeier W, Holman L, Johannesson H, Knief U, Kokko H, Larracuente AM. 2016. The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* 31:315–326.
- Lyttle TW. 1993. Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends Genet.* 9:205–210.
- Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A

- fast and versatile genome alignment system. *PLoS Comput. Biol.* 14:e1005944.
- McAllister BF, Werren JH. 1997. Hybrid origin of a B chromosome (PSR) in the parasitic wasp *Nasonia vitripennis*. *Chromosoma* 106:243–253.
- Nur U. 1962. A supernumerary chromosome with an accumulation mechanism in lecanoid genetic system. *Chromosoma* 13:249–271.
- Nur Uzi. 1962. Sperms, sperm bundles and fertilization in a mealy bug, *Pseudococcus obscurus* Essig. (Homoptera: Coccoidea). *J. Morphol.* 111.2: 173-199.
- Nur U. 1966. Effect of Supernumerary Chromosomes on Development of Mealy Bugs. *Genetics* 54:1239;
- Nur Uzi. 1966. Harmful supernumerary chromosomes in a mealy bug population. *Genetics* 54:1225;
- Nur U. 1969. Harmful B-Chromosomes in a Mealy Bug - Additional Evidence. *Chromosoma* 28:280;
- Nur U. 1990. Heterochromatization and Euchromatization of Whole Genomes in Scale Insects (Coccoidea, Homoptera). *Development Supplement*:29–34.
- Nur U, Brett BLH. 1985. Genotypes Suppressing Meiotic Drive of a B-Chromosome in the Mealybug, *Pseudococcus-Obscurus*. *Genetics* 110:73–92.
- Nur U, Brett BLH. 1988. Genotypes affecting the condensation and transmission of heterochromtic B chromosomes in the mealybug *Pseudococcus affinis*. *Chromosoma* 96:205–212.
- Nur U, Werren JH, Eickbush DG, Burke WD, Eickbush TH. 1988. A Selfish B-Chromosome That Enhances Its Transmission by Eliminating the Paternal Genome. *Science* 240:512–514.
- Nur U-102. ris. 1980. Evolution of unusual chromosome systems in scale insects (Coccoidea: Homoptera). In: Hewitt GM, Ashburner M, editors. *Insect Cytogenetics*. Blackwell, Oxford. p. 97–118.
- Palacios-Gimenez OM, Milani D, Song H, Marti DA, López-León MD, Ruiz-Ruano FJ, Camacho JPM, Cabral-de-Mello DC. 2020. Eight Million Years of Satellite DNA Evolution in Grasshoppers of the Genus *Schistocerca* Illuminate the Ins and Outs of the Library Hypothesis. *Genome Biol. Evol.* 12:88–102.
- Pardo MC, López-León MD, Cabrero J, Camacho JPM. 1994. Transmission analysis of mitotically unstable B chromosomes in *Locusta migratoria*. *Genome* 37:1027–1034.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* 33:W116–W120.
- Reed KM, Werren JH. 1995. Induction of paternal genome loss by the paternal-sex-ratio chromosome and cytoplasmic incompatibility bacteria (Wolbachia): a comparative study of early embryonic events. *Mol. Reprod. Dev.* 40:408–418.
- Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* 17:155–158.
- Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM. 2016. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* 6:28333.
- Sahlin K, Vezzi F, Nystedt B, Lundberg J, Arvestad L. 2014. BESST-efficient scaffolding of large fragmented assemblies. *BMC Bioinformatics* 15:281.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
- Skalska L, Stojnic R, Li J, Fischer B, Cerda-Moya G, Sakai H, Tajbakhsh S, Russell S, Adryan B, Bray SJ. 2015. Chromatin signatures at Notch-regulated enhancers reveal large-scale changes in H3K56ac upon activation. *EMBO J.* 34:1889–1904.
- Smit A, Hubley R, Green P. 2019. 2013–2015. RepeatMasker Open-4.0.
- Swim MM, Kaeding KE, Ferree PM. 2012. Impact of a selfish B chromosome on chromatin dynamics and nuclear organization in *Nasonia*. *J. Cell Sci.* 125:5241–5249.
- Szathmáry E, Smith JM. 1995. The major evolutionary transitions. *Nature* 374:227–232.

- Vea IM, Tanaka S, Shiotsuki T, Jouraku A, Tanaka T, Minakuchi C. 2016. Differential Juvenile Hormone Variations in Scale Insect Extreme Sexual Dimorphism. *PLOS ONE* 11:e0149459.
- Wang L, Xu Z, Khawar MB, Liu C, Li W. 2017. The histone codes for meiosis. *Reproduction* 154:R65–R79.
- Werren JH. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci.* 108:10863–10870.
- Xiao S-J, Zhang C, Zou Q, Ji Z-L. 2010. TiSGeD: a database for tissue-specific genes. *Bioinformatics* 26:1273–1275.