Genetics and Genomics of an Unusual Selfish Sex Ratio Distortion in an Insect

Phineas T. Hamilton, Christina N. Hodson, Caitlin I. Curtis, Steve J. Perlman

December 2018

Accepted manuscript:

Hamilton PT, Hodson CN, Curtis CI, Perlman SJ. 2018. Genetics and genomics of an unusual selfish sex ratio distortion in an insect. *Current Biology. **28**, 3864–3870.*

Genetics and genomics of an unusual selfish sex ratio distortion in an insect

Phineas T. Hamilton[1,2*], Christina N. Hodson[1,3], Caitlin I. Curtis[1], Steve J. Perlman[1*]

[1]Department of Biology, University of Victoria, Victoria, British Columbia, Canada
[2]The Deeley Research Centre, BC Cancer, Victoria, British Columbia, Canada

[*]Correspondence: stevep@uvic.ca; phin.hamilton@gmail.com

Abstract/Summary:

Diverse selfish genetic elements have evolved the ability to manipulate reproduction to increase their transmission, and this can result in highly distorted sex ratios. Indeed, one of the major explanations for why sex determination systems are so dynamic is because they are shaped by ongoing coevolutionary arms races between sex ratio distorting elements and the rest of the genome. Here, we use genetic crosses and genome analysis to describe an unusual sex ratio distortion with striking consequences on genome organization in a booklouse species, *Liposcelis* sp. (Insecta: Psocodea), in which two types of females coexist. Distorter females never produce sons but must mate with males (the sons of nondistorting females) to reproduce. Although they are diploid and express the genes inherited from their fathers in somatic tissues, distorter females only ever transmit genes inherited from their mothers. As a result, distorter females have unusual chimeric genomes, with distorter-restricted chromosomes diverging from their nondistorting counterparts and exhibiting features of a giant nonrecombining sex chromosome. The distorter-restricted genome has also acquired a gene from the bacterium *Wolbachia*, a well-known insect reproductive manipulator; we found that this gene has independently colonized the genomes of two other insect species with unusual reproductive systems suggesting possible roles in sex ratio distortion in this remarkable genetic system.

Keywords: Sex ratio, *Wolbachia*, Paternal genome elimination, Selfish genetic elements, Genetic conflict, Sex determination, Lice, Booklice

Results

Distorter females only transmit genes they inherit from their mother

We previously found an unusual sex ratio distortion in a booklouse, *Liposcelis* sp., whereby some females, which we call distorters, produce daughters exclusively [1]. This trait is inherited strictly maternally but microbes are not involved. To understand the genetic basis of this sex ratio distortion, we genotyped offspring from five F1 crosses (N = 69; Figure 1, Table S1) and eight F2 crosses (N = 101 from four F1 families; Table S2), using a polymorphism at the *Phos1* gene [2]. We found only one of the two expected alleles present in the offspring of distorter females ($P < 0.005$ for all crosses), and that all distorter females always transmitted the same allele to offspring (A' allele in Tables S1 & S2). This means that this locus is strictly maternally transmitted in distorters but is transmitted in a Mendelian manner in nondistorter females [2]).

Distorter genomes are chimeras

We previously found that the mitochondrial genomes of distorter and nondistorter females have radically different architectures and encode highly divergent proteins [1]. During assembly of Illumina reads to study this architecture, we noticed additional pervasive and unusual differences, with the nondistorter genome assembling at a much smaller size (260 vs 330 Mb) and a much higher N_{50} (4125 vs 733 bp), despite similar read abundance (Table S3), suggesting substantial additional complexity in the distorter genome.

To investigate this more rigorously, we undertook a comprehensive assembly with additional Illumina reads from nondistorter females, males, and distorter females, and PacBio long reads from distorter females to produce a 'pan-genome' that encompassed nondistorter and distorter females (STAR methods; Table S3). We supplemented this assembly with transcriptome sequencing and assembly for gene prediction using replicate samples of nondistorter and distorter females (N = 3 per female type), yielding a pan-genome of 391 Mb with a contig N_{50} of 71.3 kb and 13,314 predicted protein coding

genes. Because of the complexity of this genome, we emphasize its draft nature and restrict our interpretation accordingly.

We compared the distribution of predicted genes across distorter and nondistorter females by calculating the coverage for genes in each female type (as FPKM DNA reads mapping to exons). We found that a large proportion of predicted genes appeared to be entirely restricted to the distorter or showed much lower coverage in the nondistorter, suggesting that the distorter female possesses additional gene content that is effectively unique to it. Lower, but non-zero, read mappings from nondistorter females relative to distorters for some genes led us to suspect that many of these distorter-restricted genes represented homologs of nondistorter counterparts. Since we observed distorter-restricted transmission of the maternal genome (for the *Phos1* gene), we reasoned that this pattern could be explained by a separate evolutionary trajectory for the distorter-restricted portion of genome and that nondistorter (i.e. paternal) and distorter (i.e. maternal/female-limited) components of the distorter genome have diverged over time to the extent that at least some regions assemble separately; our further analyses prove consistent with this hypothesis.

We first examined kmer frequencies for a subset of reads from distorter and nondistorter females and found stark differences, with a single peak in the nondistorter female, but multiple peaks in the distorter, showing a larger apparent genome size with a large portion of genomic content at ~½ coverage (Figure S1). We likewise found pronounced differences in read coverage across genes in $\log_2$ distorter/nondistorter coverage, with a diffuse peak highly enriched in the distorter, and clear peaks at $\log_2(D_{cov}/N_{cov})$ of ~ -1 and 0 (Figure 2A). We infer that these three groupings represent distorter-restricted alleles, nondistorter alleles (that are homologous to distorter-restricted alleles), and alleles that are present on homologous chromosomes in both distorter and nondistorter females (and similar enough to assemble and map as the same gene in distorters), respectively. We expect that alleles present at similar coverage in both distorters and nondistorters are 'conserved' genes for which purifying selection has prevented divergence of the distorter allele. Supporting this, we found that 'conserved' genes frequently co-occur on contigs

with nondistorter or distorter genes (Figure S2), confirming they are not confined to a distinct region of the genome. Conversely, distorter and nondistorter genes, as expected, very rarely co-occur; we expect that these co-occurrences represent infrequent misassemblies (Figure S2).

BLAST searches of predicted coding sequences support this interpretation. Strikingly, both all-by-all reciprocal tBLASTx searches including all genes, as well as homology searches after splitting genes by potential nondistorter or distorter allele (determined by $\log_2(D_{cov}/N_{cov})$ < or > 0, respectively) yielded pervasive matches. The latter, for instance, revealed ~2900 gene pairs encompassing >40% of overall predicted gene content (Figure 2B), with a median predicted amino acid similarity of 93.7% (Figure 2C).

We also explored broad expression patterns, although the vastly different gene content of nondistorter/distorter genomes precluded formal differential expression analysis. Interestingly, expression strongly tracked copy number, as the relative expression ($\log_2 D_{expr}/N_{expr}$) of predicted nondistorter alleles ($\log_2(D_{cov}/N_{cov}) < -0.5$) in the distorter genome was almost precisely ½ that of conserved alleles ($-0.5 < \log_2(D_{cov}/N_{cov}) < 1$; Figure 2D; $\log_2$ difference = -0.96, $P < 0.001$), showing an apparent lack of dosage compensation favouring nondistorter alleles in the distorter's genome.

The distorter-restricted genome behaves like a non-recombining sex chromosome

The inheritance of sex-restricted chromosomes such as the male Y chromosome leads to weakened purifying selection, the accumulation of deleterious mutations, and gene loss [3]. The mode of inheritance and pattern of sequence divergence of distorter-restricted genes suggest that the distorter-restricted genome is in a similar state of deterioration. Consistent with this expectation, although we often observed clear synteny in contig structure, distorter-restricted contigs exhibit divergence from their nondistorter homologs in the form of inversions, indels, and SNPs (Figures 3A and 3B). We calculated dN/dS for distorter and nondistorter allele sets relative to ORFs predicted from *Liposcelis entomophila*, a related sexual species with a publicly available transcriptome [4].

Distorter-restricted alleles had significantly higher dN/dS ratios relative to *L. entomophila* than their nondistorter counterparts (Figure 3C; N = 1091; Wilcoxon signed rank test, $P < 0.001$; mean dN/dS: 0.105 vs. 0.095), suggesting relaxed selection. Distorter ORFs were also significantly shorter than their nondistorter counterparts (Figure 3D; Wilcoxon signed rank test, $P < 0.001$). Furthermore, genes binned as 'conserved' genes, had significantly lower dN/dS (mean = 0.088) than either the nondistorter or distorter allele sets (Mann-Whitney test; N = 3396 with detected orthologs; $P < 0.01$ and $P < 10^{-5}$, respectively), relative to *L. entomophila*, consistent with our hypothesis of generally stronger purifying selection acting at these loci.

Finally, to estimate the age of this peculiar sex-ratio distortion, we used human and chimpanzee lice (*Pediculus humanus* and *P. schaeffi*) sequence [5], as these are relatively closely related to *Liposcelis* [6], and have a known divergence time (i.e. that of their hosts). We identified orthologs in *P. humanus* and *P. schaeffi*, with corresponding distorter and nondistorter *Liposcelis* sequences, resulting in a set of 1008 genes/alleles for analysis. Computing dS yielded an average rate of divergence between *Liposcelis* alleles at ~8 % of that of the *Pediculus* species. Making simplifying assumptions (constant dS across lice, divergence of *Pediculus* with their hosts ~6-10 mya [7]) we estimate the age of the distorter-nondistorter split to be ~450-820,000 years (Figure 3E).

A horizontally transferred Wolbachia gene is a candidate for sex ratio distortion

Some maternally-inherited bacteria, such as *Wolbachia* and *Cardinium,* distort host sex ratios towards females to favour their own transmission [8, 9]. As validation of our earlier findings that microbes are not involved in this distortion [1] we found no evidence of a symbiont that might contribute to distortion, searching predicted protein sequences against NCBI databases. Intriguingly, we did find a total of 103 genes that appeared to be of bacterial origin and that might represent contamination or HGT events, which we examined in detail. We were particularly interested in genes that were restricted to distorters. Of the 103 genes, 21 were most closely related to genes from *Wolbachia*, *Rickettsia*, or *Cardinium*. Of these, we were able to recover 9 from *L. entomophila* and *L.*

bostrychophila transcriptomes, suggesting they were acquired early in *Liposcelis* evolution, and thus unlikely to be causes of distortion. Supporting this, a number were clearly distorter/nondistorter pairs in our assembly.

Of the remaining genes, three related sequences appear to have been acquired at one time from *Wolbachia* and expanded in the distorter genome, and represent intriguing candidates for involvement in sex ratio distortion. These sequences are encoded in large contigs (~70-100 kb), are clearly expressed in the polyA-primed mRNA sequencing, and contain putative introns. The function of these genes and their *Wolbachia* orthologs is not known, although they contain predicted NB-ARC and tetratricopeptide repeat domains (Figure S3). We also found that this gene was independently acquired by at least three other insect species (Figure 4), including two with unusual modes of reproduction. We name this gene *Odile*, for 'Only Daughters in *Liposcelis*-associated Element'. Odile is also the name of the Black Swan from Tchaikovsky's ballet, Swan Lake; she tries to steal the Prince from her almost-twin, the White Swan.

Although several lines of evidence make the *Odile* gene an intriguing candidate for distortion, we have not yet demonstrated a causal role. It is also possible that distortion is caused by other genes that are either unique to the distorter or that are divergent distorter-restricted alleles; we identified a handful of these types of genes that appear to have a putative connection to mitosis or meiosis (see Table S4) and that warrant further study. It is also possible that these genes have evolved as a consequence of relaxed selection for meiotic or male reproductive function.

Discussion

A major question in evolutionary biology is why sex determination systems evolve so rapidly [10, 11]. One of the main hypotheses to explain the dynamic turnover of sex determination systems and sex chromosomes is that it is shaped by conflicts over transmission. Selfish genetic elements that distort sex ratios have been described across a

wide range of organisms, including rodents, insects, and flowering plants [12, 13]. Coevolutionary arms races between these elements and the rest of the genome are thought to play a major role in driving transitions in sex determination [11, 14, 15]. Indeed, the study of sex ratio distortion has uncovered some of the most iconic selfish genetic elements, such as PSR, a selfish chromosome that distorts sex ratios in haplodiploid wasps by destroying all of the chromosomes that were inherited alongside it, so that females that fertilize eggs with the sperm of PSR males produce PSR sons anew [16]. Currently, there is great interest in using natural distorting systems, and learning from them, to control pests and disease vectors [17].

Here we report the discovery of an unusual mode of sex determination and selfish sex ratio distortion in *Liposcelis* booklice with accompanying profound genomic consequences. While nondistorter females from an as yet unnamed species from Arizona produce both sons and daughters, distorter females never produce sons, although they must mate with males (i.e. the sons of nondistorters) in order to reproduce [1]. The genes that distorter females inherit from their fathers are expressed in their somatic tissues, but are not transmitted to the next generation. Because they never transmit paternal DNA to the next generation, distorters are essentially parasitizing males.

We are not aware of another example of this kind of sex determination, although it is perhaps most similar to hybridogenesis. In this mode of sex determination, which has been best studied in amphibians, hybridogenetic lineages typically arise from hybridization between two distinct species, and mate every generation with individuals from one of the parent species, whose genome is excluded from the hybridogen's germline [13, 18]. Unlike hybridogens, distortion in *Liposcelis* does not involve hybridization between different species, nor does it involve polyploidy. The *Liposcelis* distorter genome evolved from its nondistorting counterpart, and our analysis of sequence divergence suggests that this happened fairly recently, approximately 450-820,000 years ago.

So how did distortion evolve in *Liposcelis*? In booklice and their relatives, including parasitic lice, the baseline mode of sex determination is paternal genome elimination (PGE) [2, 19], which has evolved independently in a number of terrestrial microarthropods [20]. Booklice that will develop into males fail to transmit any chromosomes that they inherit from their father. Distorter females may have thus hijacked PGE, and by transmitting only their maternally inherited chromosomes, are behaving like feminized males. Interestingly, one of the main hypotheses as to why PGE evolved in the first place involves transmission distortion [21]. In this hypothesis, PGE arose when an XX/XO sex determination system was invaded by a meiotic drive X chromosome and the autosomes evolved to hitchhike alongside it.

Distortion in *Liposcelis* has had dramatic consequences for genome structure and evolution. Distorters have chimeric genomes – half of their genome comes from their nondistorter fathers, while the other half is restricted to distorter females. This distorter-specific part of the genome shows the hallmarks of one giant nonrecombining chromosome, with altered selection on distorter-restricted alleles. Interestingly, distorters live much shorter than their nondistorter counterparts in the lab [1], which helps explain how the distorter polymorphism may persist in nature; without fitness costs, we might expect the distorter to reach high frequencies and cause population extinctions.

Comparing distorter and nondistorter genomes yielded an intriguing candidate for the genetic basis of distortion, a gene of unknown function that has been acquired from *Wolbachia*, a bacterial symbiont of arthropods well known for its ability to manipulate reproduction and chromosome transmission [22, 23], although a causal role for this gene has not yet been demonstrated. Indeed, without functional experiments we also cannot rule out the possibility that distortion is caused by neo-functionalization of a divergent distorter-restricted gene.

We were surprised to find that our candidate *Wolbachia* gene has independently been integrated into at least four insect genomes, three of which, including the *Liposcelis* sp. distorter, have unusual modes of reproduction. This gene is also integrated in the genome

of *Liposcelis bostrychophila*, a parthenogenetic stored grain pest. The two *Liposcelis* species are not closely related [24], and it is likely that each acquired the *Wolbachia* gene independently; also, in addition to being absent from the nondistorter, we did not detect it in the transcriptome of the sexual *L. entomophila*. This gene is also integrated in the genome of the little fire ant *Wasmannia auropunctata,* which has an odd mode of reproduction [25], with both queens and males reproducing clonally, via an unusual form of maternal genome elimination. Worker ants are produced as a result of sexual reproduction between queens and males, but because workers are sterile, the genomes of queens and males are diverging, a situation reminiscent of distortion in *Liposcelis*. It was likewise recently shown that the horizontal transfer of an entire *Wolbachia* genome into the pillbug *Armadillidium vulgare* has resulted in the evolution of a new sex chromosome [26]. Genes acquired from *Wolbachia* and other facultative inherited symbionts [27], which are widespread in arthropods [28, 29], may be an important generator of animal reproductive diversity.

Acknowledgements:

This work was funded by an NSERC Discovery Grant. SJP also acknowledges support from the Canadian Institute for Advanced Research's Integrated Microbial Biodiversity Program. We thank Compute Canada for access to computational resources that enabled this study, and Jong Leong and David Minkley for method discussion and advice. PTH is supported by a CIHR postdoctoral fellowship.

Author Contributions:

PTH, CNH, and CIC performed experiments; all authors analyzed the data; PTH and SJP wrote the paper, with comments and input from CNH and CIC.

Declaration of Interests:

The authors declare no competing interests.

STAR Methods

1 Contact for reagents and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the senior author, Steve Perlman (stevep@uvic.ca).

2 Experimental models and subject details

Colony Information

Liposcelis sp. was initially collected from the Chiricahua Mountains, Arizona, in 2010 [1]. Individuals from our lab culture have been deposited in the insect collection at the Royal British Columbia Museum, Victoria, BC, while this species awaits formal description. We keep *Liposcelis* sp. colonies in small glass canning jars (125ml) with the lid replaced with 70mm Whatman filter paper (Sigma-Aldrich) and rear them on a diet of 1:10 (weight:weight) mixture of Rice Krispies (Kellogg's) to cracked red wheat (Planet Organic). Colonies are maintained at 27°C and 75% relative humidity. We check the colonies every second week and replace food with new food as needed. We keep colonies containing distorter females separate from those containing nondistorter females, adding males to distorter female colonies every second week.

3 Methods details

Distorter Inheritance Experiment

We conducted a two-generation crossing experiment, tracking the cAMP-specific IBMX-insensitive 3',5'-cyclic phosphodiesterase gene (*Phos1* for short) in the distorter lineage of our lab culture of *Liposcelis* sp. This marker was previously used in inheritance experiments to determine that males of this species exhibit paternal genome elimination [2]. By tracking the *Phos1* marker over two generations, we were able to test for departures from Mendelian inheritance in the distorter female lineage.

Experiments were run in 35mm petri dishes containing 0.5g of booklouse food. For the first generation crosses, we paired 20 distorter females with a male from the nondistorter colony. We left them for approximately three weeks, transferring them into a new dish once in this period, after which we took the male out for DNA extraction and left the female to lay eggs for another 3 weeks (again transferring her to a new dish once in this period), before removing her for DNA extraction. We monitored the containers containing F1 offspring three times a week. Once individuals matured into adult females, we moved them to their own petri dish and paired them with a male from the colony. We carried out the same procedure outlined for the first generation above and left the F2 offspring for at least a week to develop before terminating the experiment and extracting their DNA. We extracted DNA from 10 to 15 offspring from both the first and second generations of the crosses. We sequenced the DNA at Sequetech (California, USA). Since males only transmit the allele they inherit from their mother [2] we analyzed the data with the expectation that heterozygous males would transmit the same allele to all their offspring (i.e. exhibit the same transmission dynamics as homozygous males).

DNA and RNA extraction and high throughput sequencing

DNA for genome sequencing was isolated from pooled individuals derived from laboratory lines of *Liposcelis* sp. Depending on sequencing technology either ~80 individuals (Illumina) or ~300 individuals (PacBio; ~150 individuals/extraction) were extracted using a Qiagen DNeasy Kit with an additional 1 sec bead beating step (BioSpec Beadbeater 16, 3.5mm glass bead) during the lysis stage to increase yield. DNA samples were purified using AmpPureXP and eluted in Qiagen buffer EB before library construction and sequencing by Genome Quebec. Illumina Truseq libraries were constructed for each of the nondistorter female, male, and distorter genotypes and sequenced in one lane of Illumina Hiseq 2500 with 100 bp PE reads. An additional library was constructed for PacBio sequencing, with 5 smart cells sequenced, and this sequencing was combined with existing Illumina sequence [1] for assembly.

We assembled a draft genome using DBG2OLC [30], which uses an input de Bruijn graph assembly (based on Illumina reads and generated by Ray v. 2.2.0; k = 31 [31]) and PacBio reads to assemble larger genomes with lower PacBio coverage. Contig consensus sequences were called using the DBG2OLC sparc module, and polished to improve error-rates using two iterations of Pilon [32] following mapping of Illumina short-reads to the assembly using bwa mem [33].

We generated a transcriptome assembly by combining reads from distorter and nondistorter females to examine transcribed genes and gene expression levels across the two female types. To prevent mating, females were collected in their last instar stage and reared until they were 1-7 day old adults, at which point they were pooled in samples of 20 individuals for replicate RNA extractions (3 replicates per female type; N = 6). Total RNA extractions were done in 300 uL TriZol reagent as per the manufacturer's instructions, including an initial bead-beating step for 6 seconds during the lysis stage. Samples were purified using AmpPure XP, flash frozen, and provided to Genome Quebec for QC (Agilent BioAnalyzer), library construction (TruSeq mRNA stranded), and sequencing (6 libraries in one lane of Illumina Hiseq 2500, 125 bp PE reads).

Raw reads were assembled into a transcriptome of combined female types by merging a Trinity *de novo* (default parameters; reads cleaned by Trimmomatic with *in silico* read normalization [34]) with a Trinity genome guided assembly based on the above draft genome assembly using Hisat2 [35] for read alignment (default parameters). Assemblies were merged using the tr2aacds.pl script of Evidential Gene and passed to Maker2 for gene annotation/prediction [36]. Kmer frequency plots were generated by kmer counting on a subset (40 M) of raw reads using jellyfish [37].

We estimated genome, transcriptome, and gene prediction completeness using BUSCO [38], against the insect ortholog set. This yielded completeness estimates of 94.0%, 96.4% and 76.4% for the genome, transcriptome (Evidential Gene merged transcript set) and gene models (Maker-predicted), respectively.

Read abundance estimation

To infer gene copy number (for DNA) and the expression level (for RNA) for assembled genes across the two female types, we mapped raw reads to the draft genome using bwa-mem (for DNA reads; mapq > 30) or Hisat2 (for gapped RNA-seq reads; default parameters). Mappings to exons were quantified as fragment counts using the featureCounts utility of subRead [39] (unstranded library type for DNA; reverse stranded for RNA) and subsequently as FPKM using *edgeR* [40].

For assessing the variance in coverage across genes in a contig, we filtered the assembly contig set to retain those encoding more than five predicted genes.

Ortholog detection, selection analysis, and dating distorter

All-by-all BLASTs to identify gene pairs in transcript sets were done using tBLASTx (evalue < 10^{-5}), with hits-to-self filtered, and only alignments > 100 bp and spanning > 60% of the query length accepted; the single best non-self match for each query was retained. This allowed some genes to participate in multiple gene-pairs as reciprocal best hits were not enforced, although this was uncommon. This approach yielded ~2700 gene pairs encompassing ~5400 genes. We also performed reciprocal-best-hit searches between inferred potential distorter and nondistorter alleles ($\log_2 D_{cov}/N_{cov} < 0$ or > 0, respectively); this binning threshold was chosen to be inclusive (i.e. overlapping predicted 'conserved' gene coverage (below)) to prevent drop-out of more conserved allele pairs, although this approach could detect paralogs among 'conserved' genes as well. Inspection of pairings showed the vast majority however, represented likely nondistorter-distorter pairings.

dN/dS analysis analysis used *ortholgr* [41], an R package that wrapped reciprocal-best-hit identification (BLAST), codon-based alignment (clustalw), and dN/dS inference, using the model of Yang and Nielsen [42]. For analysis of dN/dS putatively 'conserved' genes were defined as having $\log_2 D_{cov}/N_{cov} > -0.5$ and < 1 and not having an identified reciprocal best hit (i.e. which would characterize them as a nondistorter-distorter pair in

the above analysis). As input to ortholog detection, we first trimmed transcript sequences to the longest predicted ORF per transcript to remove non-coding sequence (or trimmed genomic sequence to predicted ORFs; start to stop codon) using EMBOSS (ORFs > 300 nt) [43].

We defined gene sets orthologous between *P. humanus*, *P. schaeffi*, and the distorter and nondistorter *Liposcelis* sp. similarly, based on the publicly available *P. humanus* (PhumU2.2) transcript set and a reassembly of *P. schaeffi* genomic sequence (from PRJNA230884) using Ray. We binned *Liposcelis* alleles as distorter or nondistorter based on log coverage ratios as above prior to ortholog detection, and extracted dS from homologous gene pairs as a measure of divergence. We generated a phylogeny to display these relationships by concatenating the longest 25 codon-based alignments (using MAFFT) of the ortholog sets, manually trimming regions with large gaps in Geneious. We used Fasttree 2.1.18 with the GTR model and optimized gamma likelihood to construct a maximum-likelihood phylogeny for display [44]. *ggtree* was used to visualize the resulting phylogeny [45].

HGT screens

To screen for contamination/HGT events we used BLASTp, searching predicted proteins against the non-redundant (nr) NCBI database (evalue < 10^{-5}).

Phylogenetics and sequence analysis of a horizontally-transferred Wolbachia gene

For comparison of specific genes from *Liposcelis sp*. with sister lineages – specifically a putative HGT event from *Wolbachia* to booklice – raw Illumina RNA read sets for asexual *L. bostrychophila* (PRJNA188391), a parthenogenetic species, and *Liposcelis entomophila* (PRJNA214735), a sexual species, were downloaded from the NCBI and assembled using Trinity (default *de novo* parameters) and Ray (k = 31); we used Ray-assembled sequences for phylogenies, as we found our gene of interest to assemble more completely with Ray.

BLASTp searches using the *Liposcelis* sp. gene of *Wolbachia* origin against the nr database recovered predicted proteins from *Wolbachia*, as well as two other predicted proteins from insects that are not likely to be bacterial contamination as they both contain an intron and are encoded on large contigs (128 kb and 5.9 Mb). *L. bostrychophila* contained homologous sequences of the putative HGT event from *Wolbachia* while none were found in *L. entomophila*. An alignment of amino acid sequences was generated using Geneious alignment with automated parameterization in Geneious 7. We performed alignment character trimming with BMGE 1.12 [46] weighted with the BLOSUM35 similarity matrix and used Hyphy [46] to estimate the optimal model of amino acid substitution (WAG). Phylogenies were generated with PhyML 3.0 [48] in SeaView 4.6.2, [49], bootstrapped with 1000 replicates and visualized in Figtree.

We identified four paralogs of the candidate *Wolbachia* gene in the distorter booklouse transcriptome. Three are on a single 77757 bp contig; there are no other genes on this contig. The fourth paralog (*Odile4*) lies on a 91534 bp contig that contains an insect gene (ortholog of *Pediculus humanus* ser/thr protein kinase-trb3, putative); these two genes are ~30kb apart. We identified putative introns based on transcript alignments to genomic contigs and the presence of donor and acceptor splice-sites in three of the four *Wolbachia* gene paralogs (Figure S3). We searched for conserved domains with CDD [50] and found that the *Wolbachia* homologs contained ankyrin repeat (ANK) domains at the 5' end as well as NB-ARC domains and tetratricopeptide repeat domains (TPRs). Both ANK and TPR repeat regions mediate protein-protein interactions, while NB-ARC is thought to bind and hydrolyse ATP ([51]). None of the *Wolbachia* genes in the distorter booklouse, asexual booklouse or the two other insects contained the ANK domains, but did contain both NB-ARC and TPRs (Figure S3). The nature of the tetratricopeptide repeat regions along with the paralogous gene copies makes Sanger sequencing difficult, but we have been able to confirm the main *Wolbachia* candidate gene sequence (*Odile1*, Figure S3) with Sanger sequencing.

Additional screens for distorter-restricted distortion candidates

We identified distorter-unique genes based on the criteria of expression in the distorter (mean FPKM > 2.5), not in the nondistorter (mean FPKM < 1), and few DNA reads mapping from nondistorter (mean DNA FPKM in nondistorter female and male < 5) (yielding N=1297 candidates). We filtered any of these that had a pairing from the ortholog detection (above; N=290 remaining) and conducted a secondary tBLASTx search against the more expansive full Evidential Gene transcript set with the remaining genes to further remove any potential homologs. This yielded 39 candidate genes, including *Odile*, only one of which appeared to have function associated with meiosis or mitosis, an ortholog of *geminin,* an inhibitor of DNA replication and a cell cycle regulator [52] (Table S4A).

To examine highly divergent gene pairs, we examined the most divergent nondistorter-distorter allele pairings with the highest relative expression in the distorter (as $\log2(D_{expr}/N_{expr})$). We selected the 100 most divergent pairings from the top 10% of distorter-expressed for manual curation; of these we recovered 6 that have putative function in mitosis or meiosis, presented in Table S4B.

4 Quantification and statistical analyses

Statistical analyses were conducted using R/Bioconductor v.3.4.1. For quantifying departures from Mendelian inheritance we used Chi square tests on each cross. We analyzed genome coverage and gene expression using linear models, where appropriate, after transforming raw counts as $\log_2(FPKM + 1)$.

5 Data and software availability

Raw sequence data are deposited at NCBI (Bioproject PRJNA355858 and Genbank accessions MH764403 for the *Phos1* distorter allele and MH751905 for *Odile1* sequence confirmed by Sanger). Compiled data to reproduce analyses are available at the Dryad data repository (Need to get this accession).

References

1. Perlman, S.J., Hodson, C.N., Hamilton, P.T., Opit, G.P., and Gowen, B.E. (2015). Maternal transmission, sex ratio distortion, and mitochondria. Proc. Natl. Acad. Sci. USA *112*, 10162–10168.

2. Hodson, C.N., Hamilton, P.T., Dilworth, D., Curtis, C.I., and Perlman, S.J. (2017). Paternal genome elimination in *Liposcelis* booklice (Insecta: Psocodea). Genetics *206*, 1091-1100.

3. Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. *14*, 113-124.

4. Wei, D.D., Chen, E.H., Ding, T.B., Chen, S.C., Dou, W., and Wang, J.J. (2013). *De novo* assembly, gene annotation, and marker discovery in stored-product pest *Liposcelis entomophila* (Enderlein) using transcriptome sequences. PLoS ONE *8*, e80046.

5. Johnson, K.P., Allen, J.M., Olds, B.P., Mugisha, L., Reed, D.L., Paige, K.N., and Pittendrigh B.R. (2014). Rates of genomic divergence in humans, chimpanzees and their lice. Proc. Biol. Sci. *281*, 20132174.

6. Yoshizawa, K., and Johnson, K.P. (2003). Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. Mol. Phylogen. Evol. *29*, 102-114.

7. Moorjani, P., Amorim, C.E.G., Arndt, P.F., and Przeworski, M. (2016). Variation in the molecular clock of primates. Proc. Nat. Acad. Sci. USA *113*, 10607-10612.

8. Werren, J.H., and O'Neill, S.L. (1997). The evolution of heritable symbionts. In Influential Passengers: Inherited Microorganisms and Arthropod Reproduction, S.L. O'Neill, A.A. Hoffmann, and J.H. Werren, eds. (Oxford University Press), pp.1-42.

9. Engelstädter, J., and Hurst, G.D.D. (2009). The ecology and evolution of microbes that manipulate host reproduction. Annu. Rev. Ecol. Evol. Syst. *40*, 127-149.

10. Bachtrog, D., Mank, J. E., Peichel, C.L., Kirkpatrick, M., Otto, S.P. *et al.,* (2014). Sex determination: why so many ways of doing it? PLoS Biol. *12*, e1001899.

11. Beukeboom, L.W., and Perrin, N. (2014). The Evolution of Sex Determination. (Oxford University Press).

12. Hurst, L.D. (1993). The incidences, mechanisms and evolution of cytoplasmic sex ratio distorters in animals. Biol. Rev. *68*, 121-194.

13. Burt, A., and Trivers, R. (2006). Genes in Conflict: the Biology of Selfish Genetic Elements. (Harvard University Press).

14. Kozielska, M., Weissing, F.J., Beukeboom, L.W., and Pen, I. (2010). Segregation distortion and the evolution of sex-determining mechanisms. Heredity *104*, 100–112.

15. Werren, J.H., and Beukeboom, L.W. (1998). Sex determination, sex ratios, and genetic conflict. Annu. Rev. Ecol. Evol. System. *29*, 233-261.

16. Nur, U., Werren, J.H., Eickbush, D.G., Burke, W.D., and Eickbush, T.H. (1988). A "selfish" B chromosome that enhances its transmission by eliminating the paternal genome. Science *240*, 512-514.

17. Bull, J.J., and Malik, H.S. (2017). The gene drive bubble: new realities. PLoS Genet. 13: e1006850.

18. Stöck, M., Ustinova, J., Betto-Colliard, C., Schartl, M., Moritz, C., and Perrin, N. (2012). Simultaneous Mendelian and clonal genome transmission in a sexually reproducing, all-triploid vertebrate. Proc. Biol. Sci. *279*, 1293-1299.

19. de la Filia, A.G., Andrewes, S., Clark, J.M., and Ross, L. (2018). The unusual reproductive system of head and body lice (*Pediculus humanus*). Med. Vet. Entomol. *32*, 226-234.

20. Gardner, A., and Ross, L. (2014). Mating ecology explains patterns of genome elimination. Ecol. Lett. *17*, 1602–1612.

21. Haig, D. (1993). The evolution of unusual chromosomal systems in coccoids: extraordinary sex ratios revisited. J. Evol. Biol. *6*, 69–77.

22. Werren, J.H., Baldo, L., and Clark, M.E. (2008). *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbiol. *6*, 741-751.

23. Stouthamer, R., and Kazmer, D.J. (1994). Cytogenetics of microbe-associated parthenogenesis and its consequences for gene flow in *Trichogramma* wasps. Heredity *73*, 317-327.

24. Feng, S.Q., Yang, Q.Q., Li, H., Song, F., Stejskal, V., Opit, G.P., Cai, W.Z., Li, Z.H., and Shao, R.F. (2018). The highly divergent mitochondrial genomes indicate that the booklouse, *Liposcelis bostrychophila* (Psocoptera: Liposcelididae) is a cryptic species. G3-Genes Genomes Genetics *8*, 1039-1047.

25. Fournier, D., Estoup, A., Orivel, J., Foucaud, J., Jourdan, H., Le Breton, J., and Keller, L. (2005). Clonal reproduction by males and females in the little fire ant. Nature *435*, 1230-1234.

26. Leclercq, S., Theze, J., Chebbi, M.A., Giraud, I., Moumen, B., Ernenwein, L., Greve, P., Gilbert, C., and Cordaux, R. (2016). Birth of a W sex chromosome by horizontal

transfer of *Wolbachia* bacterial symbiont genome. Proc. Natl. Acad. Sci. USA *113*, 15036-15041.

27. Hotopp, J.C.D., Clark, M.E., Oliveira, D.C.S.G., Foster, J.M., Fischer, P., Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S.L., Ingram, J., Nene, R.V., Shepard, J., Tomkins, J., Richards, S., Spiro, D.J., Ghedin, E., Slatko, B.E., Tettelin, H., and Werren, J.H. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science *317*, 1753-1756.

28. Zug, R., and Hammerstein, P. (2012). Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. PLoS ONE *7*, e38544.

29. Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L.Q., Engelstadter, J., and Hurst, G.D.D. (2008). The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. BMC Biol. *6*, 27.

30. Ye, C., Hill, C.M.,Wu, S., Ruan, J., and Ma, S. (2016). DBG2OLC: Efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci. Rep. *6*, 31900.

31. Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., and Corbeil, J. (2012). Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biol. *13*, R122.

32. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE *9*, e112963.

33. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1303.3997.

34. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. *15*, 644-52.

35. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods. *12*, 357–360.

36. Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491.

37. Marcais G., and Kingsford C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6): 764-770.

38. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov. E.V. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, *31*, 3210-3212.

39. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

40. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

41. Drost, H.-G., Gabel, A., Grosse, I., and Quint, M. (2015). Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. Mol. Biol. Evol. *32*, 1221-1231.

42. Yang, Z.H., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. *17*, 32-43.

43. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: The European molecular biology open software suite. Trends Genet. *16*, 276-277.

44. Price, M.N., Dehal, P.S., and Arkin A.P. (2010). FastTree 2: Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE *5(3)*, e9490.

45. Yu G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol. 8*, 28-36.

46. Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. *10*, 210.

47. Pond, S.L.K., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. Bioinformatics *21*, 676-679.

48. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307-321.

49. Gouy, M. Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. *27*, 221-224.

50. Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Lewis, Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C., and Bryant, S.H. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. *39*, D225–D229.

51. Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., … Yeats, C. (2009). InterPro: the integrative protein signature database. Nucleic Acids Res. *37*, D211–D215.

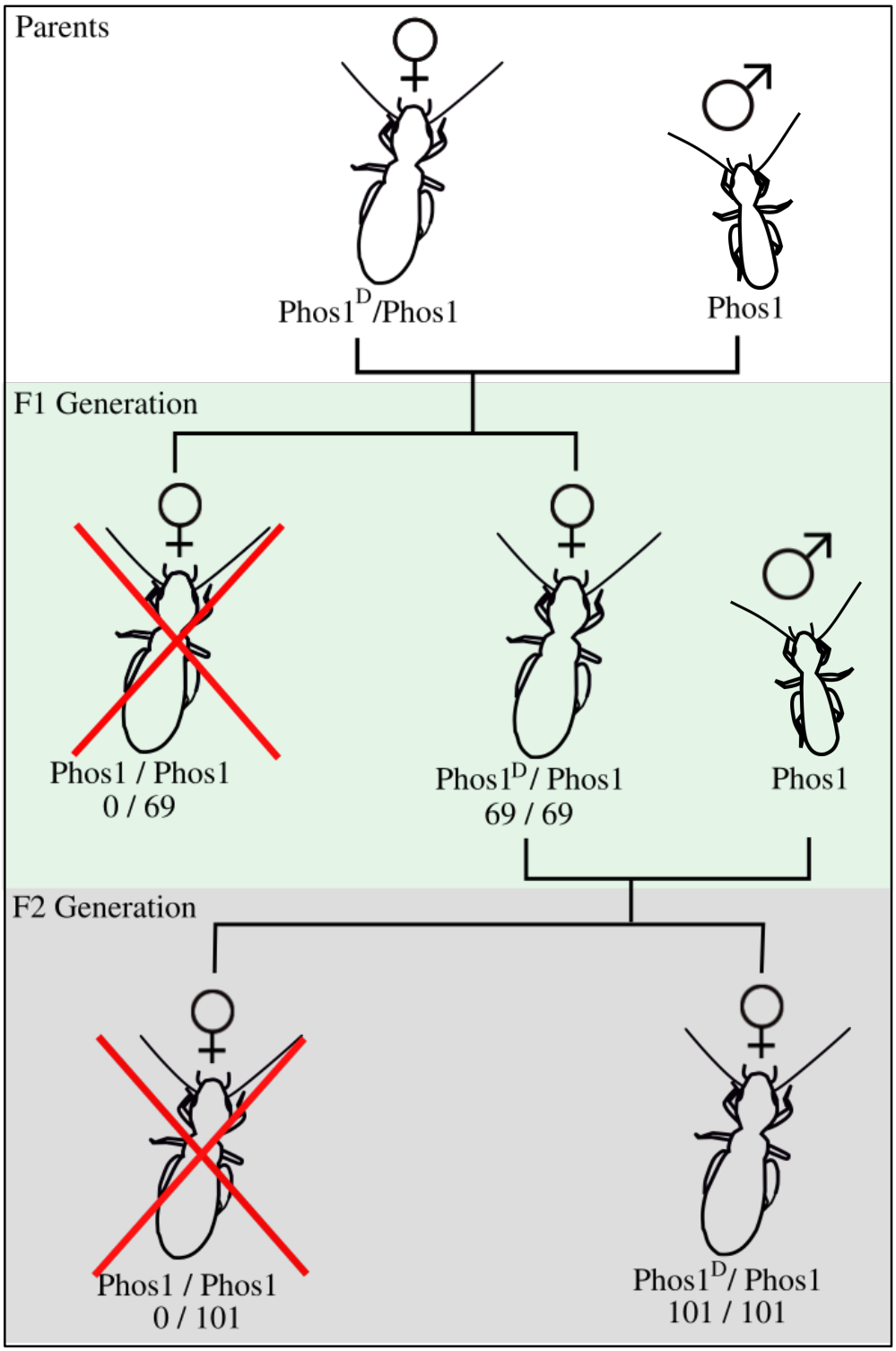
52. Quinn, L.M., Herr, A., McGarry, T.J., and Richardson, H. (2001) The Drosophila Geminin homolog: roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis. Genes Dev. *15*, 2741-2754.

Figure 1. Two-generation inheritance experiment showing that distorter *Liposcelis* sp. females always transmit the same phosphodiesterase allele (of maternal origin) to their offspring. The distorter female specific allele is labeled $Phos1^{D}$.

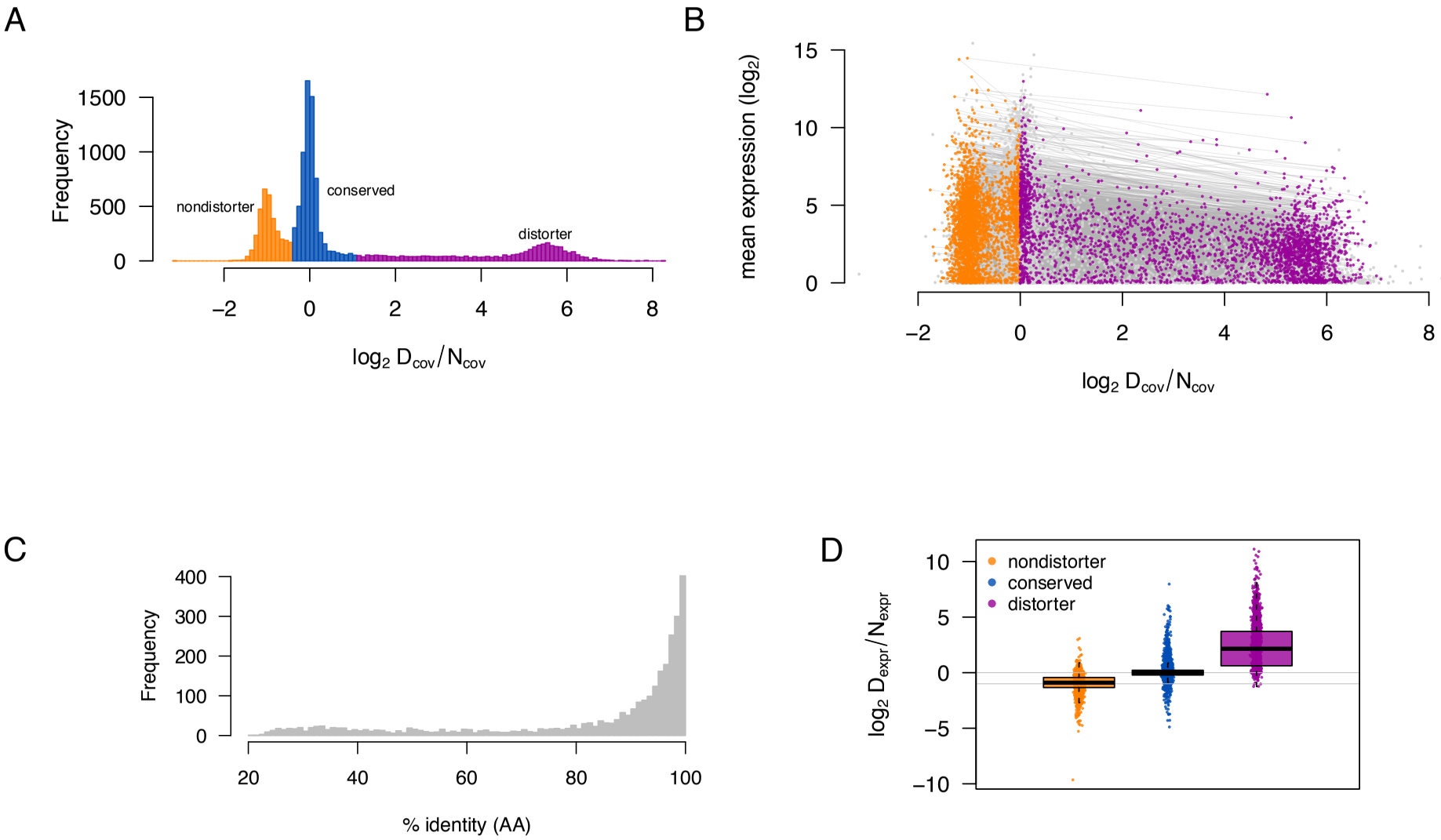
Figure 2. Distorter genomes are chimeras. **A)** Log coverage histogram of *Liposcelis* genes shows three classes of assembled gene, which we term 'nondistorter' (at half the coverage in distorter compared to nondistorter genome), 'conserved' (equal coverage) and 'distorter' (high coverage in distorter genome). **B)** *Liposcelis* gene models have best reciprocal BLAST hits (lines) in opposing gene classes, suggesting nondistorter and distorter genes are homologous and represent divergent alleles at a locus. **C)** Histogram of amino acid identity between homologous genes in distorters shows substantial divergence. **D)** mRNA expression ($\log_2(D_{expr}/N_{expr})$), shows nondistorter alleles have on average ~1/2 the relative expression of conserved alleles in distorters ($P < 0.001$).

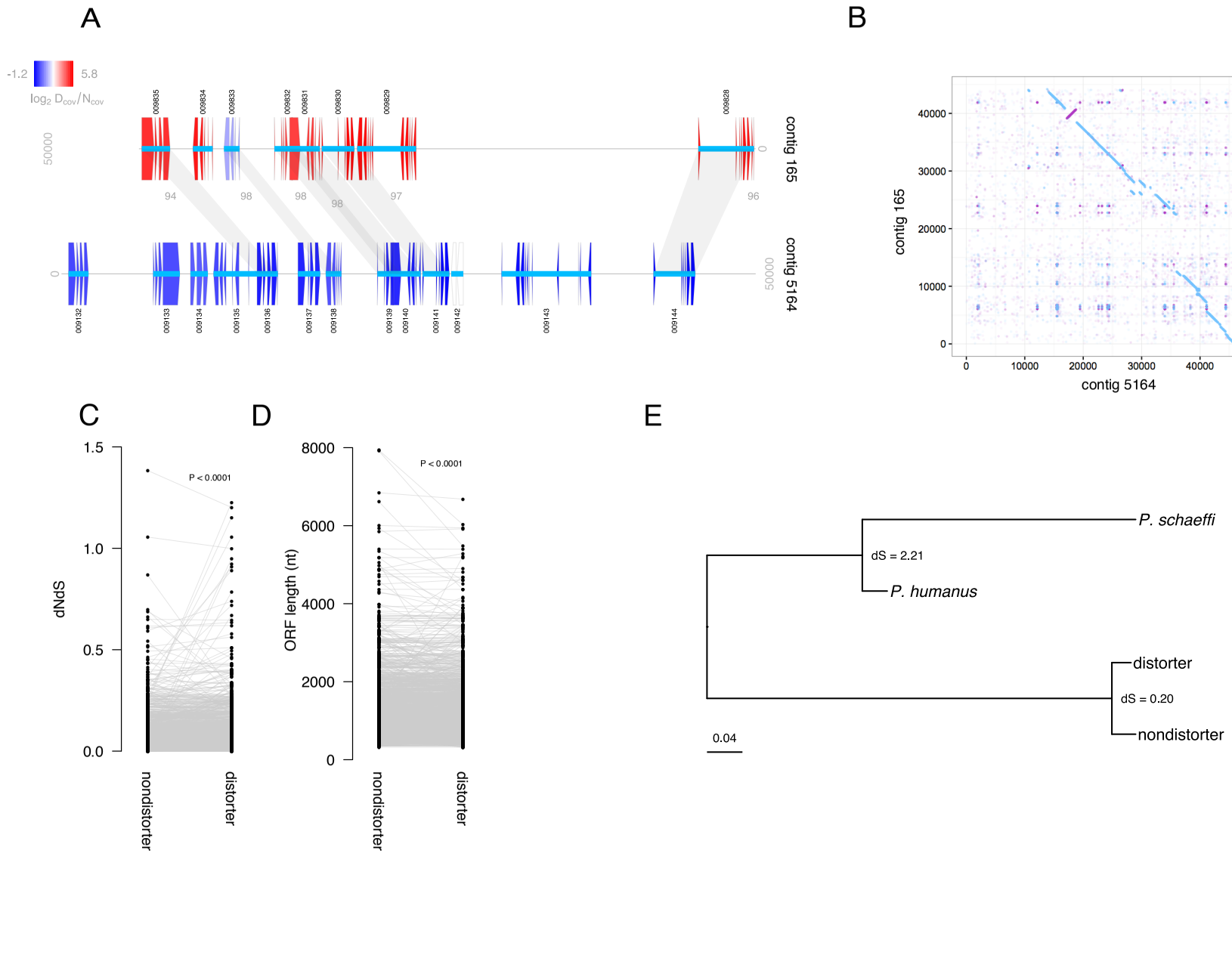
Figure 3. Distorter and nondistorter genes are largely syntenic. **A)** Representative contigs showing BLAST matches for predicted genes on two contigs, with percent identity shown for homologous genes ('alleles'; linked by gray polygons; gene colors represent $\log_2(D_{cov}/N_{cov})$). Note that stringent BLAST parameters excluded some potentially homologous gene pairs here. **B)** Dotplot of contigs shown in A) reveals divergent structure between syntenic contigs. **C)** dN/dS analysis of detected distorter and nondistorter homologous gene pairs shows elevated dN/dS in putative distorter-restricted alleles. ($P < 0.001$) **D)** distorter ORFs are shorter than corresponding nondistorter ORFs ($P < 0.001$). **E)** Synonymous mutation rates (dS) between orthologous gene sets in *P. humanus*, *P. schaeffi*, distorters, and nondistorters suggest that the nondistorter-distorter split is ~0.08 the age of the *P. humanus*-*P. schaeffi* split. Maximum-likelihood nucleotide phylogeny is based on 25 codon-aligned protein coding sequences; scale bar denotes substitutions per site.

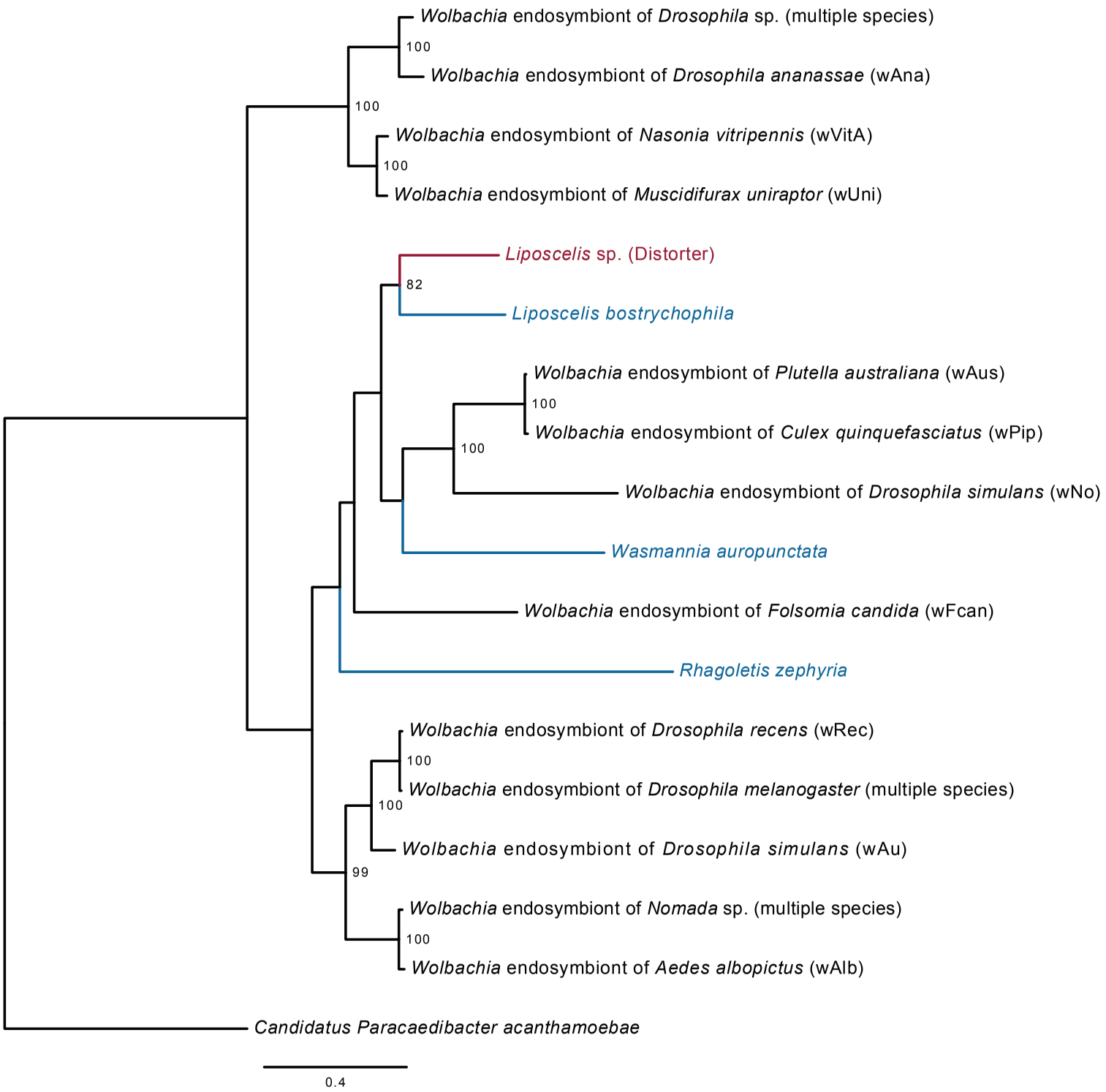
Figure 4. Maximum likelihood amino acid phylogeny of TPR-repeat containing protein (*Odile1*) from *Wolbachia* gene acquired by distorter females (red). Blue lines indicate other HGT events into arthropod genomes. Scale bar represents substitutions per site.


Parents
Phos1D/Phos1
Phos1
F1 Generation
Phos1 / Phos1
0 / 69
Phos1D/ Phos1
69 / 69
Phos1
F2 Generation
Phos1 / Phos1
0 / 101
Phos1D/ Phos1
101 / 101

729


A
Frequency
1500
1000
500
0
nondistorter
conserved
distorter
−2
0
2
4
6
8
log2 Dcov/Ncov
B
mean expression (log2)
15
10
5
0
−2
0
2
4
6
8
log2 Dcov/Ncov
C
Frequency
400
300
200
100
0
20
40
60
80
100
% identity (AA)
D
log2 Dexpr/Nexpr
10
5
0
−5
−10
nondistorter
conserved
distorter


A
-1.2
5.8
log2 Dcov/Ncov
50000
0
009835
009834
009833
009832
009831
009830
009829
009828
contig 165
contig 5164
94
98
98
98
97
96
009132
009133
009134
009135
009136
009137
009138
009139
009140
009141
009142
009143
009144
B
contig 165
contig 5164
0
10000
20000
30000
40000
50000
C
dNdS
P < 0.0001
nondistorter
distorter
D
ORF length (nt)
P < 0.0001
nondistorter
distorter
E
P. schaeffi
dS = 2.21
P. humanus
distorter
dS = 0.20
nondistorter
0.04


Wolbachia endosymbiont of *Drosophila* sp. (multiple species)
100
Wolbachia endosymbiont of *Drosophila ananassae* (wAna)
100
Wolbachia endosymbiont of *Nasonia vitripennis* (wVitA)
100
Wolbachia endosymbiont of *Muscidifurax uniraptor* (wUni)
Liposcelis sp. (Distorter)
82
Liposcelis bostrychophila
Wolbachia endosymbiont of *Plutella australiana* (wAus)
100
Wolbachia endosymbiont of *Culex quinquefasciatus* (wPip)
100
Wolbachia endosymbiont of *Drosophila simulans* (wNo)
Wasmannia auropunctata
Wolbachia endosymbiont of *Folsomia candida* (wFcan)
Rhagoletis zephyria
Wolbachia endosymbiont of *Drosophila recens* (wRec)
100
Wolbachia endosymbiont of *Drosophila melanogaster* (multiple species)
100
Wolbachia endosymbiont of *Drosophila simulans* (wAu)
99
Wolbachia endosymbiont of *Nomada* sp. (multiple species)
100
Wolbachia endosymbiont of *Aedes albopictus* (wAlb)
Candidatus Paracaedibacter acanthamoebae
0.4

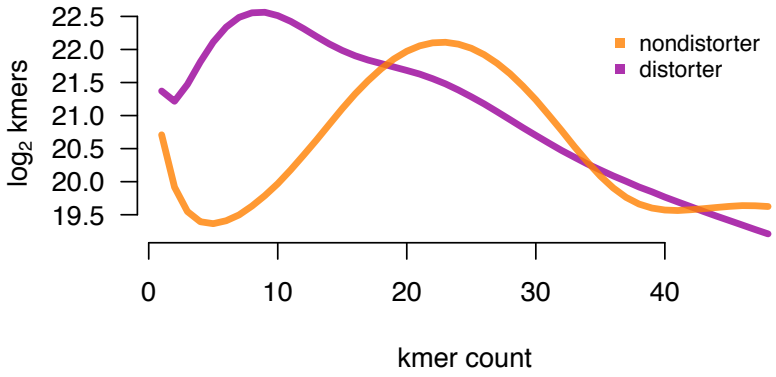


Figure S1. Kmer frequency plots of nondistorter and distorter genomes based on 40M 100bp PE reads (k = 15). Related to Figure 2.

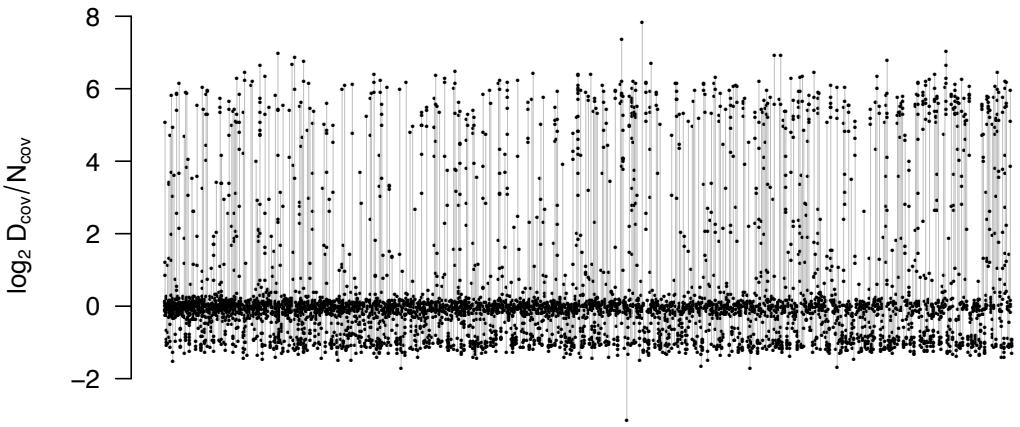


Figure S2. Coverage analysis of predicted genes on *Liposcelis* sp. assembly contigs. Related to Figure 2. Those encoding 5 or more predicted genes shown; X axis order arbitrary. Each gene is represented by a point, and each line links genes that are found on the same contig. This shows high variance in coverage among genes co-occurring on a contig and that 'conserved' genes (log coverage ratio ~ 0) are often found on the same contig as either distorter or nondistorter alleles.

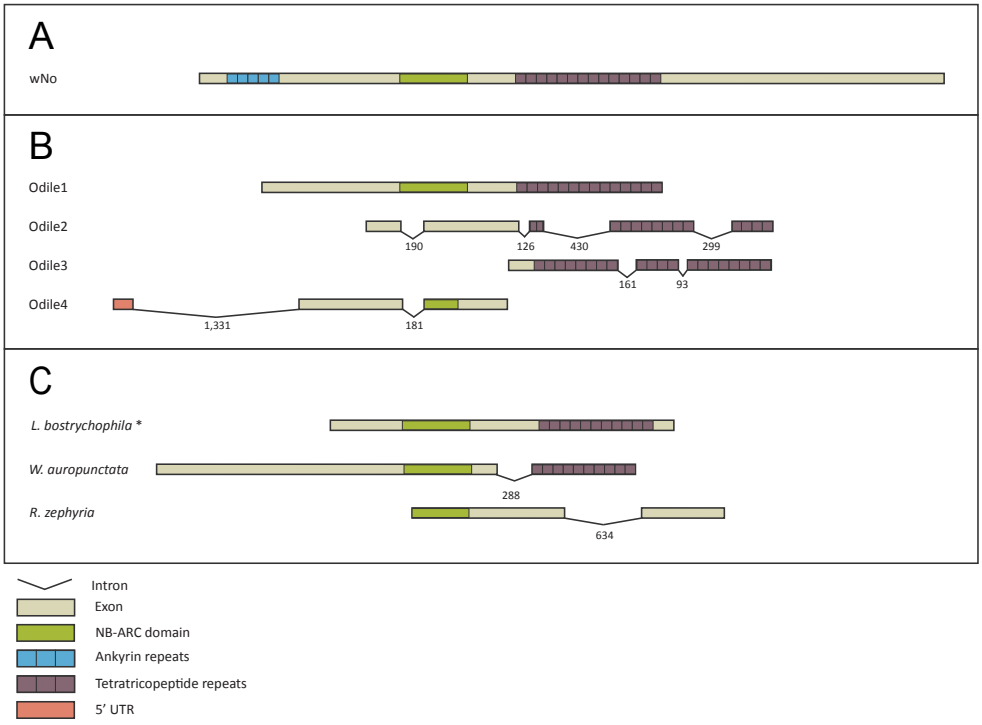


Figure S3, Nucleotide alignment schematic of the candidate gene *Odile* and homologs, showing putative introns and conserved domain locations. Related to Figure 4. *Wolbachia* homolog from *Drosophila simulans* (wNo). **B.** *Odile1* paralogs in *Liposcelis* sp. distorter. **C.** Homologs found in the insects *Liposcelis bostrychophila*, *Wasmannia auropunctata* and *Rhagoletis zephyria*. Relative domain positions were predicted by CDD. Numbers below introns represent length in nucleotides; note that *Odile4* was missed in microbial BLAST screens as the closest match was from *Wasmannia*. *Intron presence/absence is unknown for *L. bostrychophila* as only transcript sequence was available.

Female ID	Female Genotype	Male Genotype	Offspring Genotype	Total Offspring
1F	A'a	Aa	A'a	13
6F	A'A	Aa	A'A	15
8F	A'A	AA	A'A	14
12F	A'a	aa	A'a	14
15F	A'A	Aa	A'a	13

Table S1. Results of the first generation of the inheritance experiment examining allele transmission patterns at the cyclic phosphodiesterase locus in distorter females. Related to Figure 1. The A' allele was transmitted to all F1 offspring ($Phos1^{D}$ in Figure 1).

Female ID	Female Genotype	Male Genotype	Offspring Genotype	Total Offspring
6-2F	A'A	aa	A'a	10
6-3F	A'A	AA	A'A	14
8-5F	A'A	Aa	A'a	10
8-10F	A'A	aa	A'a	14
12-6F	A'a	AA	A'A	13
12-7F	A'a	Aa	A'A	12
15-1F	A'a	AA	A'A	15
15-3F	A'a	Aa	A'a	13

Table S2. Results of the second generation of the inheritance experiment examining allele transmission patterns at the cyclic phosphodiesterase locus in distorter females. Related to Figure 1. The first number in the female ID indicates which parental female the F1 mother is descended from. The A' allele was transmitted to all F2 offspring.

Assembly	Sequencing	Assembly Strategy	BUSCO completeness
Nondistorter initial genomic	37M 100bp PE reads	Ray (k=31)	N/A
Distorter initial genomic	45M 100bp PE reads	Ray (k=31)	N/A
Pan genome assembly (distorter + non-distorter)	initial reads, 53M 100bp PE reads (male), 44M 100bp PE reads (nondistorter), 91M 100bp PE reads (distorter), 5 cells PacBio (3.1 Gb longest subreads, mean length > 8kb) (distorter)	Ray (k=31), DBG2OLC, Sparc, Pilon	Genomic: 94% Gene Models: 76%
Pan transcriptome (distorter + non-distorter)	mean 44M 125bp PE per library (6 libraries)	Trinity *de novo*, Trinity genome guided, Evidential Gene	96%
Pediculus schaeffi genomic	PRJNA230884	Ray (k=31)	N/A
Liposcelis entomophila transcriptomic	PRJNA214735	Ray (k=31), Trinity (not shown)	N/A
Liposcelis bostrychophila transcriptomic	PRJNA188391	Ray (k=31), Trinity (not shown)	N/A

Table S3. Sequencing reads and assembly strategies for all assemblies. **Related to Figure 2.**