Supplementary Tables

Table S1. Genome Assembly statistics.

	Data (Gb)	N50	# Scaffolds >1000bps	Total Mbps assembled
D. busckii	9.5	36,315 bp	9,640	121
P. variegata	4.5	15,396 bp	19,486	183
B. oleae	21.2	5,992 bp	68,784	457
S. bullata	20.0	1,359 bp	184,355	752
H. illucens	11.3	2,778 bp	319,229	1,214

Table S2. P-values obtained from one-tailed Wilcoxon tests comparing the M/F coverage of each Muller element to the M/F coverage of the rest of the mapped scaffolds within a species (for instance, Element A versus Elements B, C, D, E and F)

Muller Element	*D. busckii*	*P. variegata*	*B. oleae*	*S. bullata*	*H. illucens*
A	**< 2.2e-16**	**< 2.2e-16**	0.83	0.99	0.01
B	1.00	1.00	0.99	0.81	0.13
C	1.00	1.00	0.32	0.09	1.00
D	1.00	1.00	0.55	0.53	1.00
E	1.00	1.00	0.82	0.84	0.09
F	**4.32E-08**	1.00	**< 2.2e-16**	**< 2.2e-16**	**< 2.2e-16**

Table S3. No evidence of a shared homologous Y between Drosophilids and their outgroups at the DNA level: number of homologous sequences of ancestral Drosophila Y-linked genes identified in outgroup species by a tblastn search, using the Drosophila proteins as a query. The male and female genomic coverage of each scaffold was estimated using one lane of female and one lane of male genomic sequencing data. Candidate Y-linked sequences are expected to show no coverage in females, and low coverage in males. Such sequences are only observed in *P. variegata*.

	P. variegata		*S. bullata*		*B. oleae*		*H. illucens*	
	Total	Male-specific	Total	Male-specific	Total	Male-specific	Total	Male-specific
CCY	3	**3**	1	0	1	0	1	0
ARY	8	0	10	0	6	0	33	0
KL-2	33	**13**	69	0	26	0	74	0
KL-3	27	**6**	78	0	24	0	79	0
ORY	1	0	1	0	1	0	3	0
PRY	4	0	2	0	2	0	0	0
PPR-Y	5	0	3	0	4	0	6	0

Table S4. No evidence of a shared homologous Y between Drosophilids and their outgroups at the RNA level: number of homologous transcripts of ancestral Drosophila Y-linked genes identified in outgroup species by a tblastn search, using the Drosophila proteins as a query. The male and female genomic coverage of each transcript was estimated using one lane of female and one lane of male genomic sequencing data. Candidate Y-linked sequences are expected to show no coverage in females, and low coverage in males. These sequences were not observed in either *S. bullata* or *B. oleae.*

	S. bullata		*B. oleae*	
	Total	Male-specific	Total	Male-specific
CCY	3	0	2	0
ARY	12	0	14	0
KL-2	45	0	44	0
KL-3	41	0	44	0
ORY	2	0	1	0
PRY	1	0	3	0
PPR-Y	4	0	8	0

Table S5. Mapping of candidate *S. bullata* and *B. oleae* Y-linked sequences to the *D. melanogaster* genome. In order to detect conserved Y-linkage between Drosophilids and their outgroups, known *D. melanogaster* Y-linked genes, as well as candidate Y-linked transcripts obtained in *B. oleae* and *S. bullata*, were mapped against our assembly of the *D. melanogaster* genome using tblastx with an E-value cut-off of 0.01. Male and female genomic coverage of the *D. melanogaster* genomic scaffolds (Mcov and Fcov) was obtained by aligning male and female genomic reads with bowtie2 and estimating coverage with SoapCoverage. The table shows the total number of hits in the *D. melanogaster* genome assembly, as well as the number of hits to scaffolds with male-specific coverage. All 12 *D. melanogaster* Y-linked genes mapped to scaffolds with male-specific coverage; only one candidate transcript of *B. oleae* mapped to a scaffold with male-specific coverage in *D. melanogaster*, but it was shown experimentally not to be Y-linked in either species; no *S. bullata* Y-candidates mapped to scaffolds with male-specific coverage in *D. melanogaster.*

	D. melanogaster Y linked genes	*S. bullata* Y candidates	*B. oleae* Y candidates
[a] *D. melanogaster* scaffolds	129	83	214
[b] *D. melanogaster* scaffolds; Fcov=0 and Mcov>0	40	0	1†
[c] Y-linked candidates	12	23	9
[d] % Y-linked candidates Fcov=0 and Mcov>0	100% (12/12)	0% (0/23)	11% (1/9)†

[a] Number of unique *D. melanogaster* scaffolds with significant hits
[b] Number of unique *D. melanogaster* scaffolds with significant hits, and female coverage (Fcov) =0 and male coverage (Mcov)>0
[c] Number of Y-linked sequences/candidates that map to the *D. melanogaster* genome assembly
[d] Percentage (%) of Y-linked sequences/candidates that map to *D. melanogaster* scaffolds with Fcov=0 and Mcov>0
† Transcript C89817 of *B. oleae* and genomic scaffold C1634008 of *D. melanogaster* share sequence similarity but are not Y-linked (see Fig. S10).

Supplementary Figures:

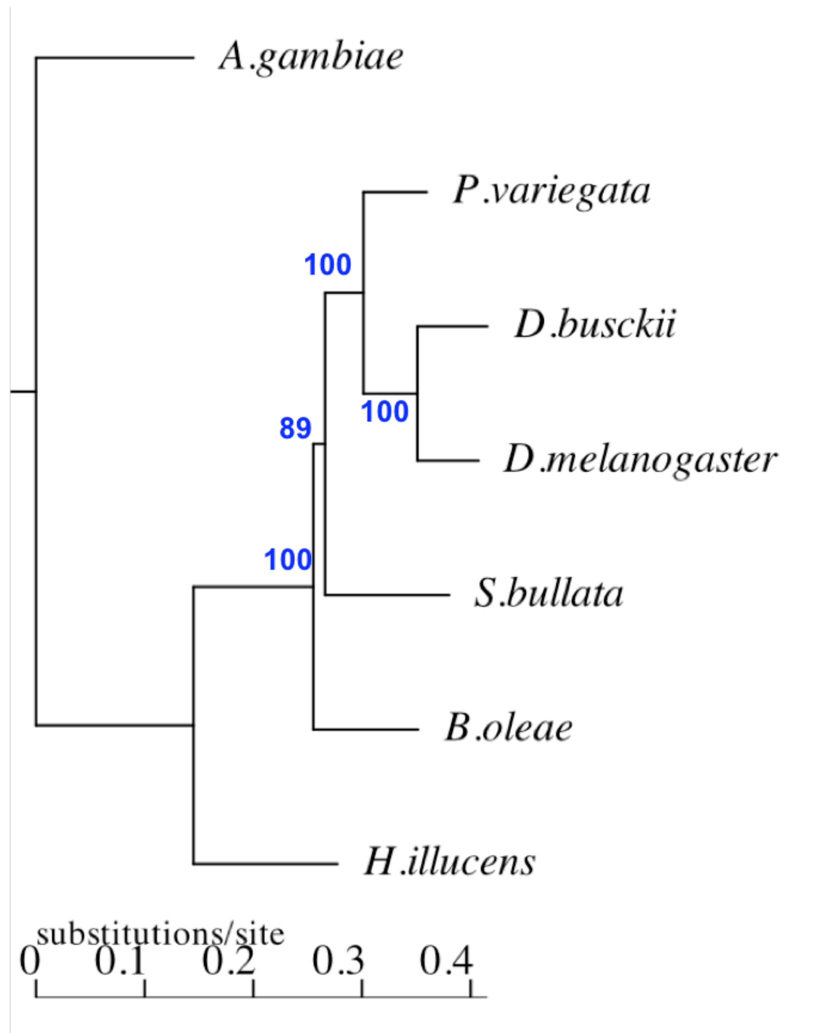

Figure S1. Phylogeny of the species using inferred protein sequence for the 185 genes that had a blat alignment score to *D. melanogaster* genes above 1000. Bootstrap values (100 replicates) are provided in blue.

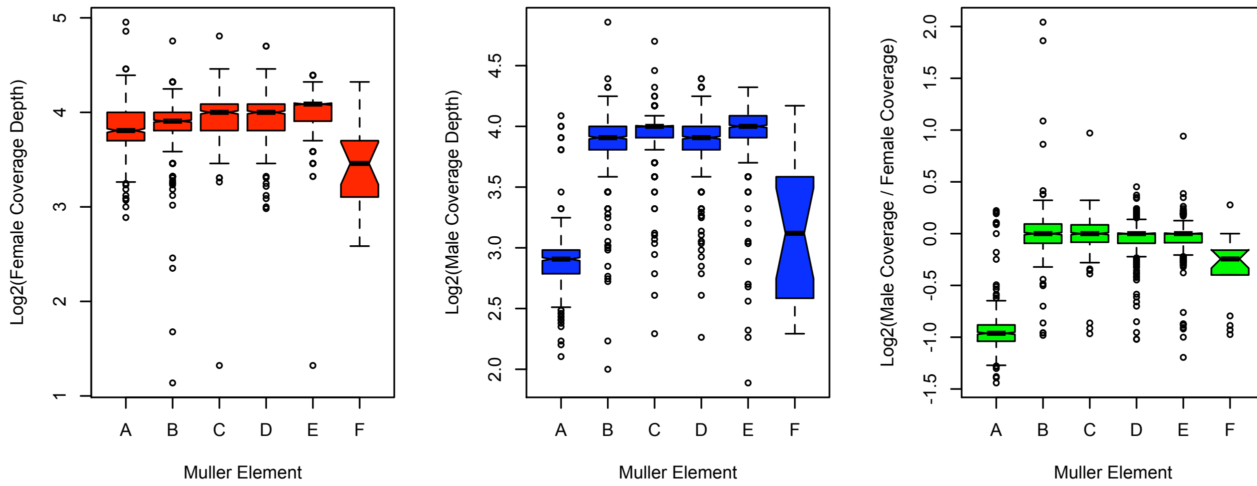

Figure S2. Log2 of Female, Male and M/F coverage for each Muller element in *D. busckii*, using a genomic assembly of male and female reads. Boxes extend from the first to the third quartile, notches to +/-[1.58*IQR/√(n)] (where IQR is the interquantile range and n the number of observations), and whiskers to the most extreme data point within 1.5 times the IQR.

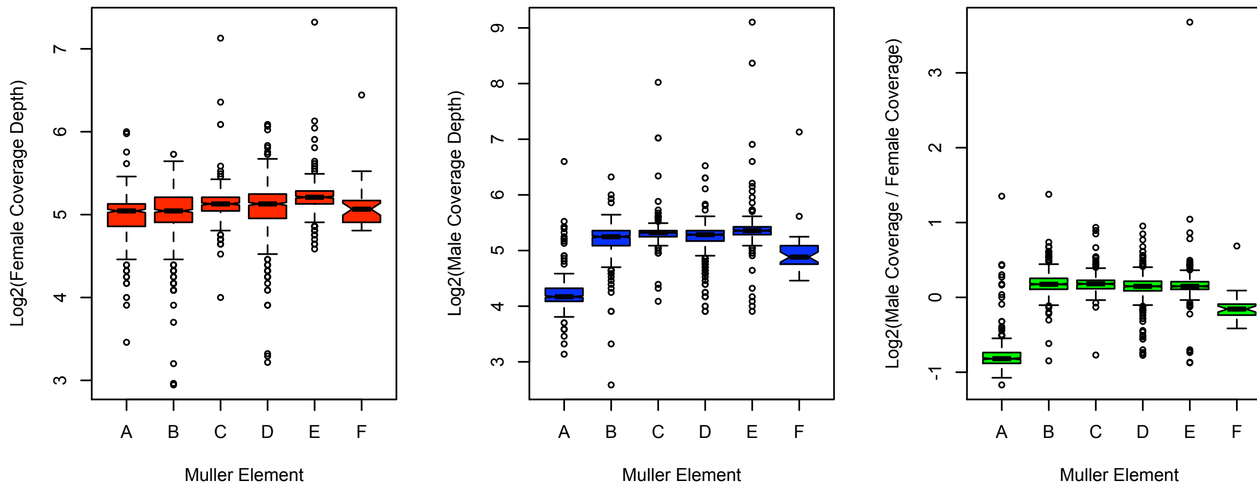

Figure S3. Log2 of Female, Male and M/F coverage for each Muller element in *D. busckii*, using a genomic assembly of only female reads. Boxes extend from the first to the third quartile, notches to +/-[1.58*IQR/√(n)] (where IQR is the interquantile range and n the number of observations), and whiskers to the most extreme data point within 1.5 times the IQR.

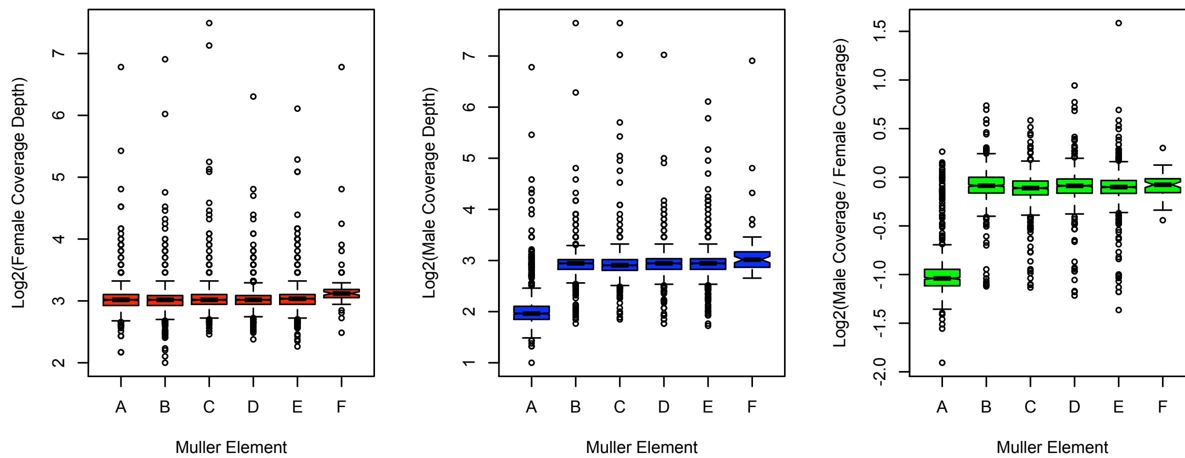

Figure S4. Log2 of Female, Male and M/F coverage for each Muller element in *P. variegata*. Boxes extend from the first to the third quartile, notches to +/-[1.58*IQR/√(n)] (where IQR is the interquantile range and n the number of observations), and whiskers to the most extreme data point within 1.5 times the IQR.

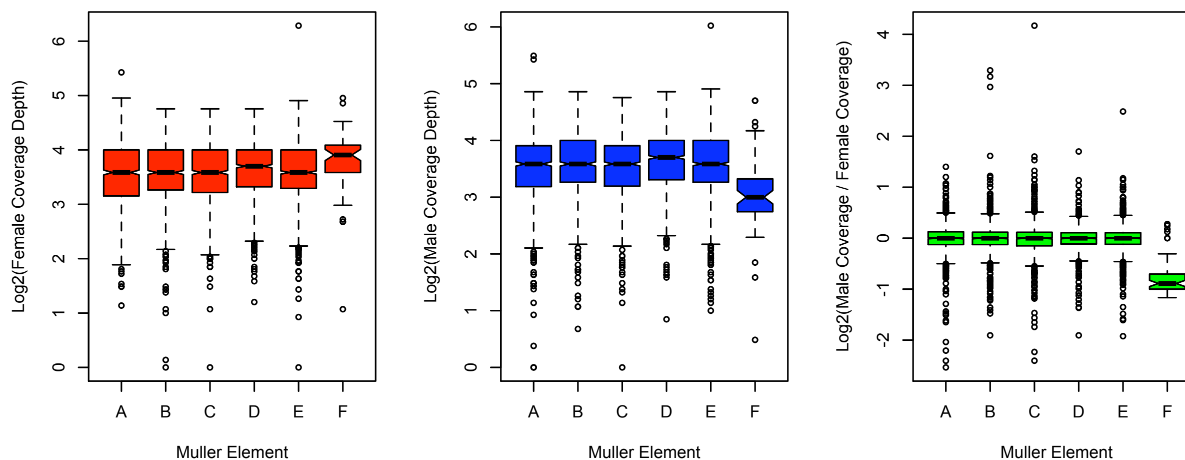

Figure S5. Log2 of Female, Male and M/F coverage for each Muller element in *S. bullata*. Boxes extend from the first to the third quartile, notches to +/-[1.58*IQR/√(n)] (where IQR is the interquantile range and n the number of observations), and whiskers to the most extreme data point within 1.5 times the IQR.

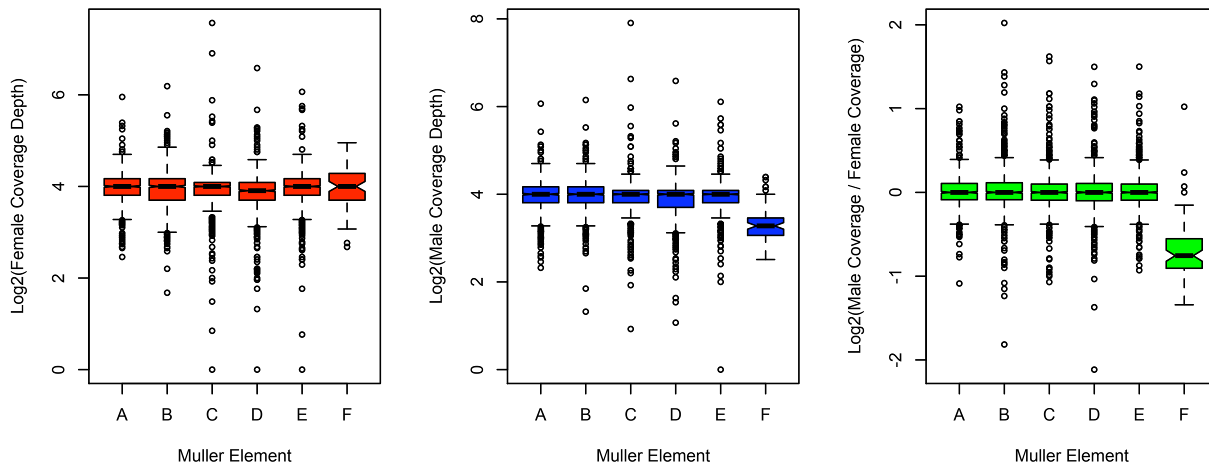

Figure S6. Log2 of Female, Male and M/F coverage for each Muller element in *B. oleae*. Boxes extend from the first to the third quartile, notches to +/-[1.58*IQR/√(n)] (where IQR is the interquantile range and n the number of observations), and whiskers to the most extreme data point within 1.5 times the IQR.

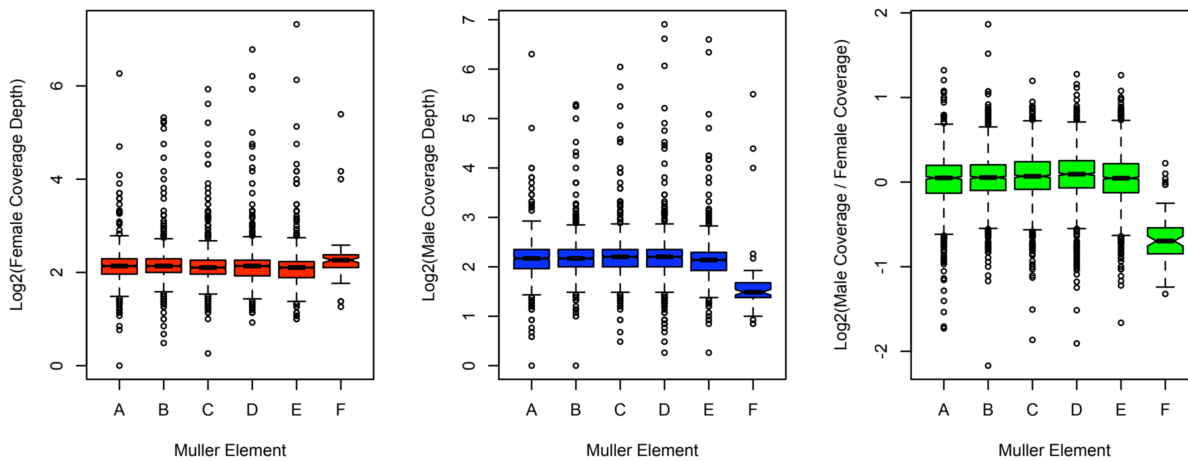

Figure S7. Log2 of Female, Male and M/F coverage for each Muller element in *H. illucens*. Boxes extend from the first to the third quartile, notches to +/-[1.58*IQR/√(n)] (where IQR is the interquantile range and n the number of observations), and whiskers to the most extreme data point within 1.5 times the IQR.

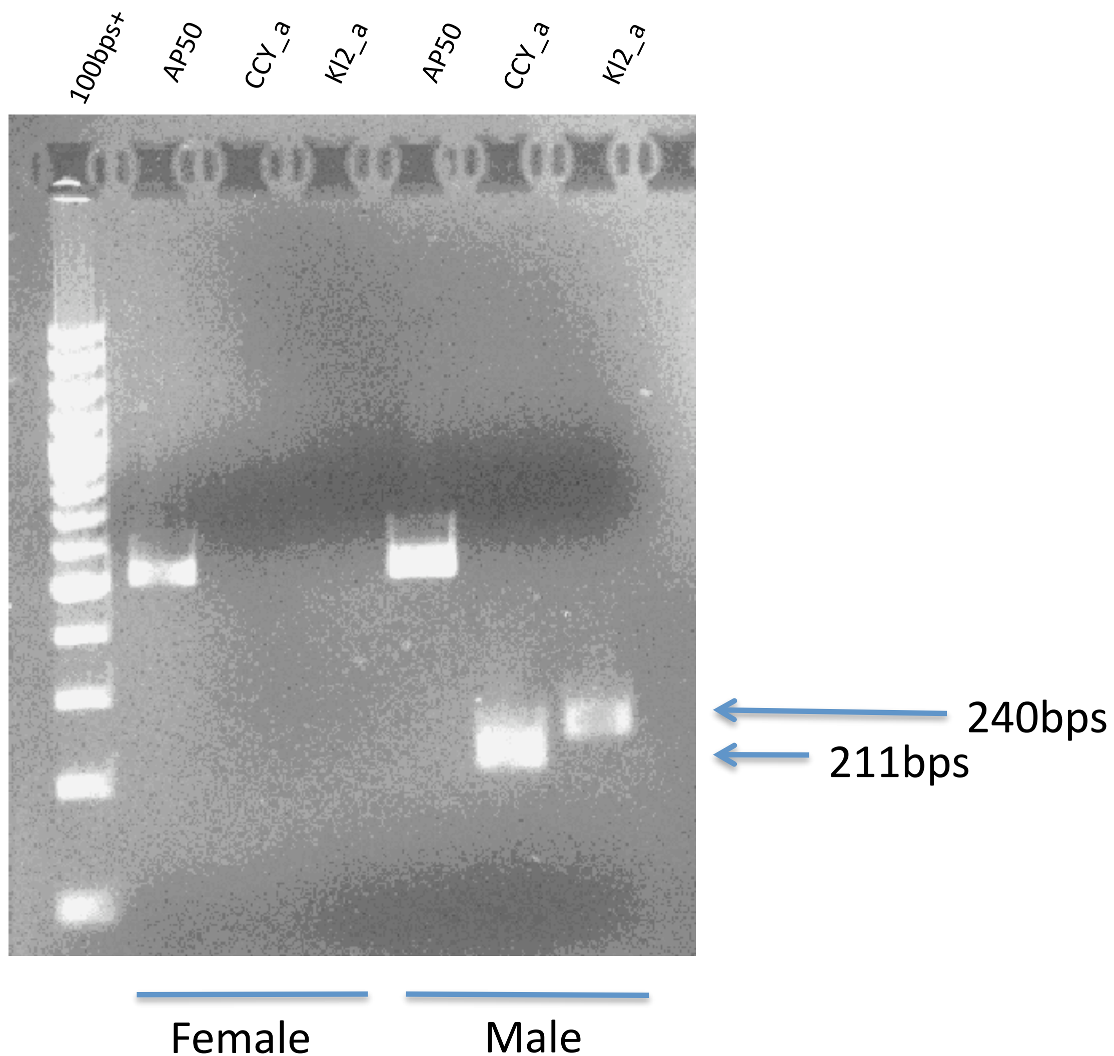

Figure S8. PCR confirmation that *CCY* and *kl-2* are Y-linked in *P. variegata*. Primers were designed to amplify fragments of the *P. variegata* male genomic scaffolds 13109 (homologous to *kl-2*; primers for: AAGTTGGCACCACTTTTTGG and rev: TCTACCAGACCCGCCAATAC) and scaffold15584 (homologous to *CCY*; primers for: AGTCGGTGTTTCCCCTTCTT, rev: GGATCCAGAAGCCATGAAAA). Primers designed to amplify a fragment of the autosomal gene *AP-50* were used as a control (for: TCCCCAAAATACCGACTCTG, rev: CGTCCGGCGGTATAAAACTA). For each set of primers, standard PCR was performed with either male or female DNA as template, and an annealing temperature of 58°C. For the *AP-50* primers, a band of expected size of 511bps was observed with both the male and the female template DNA. Expected bands of 211bps and 240bps (for *CCY* and *kl-2*, respectively) were observed when the male DNA was used as template, but not when the female DNA was used, confirming that these genes are Y-linked in *P. variegata*.

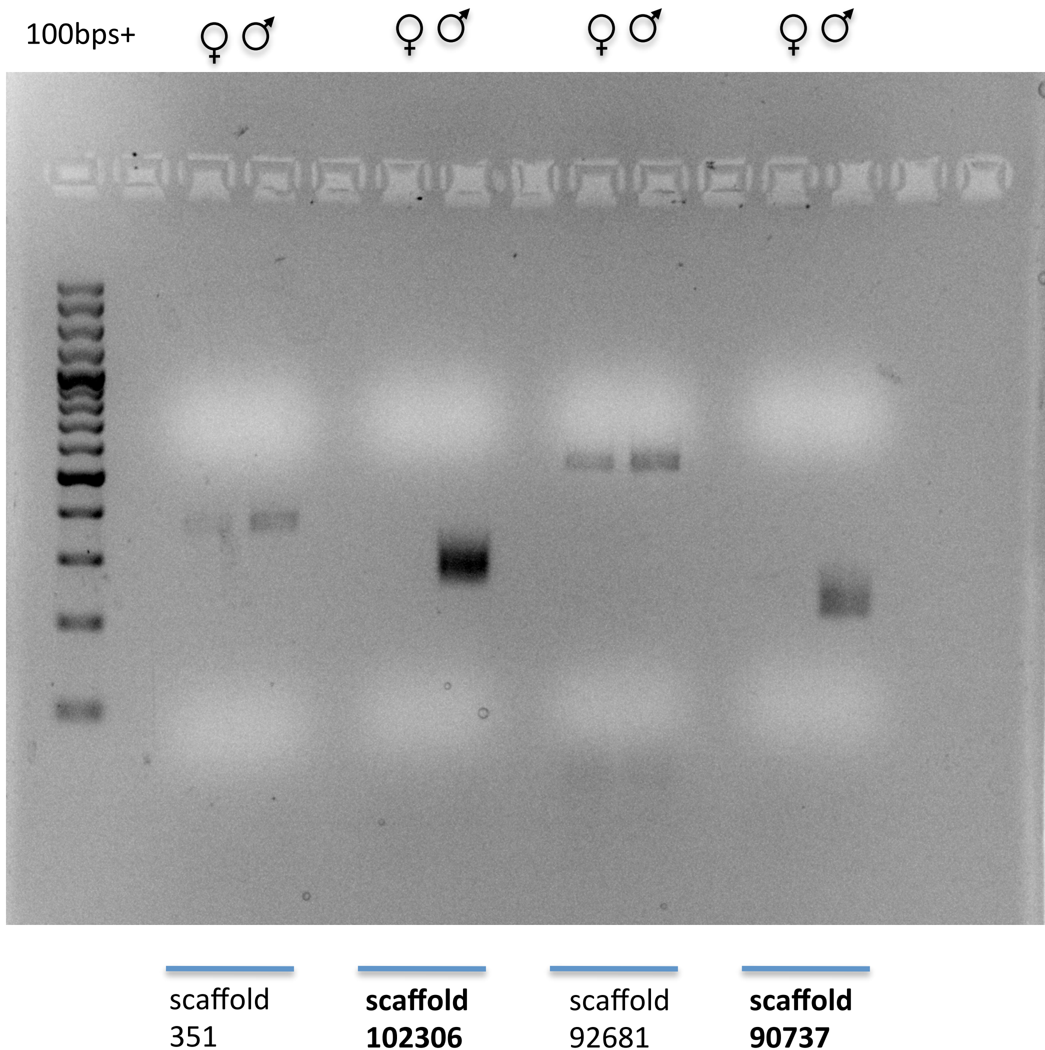

Figure S9. PCR confirmation that the male-specific transcripts scaffold102306 and scaffold90737 are Y-linked in *S. bullata*. Primers were designed to amplify fragments of the *S. bullata* male transcripts scaffold102306 (for: CCCTAAAGATAATCGCAGAATGA and rev: GCATATAGTGACGGCATATCG) and scaffold90737 (for: CAAAGGAGGGAGGGAGAGTT, rev: GCAGCTCAAGAATCATGCAA). Transcripts scaffold351 (for: TCCCGTAGAAGCGACGTTAT, rev: GGAGCATGCAATGCTAGTTG) and scaffold92681 (for: TCTTTTGCAATAGTGTTGGGTTT, rev: CTGAGAGGCTGATTACCAAGC) are not Y-linked and produce non-sex-specific bands. For each set of primers, standard PCR was performed with either male or female DNA as template, and an annealing temperature of 58°C.

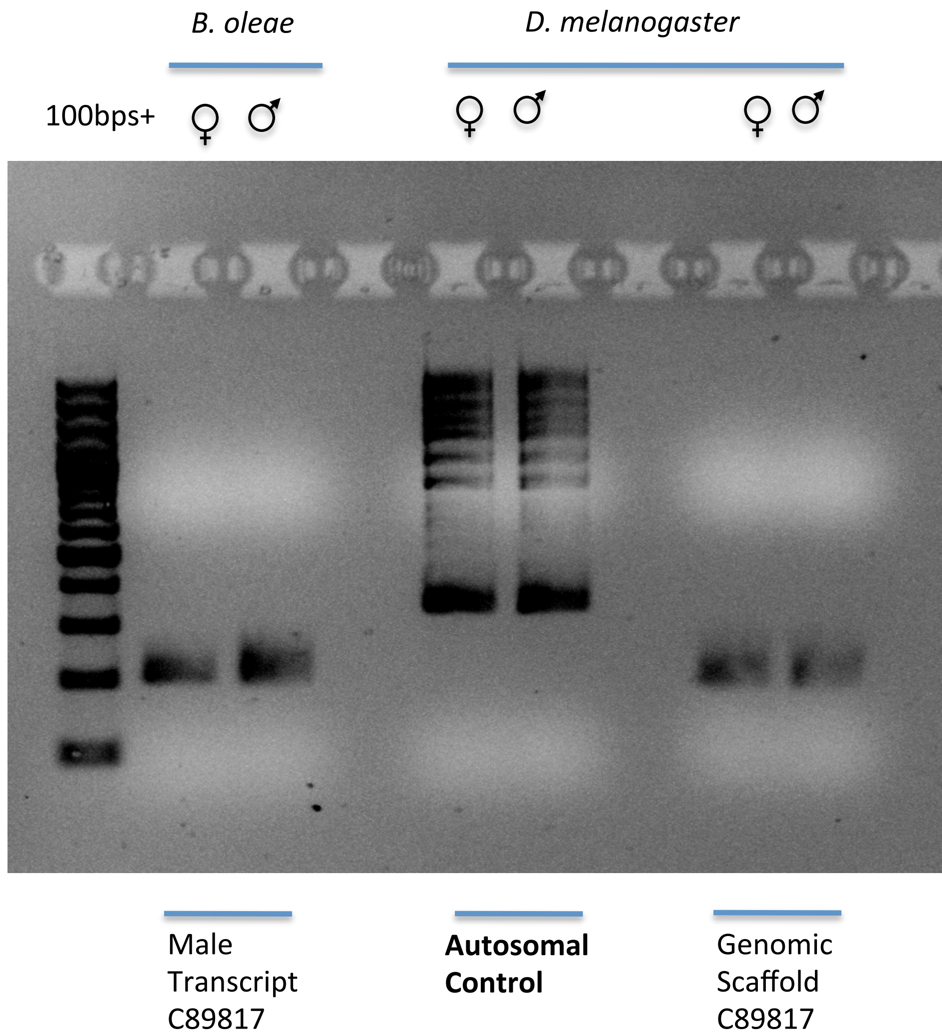

Figure S10. PCR evidence that the candidate Y-derived sequences C89817 of *B. oleae* and C1634008 of *D. melanogaster*, which share sequence similarity, are not Y-linked in either species. Primers were designed to amplify fragments of the *B. oleae* male transcript C89817 (for: TCCGCAGTAGCTTAACCCTAA, rev: AACAACAATAACTGAAAAGGCACA) and of the *D. melanogaster* genomic scaffold C1634008 (for: CCCAGAAAAATGAGAGGCAAT, rev: GCTCGATGTCCGAGCAAT). For each set of primers, standard PCR was performed with either male or female DNA as template, and an annealing temperature of 58°C.