Table of Contents

Table S2.1. The mean abundance across all MA lines and population isolates of each of 162 kmers found to have at least 2 normalized copies per sample.

kmer index	kmer sequence	mean abundance
1	A	79528.08824
2	AACCT	62281.38235
3	ACGCCAGAGCACGCCAGTGC	20047.85294
4	AATGG	18852.94118
5	AGG	8501.088235
6	AACCTTGGCG	6892.147059
7	AACAG	6845.5
8	C	6379.735294
9	ACGCCAGAGC	3562.970588
10	AAAAATAACAAC	2187.529412
11	AAG	1791.852941
12	ACGCCAGTGC	1718.705882
13	AGCCTG	883
14	AATCTGGAATGGAATGG	785.7058824
15	AAGCCAGTGCAGC	759.5882353
16	AAAG	708.2941176
17	AAGAGCACGCCAGTGCACGC	616.2941176
18	AAAATAGG	531
19	ACAGC	525.2647059
20	AAAAG	470.7941176
21	AG	407.9705882
22	AAGAG	350.7941176
23	ATCC	336.9411765
24	AAC	318.5294118
25	ACAGGAGAGC	288.9117647
26	AC	243.5588235
27	AAGGAG	230.9705882
28	AAGTGCACGCCAGAGCACGC	216.4705882
29	ACAGGAGACC	212.8235294
30	AAAAAGAAGATAGTAGG	199.8823529
31	AACAAG	195.8823529
32	AAATAGG	188.7647059
33	ACACCGAGGCTCCAGCTACT	137.9705882
34	AATGCACGCCAGAGCACGCC	126

35	AATGCACGCC	124.0588235
36	ACACCGAGTCTCCAGCTACT	112.3235294
37	AACGGTACGG	110.8529412
38	ACT	106.4117647
39	AATTCACACCAGAGCACGCC	104.0882353
40	AGC	77.85294118
41	ACTGGCATGC	73.91176471
42	AATGCACGCCAGAGCACGTC	69.79411765
43	ACTGG	66.79411765
44	AAAATTGAAGGAAGATAG	62.11764706
45	ACGAGCACGCCAGTGCACGC	57.73529412
46	AAGAT	54.41176471
47	AAAATAGGAAATAGG	52.70588235
48	ACGGG	53.97058824
49	ACGCCAGTGCATGCCAGTGC	50.05882353
50	AAATAGGG	45.35294118
51	AATGAGGAGGAGGAG	46.05882353
52	AAGAAT	40.67647059
53	AAAATGAATACATC	39.61764706
54	AACAGG	38.97058824
55	ACGCCAGTGCACGTCAGAGC	36.32352941
56	AACAAGAATAAG	35.55882353
57	AAGACG	35.82352941
58	AAGCTTCGCC	34.58823529
59	AATGCACGCCAGTGCACGCC	32.02941176
60	AACAGCACAG	32.11764706
61	ACCCG	30
62	ACACCGAGGCTCCAGCTGCT	31
63	AATCTGTAATGGAATGG	28.55882353
64	AAAATTGG	28.35294118
65	AAAAATAAGAAC	24.44117647
66	AGCAGG	22.61764706
67	AAAAAG	22.52941176
68	ACAGTGCACGCCAGAGCACG	22.73529412
69	ACGAGTAGTAGC	22.44117647
70	AAAAACCAAAATCAAAC	22.11764706
71	AAGATGGCCCGAATGGGCC	21.73529412
72	AAACAACCACGGCGTTAATG	20.20588235
73	AAGCTTCGGC	20.58823529

74	AAAAAAGATAGAATG	19.76470588
75	AACAGAACAGCACAG	19.97058824
76	ACGCCAGTGCACGCCGGAGC	19.23529412
77	AAACATCCACGGCGTTAATG	19.11764706
78	AACGCCAGAGCACGCCAGTG	18.58823529
79	ACGCCAGAGCACTCCAGTGC	16.82352941
80	AAGAGGACCG	16.82352941
81	AAGAAGGAG	16.64705882
82	AAGAGCACGC	16.41176471
83	AAACTTAATTTC	15.88235294
84	AATCTGGAATGG	15.64705882
85	AAAGCACGCCAGTGCACGCC	15.67647059
86	AACTACTACACC	14.67647059
87	ACGAGTAGCAGC	14.23529412
88	AAATTTCTATCTTCCTTC	13.61764706
89	ACACCAGAGCACGCCAGTGC	13.11764706
90	ACGCCAGTGCACGCCATAGC	12.41176471
91	ACGCCAGAGCACGGCAGTGC	12.29411765
92	AAGAGGGAGAAT	11.67647059
93	ACACCATCCACT	11.64705882
94	ACGCCAGTGCCCGCCAGAGC	11.20588235
95	ACTC	11.14705882
96	ACGCCAGAGCACGTCAGTGC	10.85294118
97	ACGACGAGG	10.52941176
98	ACACGGCAG	10.02941176
99	ACCGGCTCCTCG	10.17647059
100	AACGGAGGAGG	9.970588235
101	AAGG	9.852941176
102	AGCATG	9.441176471
103	ACGAGAGGCTCAGCT	9.264705882
104	AAAACAAAGACG	9.5
105	ACGCAGTGCACGCCAGAGC	9.235294118
106	AAGCCAGTGCAGCAAGGC	9.205882353
107	ACGCCAGAGCACGCTAGTGC	8.588235294
108	AAGTCCATGCTC	8.352941176
109	AACTCCACGATC	8.323529412
110	ACACCACGGCCTACGCT	8.294117647
111	AAGCTACTCTGCTCC	7.941176471
112	ACGCCAGAGCACGCTAGAGC	7.705882353

113	ACCCCATCCACT	7.352941176
114	AACACGTTGAGTG	7.470588235
115	AACGCCAGTGCACGCCAGAG	7.5
116	ACGCCAGTGCGCGCCAGAGC	7.441176471
117	ACCACGACTCCT	7.117647059
118	AACTGGACGACGAAGACG	7.264705882
119	ACACCTAGCT	7.235294118
120	AAGGAGGAG	7.117647059
121	AACGCCAGTG	7.088235294
122	AACTACTACCCC	6.735294118
123	ACGAGCGGAGAGAGCG	6.911764706
124	AAGCACTGGCGTGCTCTGGC	6.588235294
125	ACAGAGCACGCCAGTGCACG	6.617647059
126	ACGCCAGAGCACGCCCGTGC	6.529411765
127	AATCAGTGTCTAC	6.441176471
128	ACCATCCCCTCCCTC	6.352941176
129	AAGAGCACGTCAATGCACGC	6.235294118
130	ACGCCAGAGCACGCGAGTGC	6.117647059
131	AACAGGAGCTGG	6.029411765
132	AAGACGAGGAGG	5.911764706
133	AATATAATGGAATGGAATGG	5.705882353
134	ACAGCACGCCAGTGCACGCC	5.323529412
135	ACACCGACTCTACT	5.323529412
136	AACCTCCTCCTCCTC	5.176470588
137	AAGTACGAGACTCCC	5.352941176
138	ACACCAGTGCACGCCAGTGC	5.294117647
139	AATGCACGCCAGTGCATGCC	4.970588235
140	AAACACTTGGTCT	4.676470588
141	AGCTCC	4.558823529
142	ACCTCTGGCGTGCACTGGCG	4.617647059
143	AATGCACGCCAGAGCATGCC	4.558823529
144	AATCTGGACTGGACTGG	4.558823529
145	ACGGCCGCTGCTACT	4.647058824
146	AACAAGAACAAGAATAAG	4.323529412
147	AAGAGCACGCCAGTGCATGC	4.441176471
148	ACGATAGTGCACGCCAGAGC	4.176470588
149	AAGCTCCAGCCTACC	3.970588235
150	ACACACCTCCGGCAT	4.117647059
151	ACGCCAGTGCTCGCCAGAGC	4

152	AACTTCCAGTGACC	3.911764706
153	ACGCCAGAGCATGCCAGTGC	4.029411765
154	AACCCACCACACTACCCC	3.882352941
155	AATGACAGAGATG	3.764705882
156	AAGCTGGAGGAGAAT	3.735294118
157	AAGCCAGTGCACGCCAGAGC	3.411764706
158	ACACCAGTGCACGCCAGAGC	3.382352941
159	ACGCCAGTGCACGCTAGAGC	3.352941176
160	ACAGGAGACCACAGGAGAGC	3.323529412
161	AAAGAGCACGCCAGTGCACG	3.235294118
162	AAGTGCACGCCAGTGCACGC	2.882352941

Table S3.1. Complex repeats occupying multiple kb spans in the *Chlamydomonas reinhardtii* assembly (including mitochondrial and chloroplast DNA)). Phobos (parameters: -U 200, restricted to chromosome 1) and Tandem Repeats Finder, TRF (parameters: trf Creinhardtii_281_v5.0.chloro.mtDNA.fa 2 7 7 80 10 100 500 -h) were used. This is not a comprehensive description of all complex repeats in the genome.

Repeat sequence	Estimated Span	Program
AACCACGCTCCAGCCCCTCCCACGCACGCCAAGGTCAT CGCCCCGCCCTGCCCTCTTGCATATGGCCTTGGCTATA TGTCAGAGGGATGGGGGTGCAGTATGGCCGGGTGATT CCGTCGGGTGGGGTCACGGCAGGGGTCGAGGTGTGG ATCTGACATGGGCGGCACACCCAGCCTATGCCTT	2.1 kb, Chr1	Phobos
AAACACCCGGCCATACTGCATGCCCGTCCATATGGCAT ATAGACACGGCCTTGTGCAAGAGGGCAGGGCTGGGCG CTGACCTTTGCATGCGGGGCAGGGGCTGGAGCGTGGT CTGGGCATGGGCGGGGTTAGCCACCCATGCCACATCC ACACCTTGACCCCTGCCACGACCCCACCCGACGG	1.7 kb, Chr1	Phobos
AACAGGCGACGGAGGCTCCGGCGACGGAGGCGCCGG GCTGGGAGGCACCGGCGATGGCGCAGGCGACGGAGG CTCCGGGCTAGGCGGCGCCGGGCTGGGAGGAGAGGG GCTGGGCGGCGCGGAAGCGGGCTCCGGGCTGGGAGG	8.0 kb, Chr1	Phobos
AAGGCGCAGGCGGTGCGGGCGACGGGGGCTCCGGGC TGGGCGGCGCAGGGCTGGGAGCCGGGCTGGGAGGCG CAGGGCTCGGCGGATCGGGGCTAGGAGCCGGGCTGG GCGGATCCGGAGACGGAGGCGTCGGAGACGGCGGCG CAGGGCTGGGCGGCACCGGCG	1.2 kb, Chr 1	Phobos

AAGGCGGAGCGTCCACAGGCGCGGGGCTGGGCGGCT GAGGCGAAGGCGGCACAGGCGGTGCAGGACTGGGCG GCGCGGGGCTAGGCGGCTCCGGGCTGGGAGGCACCG GCGGCGCAGGCG	2.1 kb, Chr1	Phobos
AACGGCAGCGCAGGCCGCTGCTGGCTCTGCCGCAGAG AAGGCAGGAGCTGCGGCGTCCGCGGCGCAGTCTACGG CCGCATCTGCGGTGGAGAAGGCCAGCGGAGCGGC	1.2 kb, Chr1	Phobos
ACGCCAGGCGCGGGGCTGGGCGGCGCAGGCGACGGC GGAGGC	1.8 kb, Chr1	Phobos
AAAGGGGCTGGTGGTGCATGCAGCCGGCGCGTCACAA CAGCCACGGTGCCCAAGCTAGGTCCCGCGGGGGAACA GCTGTGCGCCGGCAGGGGGGG	16.6 kb	TRF
AACTTGCCCGCACAAGCCCCGCAGGCTCCCATGCCTTG CACAGACGCATACGAGTGCAGTGCAGGCTGGTAGACAT GCCCGCCCCTTGCCTTCCGCCCTGCCTGCC	13.2 kb	TRF
AAGGCGGCGGCGTGCCAGGCGCAGGGCTGGGAGGCG CAGGCGACGGCGGGGGCACACCCGGCGCGGGGCTGG GCGGCTCCGGCG	11.2 kb	TRF
AACACAGCTGGCCACGGAACCCAGGCACTGGGGTGCA TGGGGCACAGTGTGGGGTCCGAGGGCCCGTGTGTGG GTGGTGCGGGGTGTTGCACAGTACATGTCAAGCGCCA GGCCAGCTGTGTTGGCTGTCAATGGGGCTGTCAGCTCT GTGGTGCATCACTGGCATGCCCAGTAACGATCC	11.4 kb	TRF
AACGGCGGCGGGCTCTGGCTCCGCAGCAGGCGGCGC CGCACGAGCGGCCGACGACGCAGCGGCTGCCAC	12.9 kb	TRF
AACAGGGGTCATACCGTATAGCAGTTGCACGGCACGTT GGCTGCTGGTCCGCAGGCCGGCGCTGGCGGGGGAAC AGTCGTGCTCTCGCGCGC	46.4 kb	TRF
AAAGCAAGCAAGCAACATCTCCCACGGTCCGCCCACCG CCGCCGCCGTGTGCTGACCTCACCTCACCTCACCTCAC CTGACGCACCCATACCTGCCCTCACTCATTCACACAGG GCCATGGCTGAGGATGGCTTCATCATGACCGGCACCG ACGCTGCCACCCAGACCGAGACCGCCGCGCGCCGCGA GGCTCTGGAGGGCACCCACGGCGGCAAGAACCGCCG GTGAGTGAGCTCGGGGCGTGTGTGTGTGTGTGGTGCT TGCTGCTGTGTGTGTGCCGTGGAATAATGCGGCTGCTG GGGCTGGGCACAAGCAAGC	13.6 kb	TRF
AGCCCTGAGCCTCCCGCCCGGAGCCGCCT	11.4 kb	TRF
AAACTGCCCGCACAAGCCCCGCAGGCTCCCCATGCCTT GCACAGACGCACACGAGTGCAGTGCAGGCTGGTAGAC ATGCCCGCCCCTTGCCTTCCGCCCTTCCTGCCG	13.2 kb	TRF

Table S3.2. Repeats identified by TAREAN (Novák et al. 2017) on a subset of reads from a sequencing library used in this study. Genome percent refers to the percent of the 0.37x subsample used.

Cluster	Genome percent	N reads contributing	Consensus length (bp)	Annotation
HIGH CONFIDENCE				
CL72	0.06	299	643	5S rDNA
LOW CONFIDENCE				
CL7	0.62	3089	1190	blastn to sex determining locus (has known 16-kb stretch of repeats, Ferris et al. 2010)
CL39	0.12	617	182	no hits
CL55	0.089	443	180	no hits
CL115	0.023	115	205	blastn to sex determining locus (has known 16-kb stretch of repeats, Ferris et al. 2010)
CL119	0.022	111	180	only short hit (spurious homology, if institute > 80% of length cut-off would drop out)
CL135	0.016	81	148	ok hit to C. reinhardtii predicted protein
CL142	0.015	75	187	blastn to sex determining locus (has known 16-kb stretch of repeats, Ferris et al. 2010)

Table S5.1. Descriptions of strains used in this study.

Strain	Species	Origin	Notes
vir51	*D. virilis*	Chile	
vir52	*D. virilis*	Russia	
vir86	*D. virilis*	Mexico	
vir47	*D. virilis*	China	
vir49	*D. virilis*	Argentina	
vir85	*D. virilis*	Japan	
vir08	*D. virilis*	California, USA	
vir00	*D. virilis*	California, USA	
vir118	*D. virilis*	Rwanda, Africa	
vir48	*D. virilis*	Mexico	
vir87	*D. virilis*	Japan	inbred genome strain
vir9	*D. virilis*	Japan	
amMK1012	*D. americana*	Indiana, USA	100% X-4 fusion
amCI0518	*D. americana*	Louisiana, USA	0% X-4 fusion
amCI0515	*D. americana*	Louisiana, USA	0% X-4 fusion
amG96	*D. americana*	Indiana, USA	100% X-4 fusion, inbred genome strain
amCI0538	*D. americana*	Louisiana, USA	0% X-4 fusion
amMK0738	*D. americana*	Indiana, USA	100% X-4 fusion
amSB	*D. americana*	Iowa, USA	100% X-4 fusion
amML975	*D. americana*	Louisiana, USA	0% X-4 fusion
Gnova14	*D. novamexicana*	Utah, USA	inbred genome strain
nova13	*D. novamexicana*	Arizona, USA	
nova12	*D. novamexicana*	Colorado, USA	
nova8	*D. novamexicana*	New Mexico, USA	
nova4	*D. novamexicana*	Utah, USA	

Table S6.1. Summary of crosses for the Y fusion validation experiment. *D. novamexicana* females were crossed to vir00-Yfus males, and the male F1 progeny were backcrossed to *D. novamexicana* females. Each F2 male progeny were genotyped at each of the 4 candidate autosomal markers which are polymorphic between *D. virilis* and *D. novamexicana*. The null hypothesis was that 50% male F2s should be homozygous and 50% should be heterozygous. The chi-square p-value was only significant for vir00-Yfus on the Chr3 marker. We included crosses with vir08 instead of vir00-Yfus as a negative control. P-value indicates the p-value for a chi-square test with 1 degree of freedom. The three individuals found to be homozygous for the Chr3 marker actually had a nondisjunction event and did not contain the Y chromosome.

Strain	Chr2 het--hom	Chr3 het--hom	Chr4 het--hom	Chr5 het--hom	p-value (Chr3)
vir00-Yfus	10 -- 8	79 -- 3	9 -- 9	8 -- 10	< .00001
vir08 (control)	13 -- 8	9 -- 13	8 --14	14 -- 8	0.39

Table S6.2. Summary of crosses for the X fusion validation experiment. Vir00-Xfus females were crossed to each of three GFP-insertion lines (vir121, vir117, vir95). The resulting F1 males were crossed to GDvir (virilis genome strain vir87) females. Female progeny were scored for presence or absence of GFP signal in the larval brain. The null hypothesis was that 50% female F2s should be GFP+ and 50% should be GFP-. We included a cross with GDvir instead of vir00-Xfus as a negative control. Chrom indicates the chromosome on which the GFP insertion is located, mapped in Stern et al. 2017. P-value indicates the p-value for a chi-square test with 1 degree of freedom.

Cross	Chrom	GFP + females	GFP- females	p-value
vir00-Xfus x vir121	2	53	70	0.125
vir00-Xfus x vir117	5	103	97	0.67
vir00-Xfus x vir95	4	25	279	< 0.00001
Gdvir x vir95	4	56	55	0.92

Table S6.3. Strains used in this study. Strains were obtained from the Cornell Drosophila Species Stock Center.

Strain number	Shortform	Species	Origin	Data
15010-1051.51	vir51	*D. virilis*	Chile	sequencing
15010-1051.52	vir52	*D. virilis*	Russia	sequencing
15010-1051.86	vir86	*D. virilis*	Mexico	sequencing
15010-1051.47	vir47	*D. virilis*	China	sequencing
15010-1051.49	vir49	*D. virilis*	Argentina	sequencing
15010-1051.85	vir85	*D. virilis*	Japan	sequencing, comet assay
15010-1051.08	vir08	*D. virilis*	California, USA	sequencing, comet assay
15010-1051.00	vir00	*D. virilis*	California, USA	sequencing, comet assay, chromosome fusions
15010-1051.118	vir118	*D. virilis*	Rwanda, Africa	sequencing
15010-1051.48	vir48	*D. virilis*	Mexico	sequencing, comet assay
15010-1051.87	vir87/GDvir	*D. virilis*	Japan	sequencing, comet assay
15010-1051.09	vir9	*D. virilis*	Japan	sequencing, comet assay
15010-1051.121	vir121	*D. virilis*	Unknown	X-fusion crossing validation; genotype: Dvir\y[40a]ec[1]cv[1]v[1]si[2]dy[1]w[1]ap[40e];PBac{GreenEye.UAScnnEGFP}Dvir2
15010-1051.117	vir117	*D. virilis*	Unknown	X-fusion crossing validation; genotype: Dvir\y[40a]ec[1]cv[1]v[1]si[2]dy[1]w[1]ap[40e];PBac{GreenEye.UAStubEGFP}Dvir10
15010-1051.95	vir95	*D. virilis*	Unknown	X-fusion crossing validation; genotype: Dvir\y[40a]ec[1]cv[1]v[1]si[2]dy[1]w[1]ap[40e];PBac{5PBlueEye}Dvir1

15010-1031.14	nov14/Gdnov	*D. novamexicana*	Moab, Utah.	Y-fusion crossing validation

Table S6.4. Primer sequences used in this study.

Primer name	Forward_seq	Reverse_seq	Purpose
chr2_2	TGGAAATTTCGAGTGGTTCG	GCAAACAGTCAAGCTCGTTTC	To amplify a microsatellite locus that is polymorphic between *D. virilis* and *D. novamexicana* (Y fusion genetic validation)
chr3_2	GTGCAGCAGCCAACAGTC	ATGTCACTCACATGGCCAAA	To amplify a microsatellite locus that is polymorphic between *D. virilis* and *D. novamexicana* (Y fusion genetic validation)
chr4_2	ATTGTGCGAGTCCAGAGTCC	CATTTTGGGAGATGGCAGAC	To amplify a microsatellite locus that is polymorphic between *D. virilis* and *D. novamexicana* (Y fusion genetic validation)
chr5_1	ACACACACGACCCACCACT	GCAGCCAAGTGTTTGCTATG	To amplify a microsatellite locus that is polymorphic between *D. virilis* and *D. novamexicana* (Y fusion genetic validation)
vir_Y	CGTCATCCGTTTTGCCAGTG	AGGCATCCCGTATTAAGCGG	Y chromosome genotyping for presence/absence, designed by Yasir Ahmed-Braimah
Chr5_indel	CGCATGAACGGACAGGGTAT	AGTCAGTCTTTCTAGCAGGCG	indel genotyping to validate vir00 sublines
Chr6_indel	TGTATTCAATTTCGCGACCCT	TCCGCTTTGAGATTGGCCAA	indel genotyping to validate vir00 sublines
Chr3_indel	GCGTTATGATGCCAGGACTG	AGCAAAATGGTGGCACTAACT	indel genotyping to validate vir00 sublines
Chr2_indel	TGGACTATGCGTGTGAAGGA	CACGGACTTAAATACCTCCTTGA	indel genotyping to validate vir00 sublines

Table S6.5. Satellite DNA abundance in copy number in the vir00 substrains.

Substrain	**AAACTAC**	**AAATTAC**	**AAACTAT**
Vir00-Yfus (rep 1)	5645000	1648106	2207984
Vir00-Yfus (rep 2)	5376774	1624806	2438514
Vir00-Xfus	5502968	1508402	2342352
Vir00-Nofus	4881613	1500037	2041711

Table S6.6. Estimated rDNA copy number from read mapping of vir00 substrains and a control strain, vir08.

Substrain	Estimated rDNA copy number
Vir00-Yfus (rep 1)	625.2
Vir00-Yfus (rep 2)	634.7
Vir00-Xfus	577.8
Vir00-Nofus	637.0
Vir08	550.4

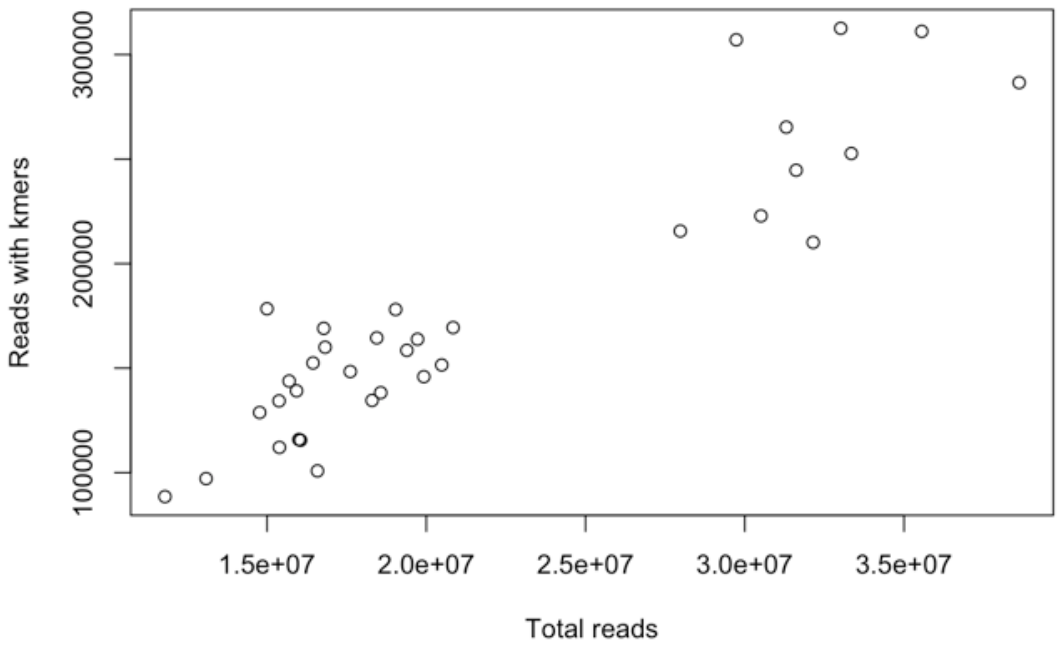

Figure S2.1. Total number of reads analyzed with k-seek versus the number of reads that were found to contain tandemly repeated kmers occupying over 50 bp of the read in tandem.

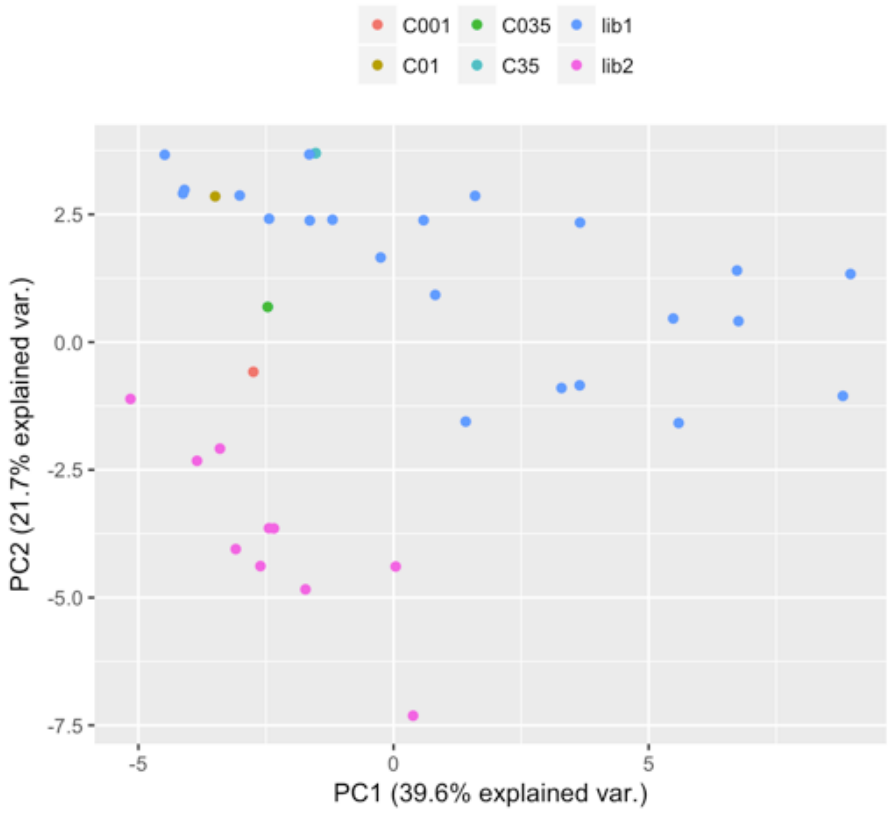

Figure S2.2. Principal Components Analysis (PCA) of samples based on the abundance of the top 39 kmers. Samples are coloured based on which library batch they were prepared in. The technical replicates, C035/C0035 and C01/C001 are labelled separately, but C01 and C035 were sequenced in batch 1 and C001 and C035 were sequenced in batch 2.

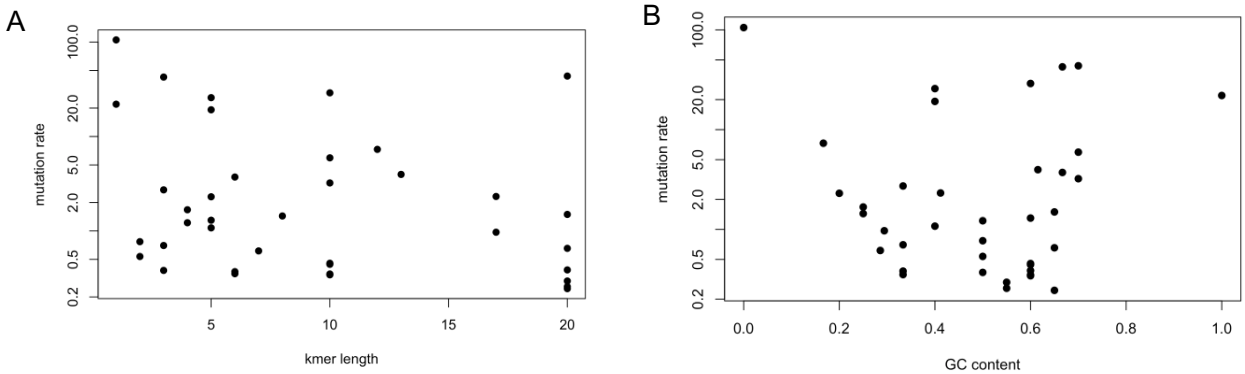

Figure S2.3. Plots comparing the mutation rate versus the length of the kmer unit (A), and the GC content of the kmer unit (B).

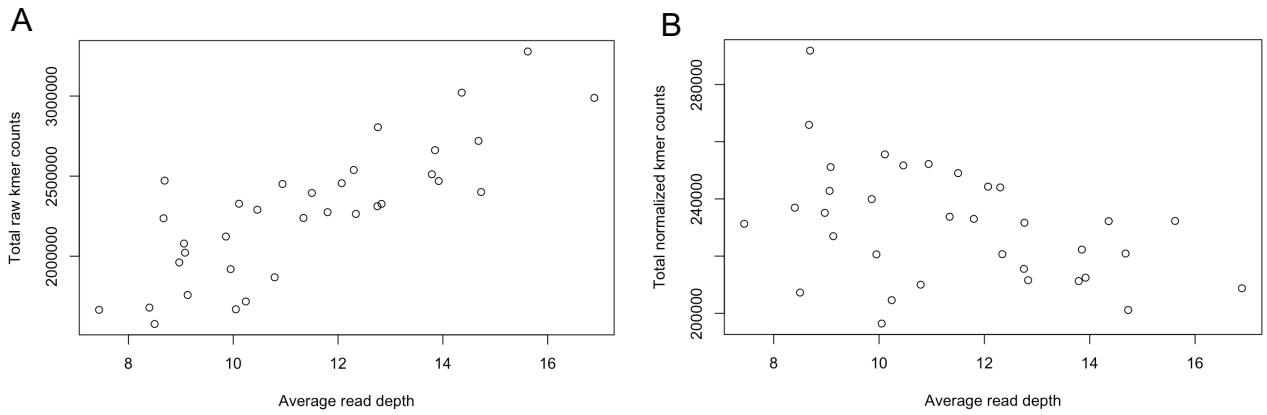

Figure S2.4. Plots comparing the kmer counts per sample versus overall sequencing depth (A) before, and (B) after normalization by GC content and read depth.

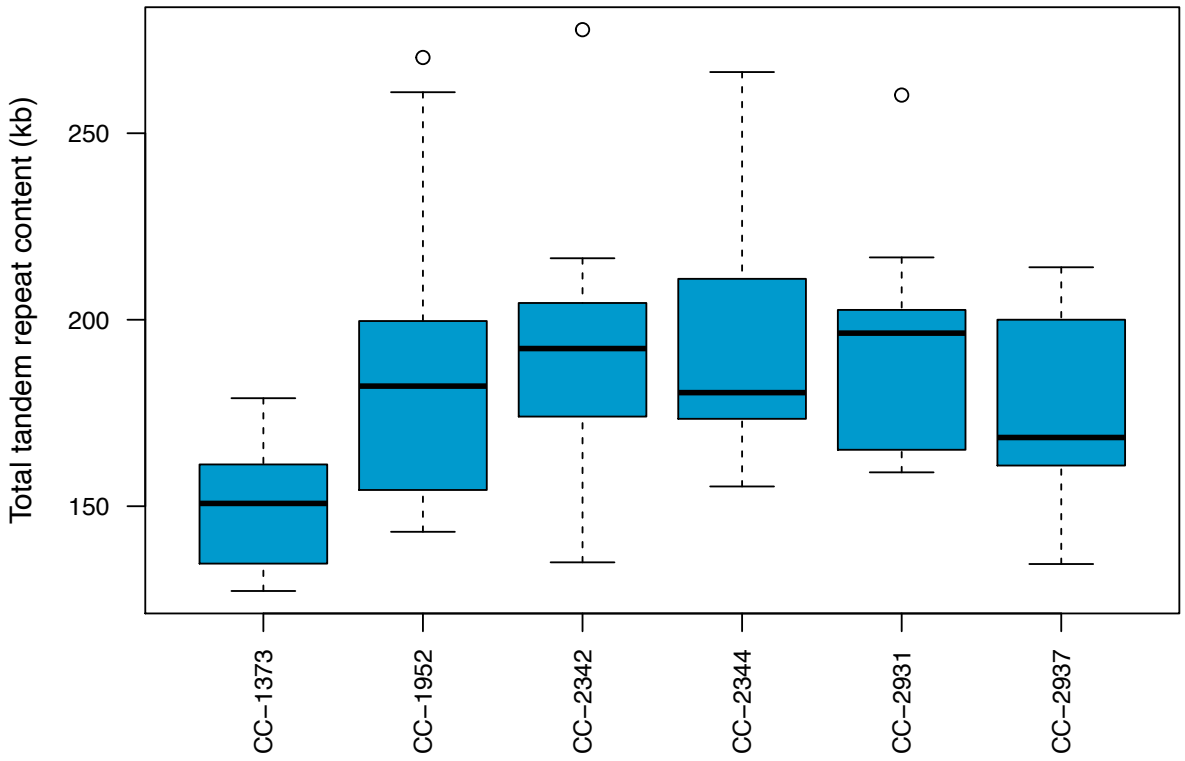

Figure S3.1. Boxplot of total genome-wide span of simple tandem repeats in the 13-15 mutation accumulation (MA) lines of six strains of *C. reinhardtii.* Number of lines per ancestral strain: CC-1373, 13; CC-1952, 14; CC-2342, 14; CC-2344, 15; CC-2931, 14; CC-2937, 15.

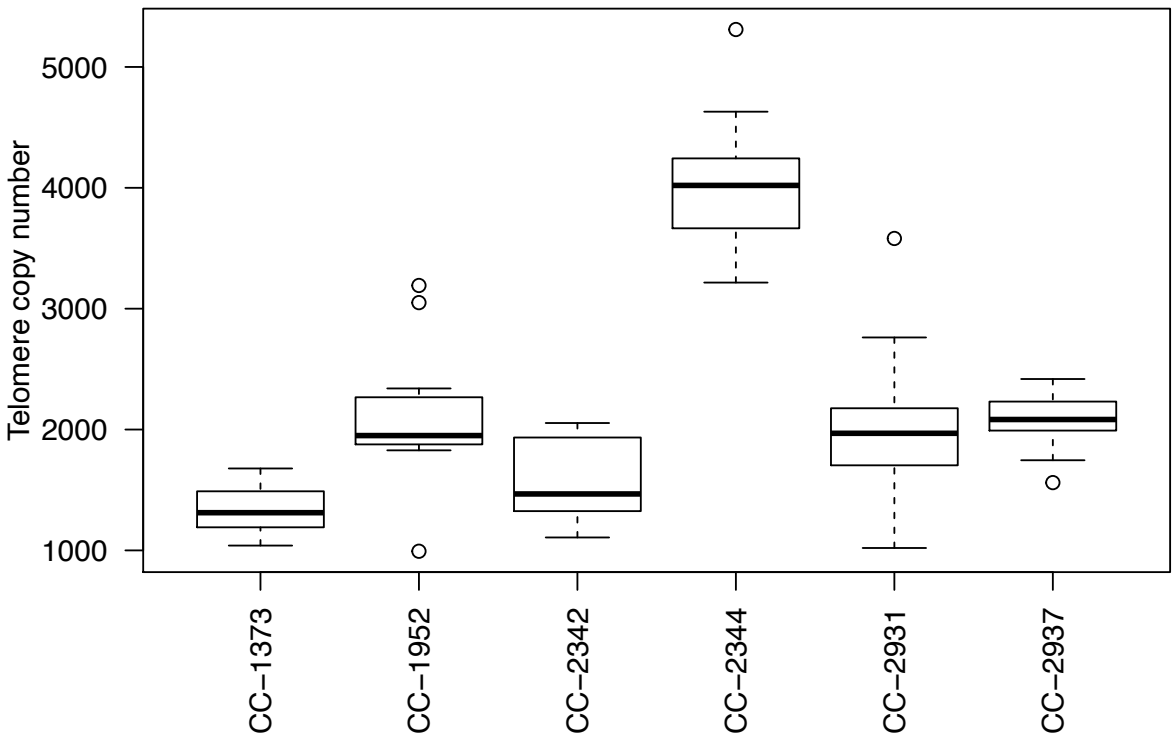

Figure S3.2. Boxplot of telomere repeat (AAAACCCT) copy number in the 13-15 MA lines of each of the *C. reinhardtii* ancestral strains. Strain CC-2344 has a significantly higher copy number than the other strains (ANOVA $p < 2 \times 10^{-16}$, posthoc Tukey test $p < 10^{-7}$).

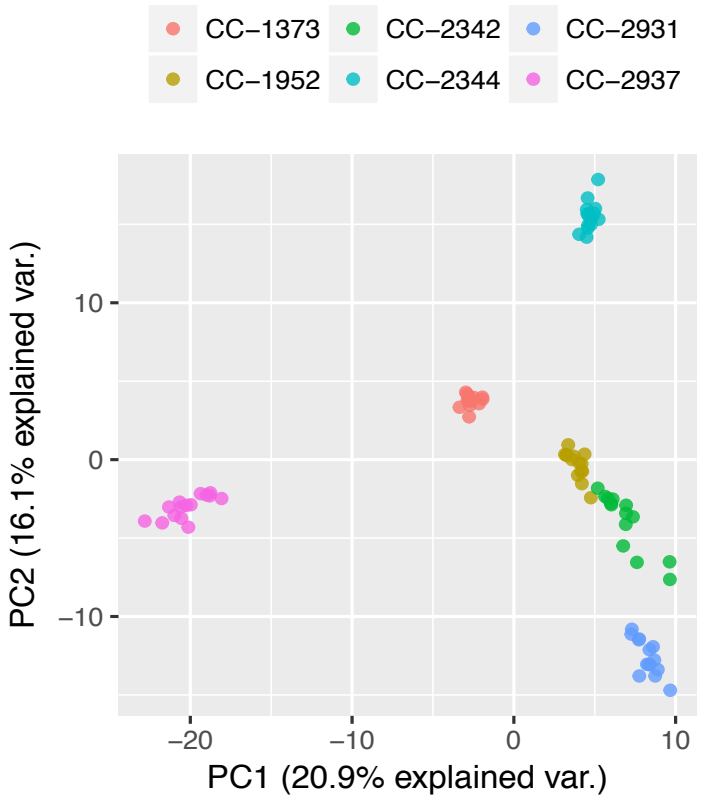

Figure S3.3. Principal components analysis (PCA) of kmers >= 2 copies per ancestral strain (including kmers that are absent from some ancestors). Repeats assessed from sequenced mutation accumulation (MA) lines cluster by ancestral strain on PC1 and PC2. Each point represents a single line. Colors represent the ancestral strains. Number of lines per ancestral strain: CC-1373, 13; CC-1952, 14; CC-2342, 14; CC-2344, 15; CC-2931, 14; CC-2937, 15.

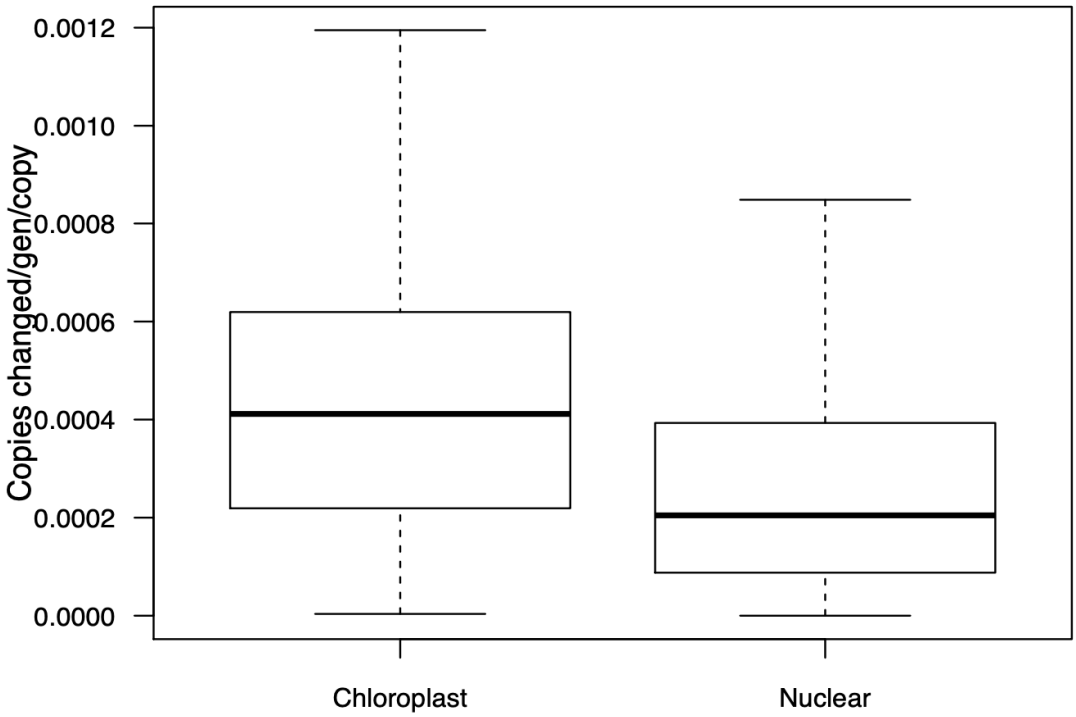

Figure S3.4. Mutation rates of copy number change in inferred in long AT-rich (putative chloroplast) repeats versus inferred nuclear repeats. Repeats were inferred to be putative chloroplast repeats if they had a GC content < 35% and were of 10 bp unit length or longer. ANOVA indicated a significant difference between these two groups ($p = 2.5 \times 10^{-11}$).

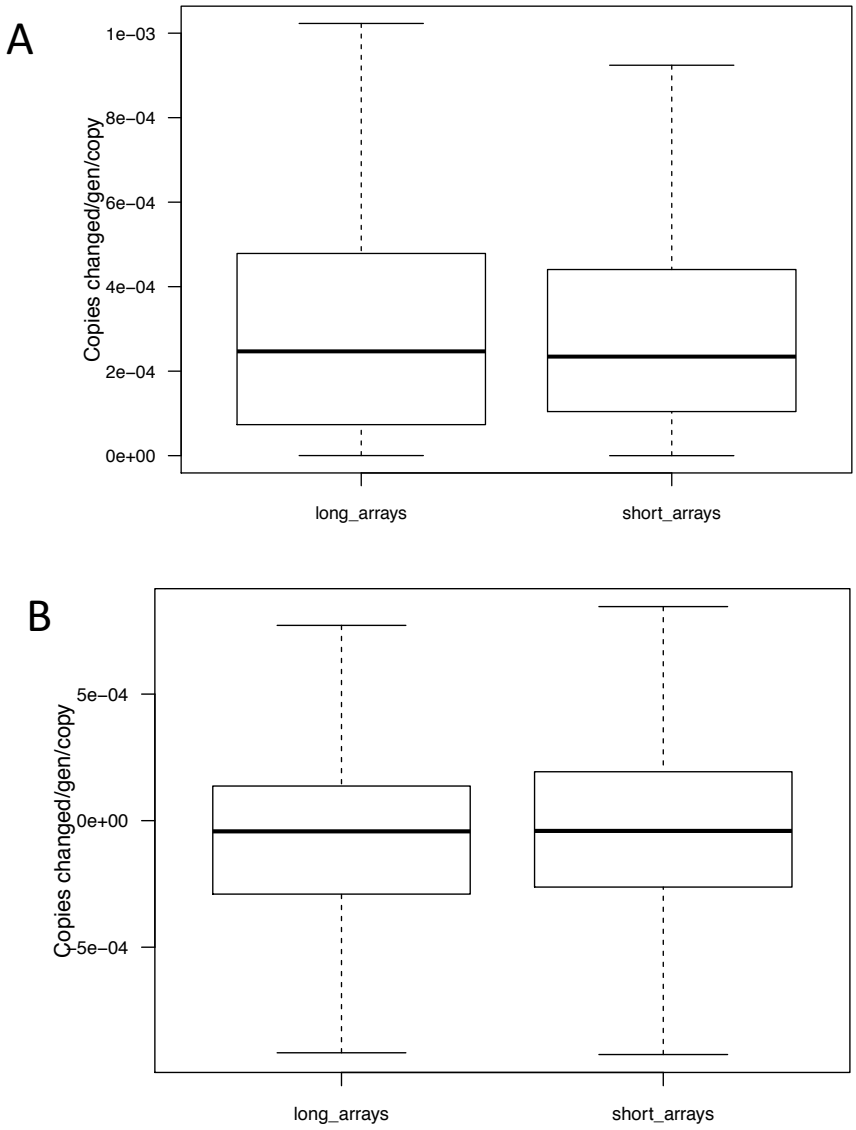

Figure S3.5. Mutation rates of copy number change in kmers in long arrays versus those in short arrays. A) Shows no significant difference in the absolute mutation rate (ANOVA P = 0.87) and B) shows no significant difference in the directional (+ expansions, - contractions) mutation rate (ANOVA P = 0.92). Kmers were determined to be in long arrays if some reads from both mate pairs of paired-end reads contained the same kmer.

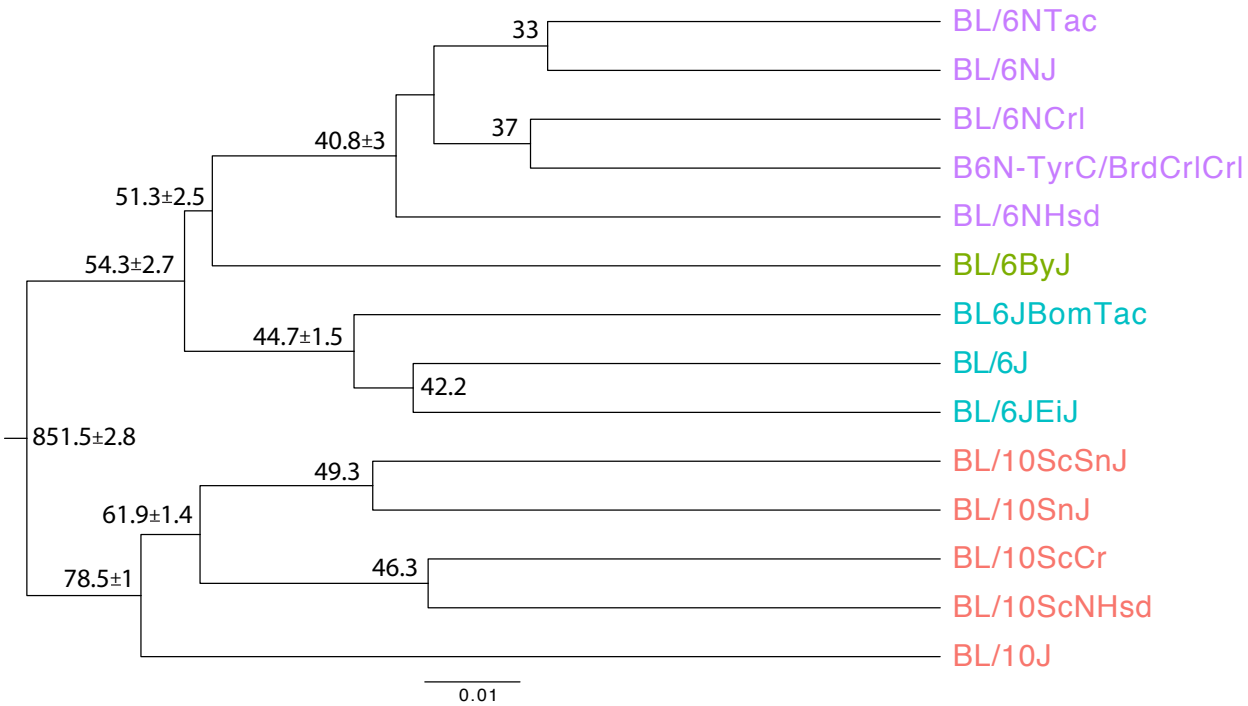

Figure S4.1 Phylogenetic reconstruction from SNP data. Variants were called with GATK and the R package SNPRelate was used to calculate a distance matrix based on identity-by-state, followed by hierarchical clustering and producing a tree. The SNP data was able to distinguish the B6 and B10 clades, but not all relationships within them. Branch length estimates in generations are written at each node. We used the tree produced here to estimate the branch length between B6 and B10 as ~850 generations.

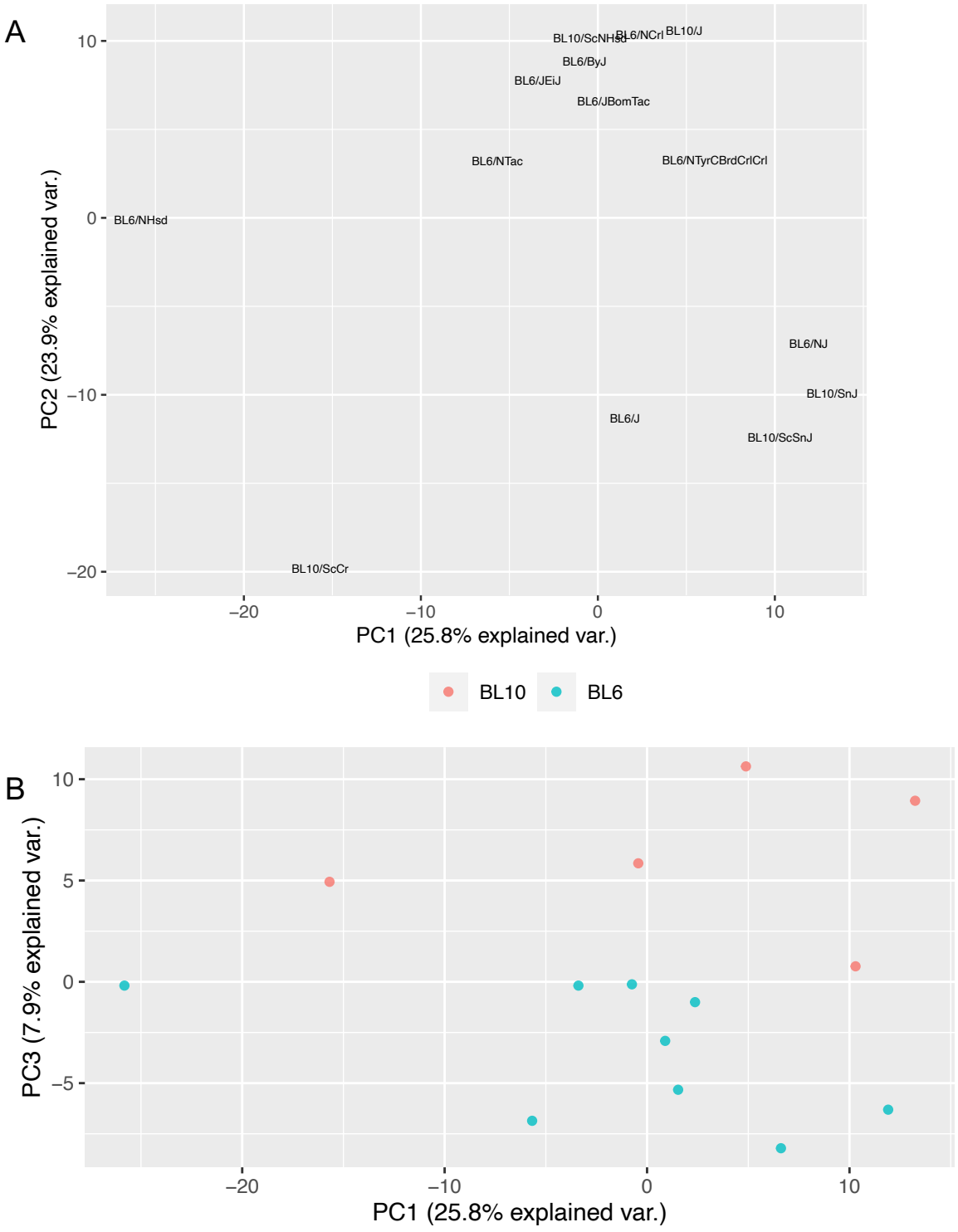

Figure S4.2. Principal components analysis (PCA) of kmer abundances. A) PC1 and PC2 do not show separation between the two major clades BL6 and BL10. B) PC1 and PC3; there is some grouping of BL6 and BL10 clades on PC3.

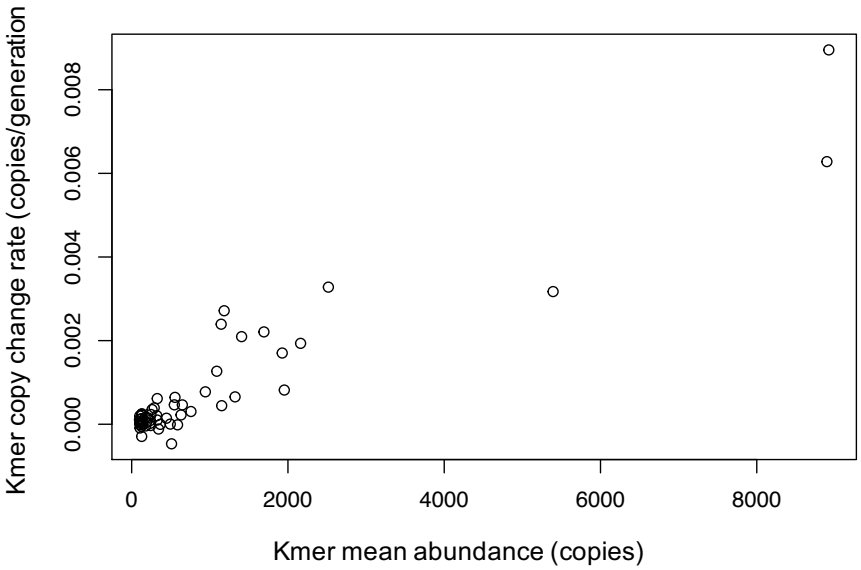

Figure S4.3. The copy number change rate is positively correlated with the kmer's mean abundance. The 6 most abundance kmers were removed from the plot to reduce skewing.

Figure S5.1. NCRF Simulation results. We ran NCRF with a range of max error cutoffs on our PBSIM simulated PacBio reads from a Mock genome. The mock genome contained 109 Mb AAACTAC, 14 Mb AAACTAT, and 17 Mb AAATTAC.

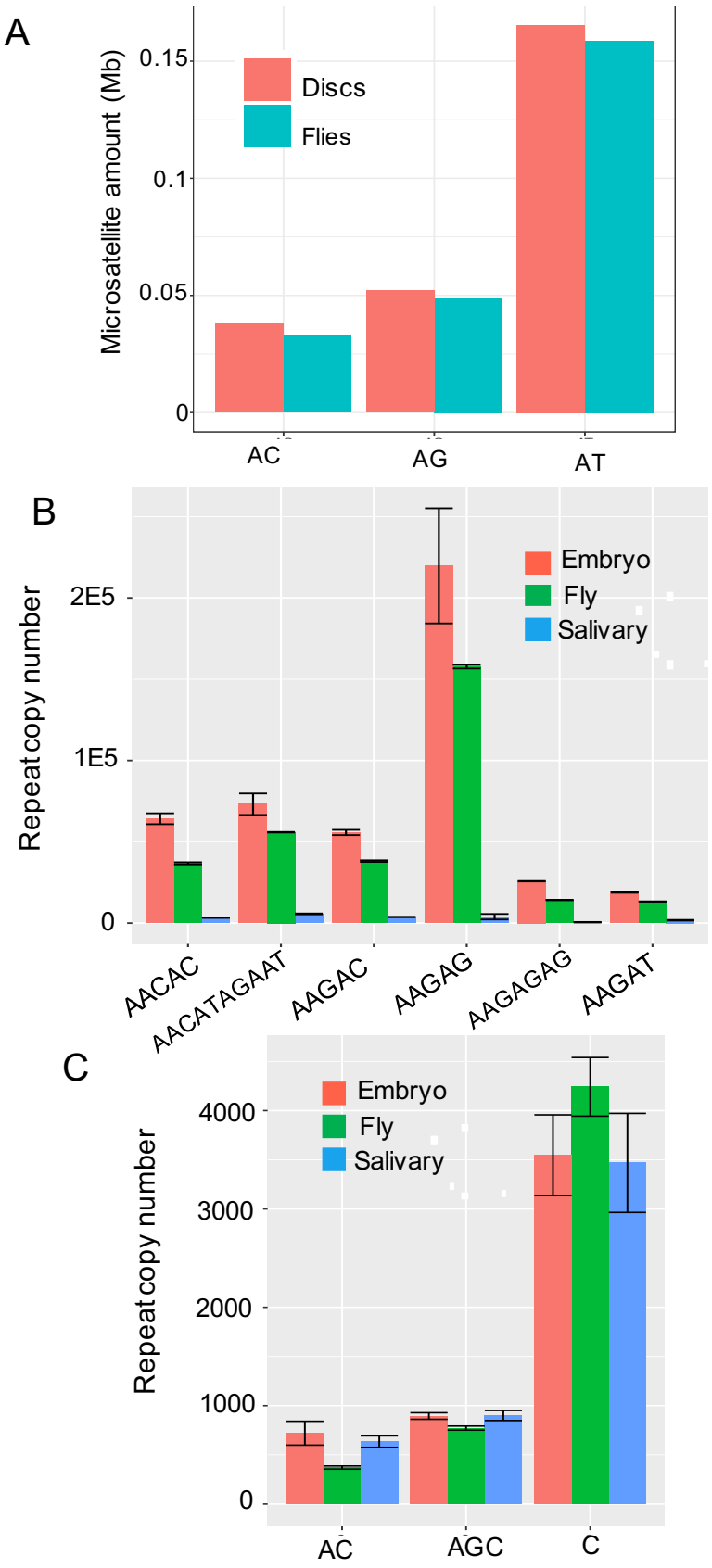
A
Microsatellite amount (Mb)
0.15
0.1
0.05
0
Discs
Flies
AC
AG
AT
B
Repeat copy number
2E5
1E5
0
Embryo
Fly
Salivary
AACAC
AACATAGAAT
AAGAC
AAGAG
AAGAGAG
AAGAT
C
Repeat copy number
4000
3000
2000
1000
0
Embryo
Fly
Salivary
AC
AGC
C

Figure S5.2. Polyteny vs. diploid analysis. (A) Microsatellites are not exclusively present in pericentromeric heterochromatin like the abundant AAACTAC family satellites. Therefore, we would not expect a difference in abundance between imaginal disc and polytene tissues. This serves as a control to Fig 1D to show that there was not a global bias against simple repeats in the sequencing, and the effect on the heterochromatic satellites is truly due to polyteny. (B) Analysis of *D. melanogaster* sequencing data from embryos, flies, and salivary glands. Results are shown for several satellites present in the pericentromeric heterochromatin as in Jagganathan et al (2017). (C) Microsatellites in the same *D. melanogaster* data. Again, this acts as a control for Fig S2B because microsatellites are not exclusively present in pericentromeric heterochromatin like the abundant AAACTAC family satellites.

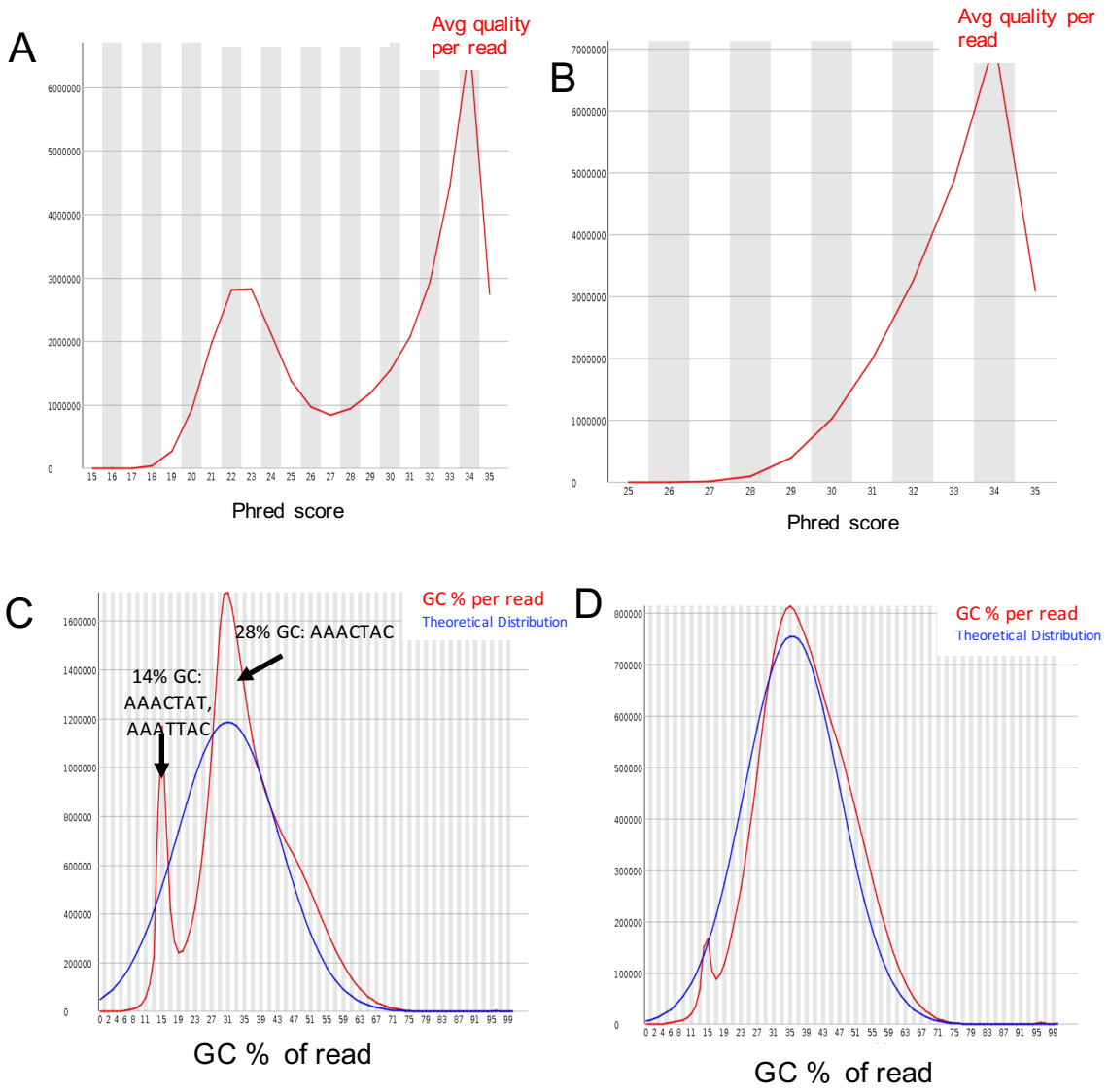

Figure S5.3. Satellite containing reads are enriched for low quality scores in Illumina data. A) Quality score distribution in the raw reads. (B) Quality score distribution after quality filtering. (C) GC distribution of raw reads. (D) GC distribution of reads after quality filtering.

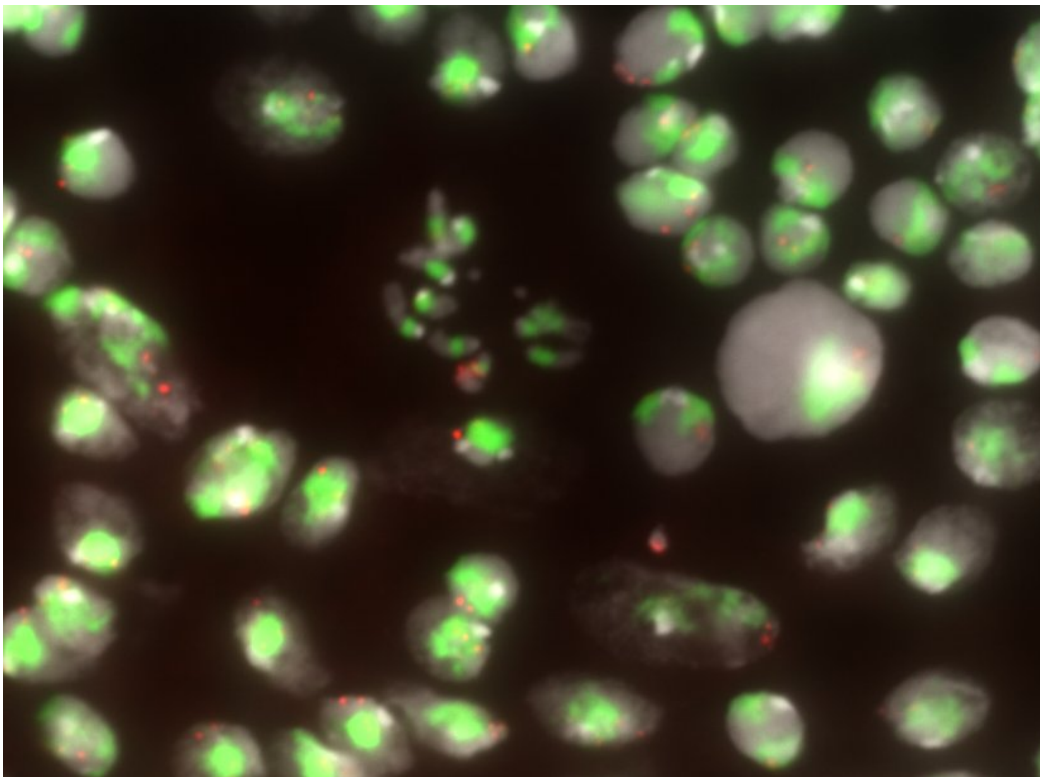

Figure S5.4. FISH image of *D. virilis* male. The AAACAAC satellite appears on a single chromosome pair. This is clear from the metaphase chromosomes (middle) as well as the interphase cells where two distinct foci are localized. The distinct foci in interphase cells is in contrast to the AAACTAC satellite, which takes up a large portion of the nucleus.

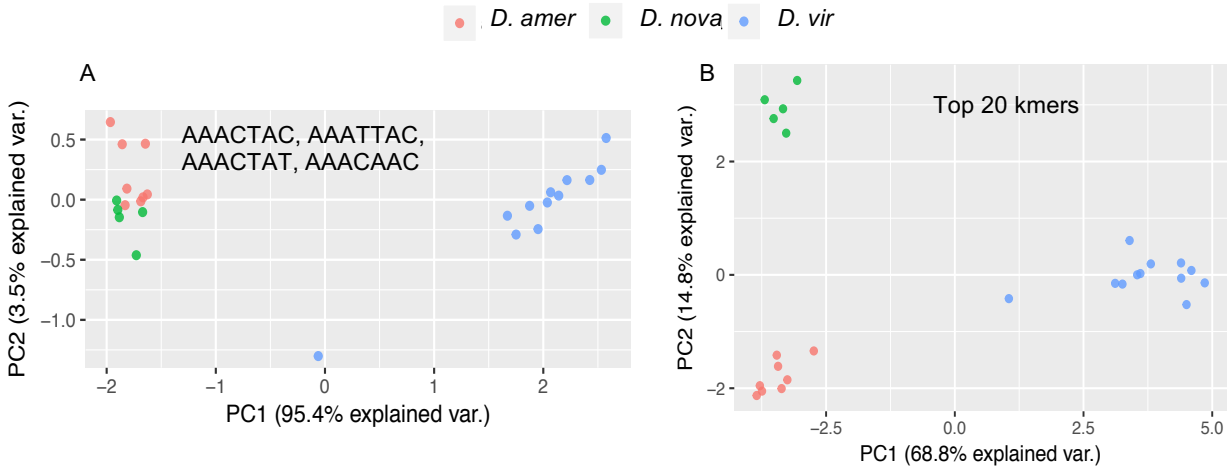

Fig S5.5. PCA of satellite DNA copy number of *D. virilis* group strains. (A) Using only the abundances of the satellites AAACTAC, AAATTAC, AAACTAT, AAACAAC. (B) Using the abundances of the 20 most abundant simple satellites.

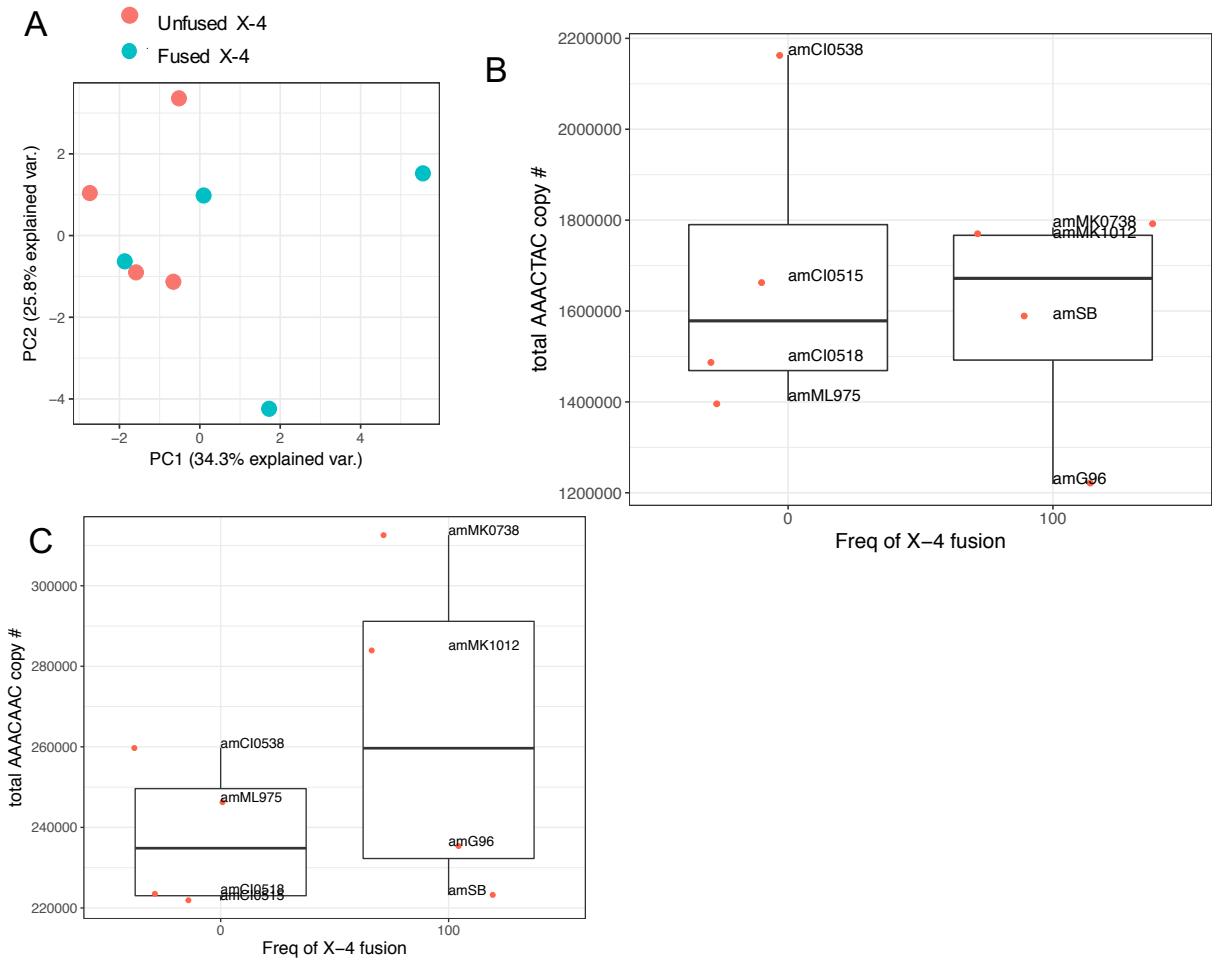

Figure S5.6. Comparison of *D. americana* strains with and without the X-4 chromosomal fusion. (A) PCA using the top 20 simple satellites. (B) Boxplot of AAACTAC copy number. (C) Boxplot of AAACAAC copy number.

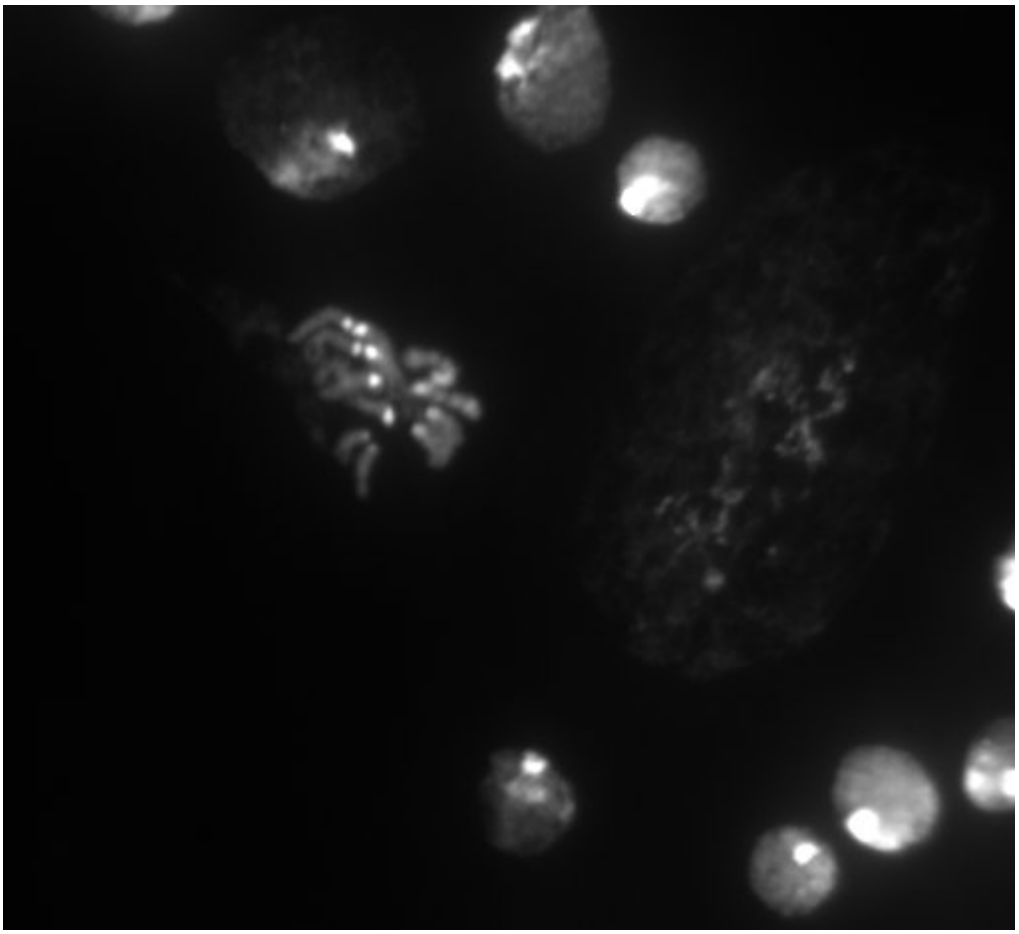

Figure S6.1. XXYY karyotype in vir00-Xfus female larva neuroblast.

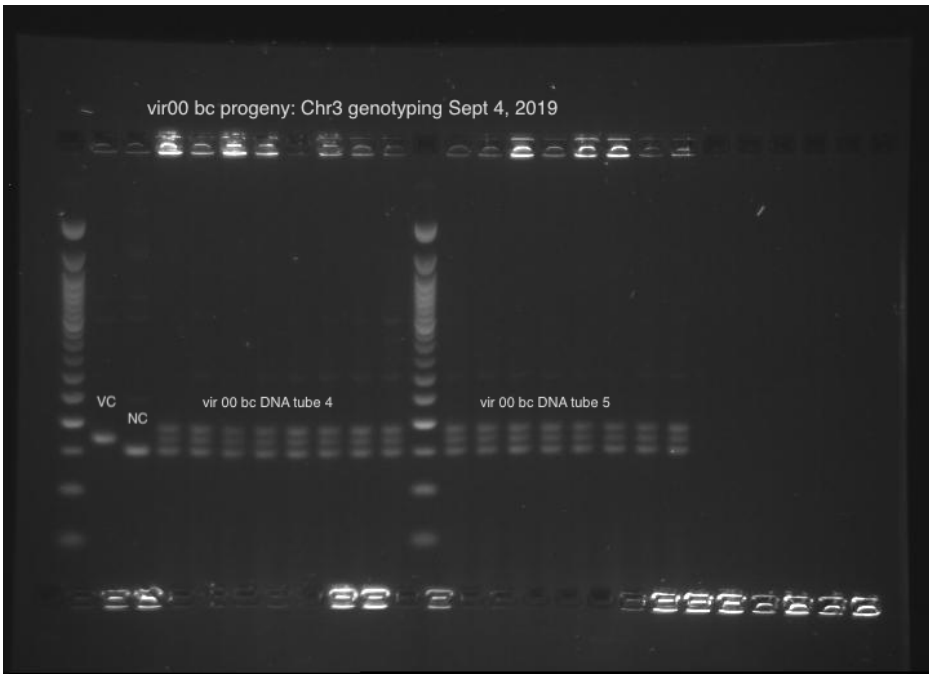

Figure S6.2. Sample genotyping results for Chr3. The first two lanes after the ladder are controls for the *D. virilis* and *novamexicana* allele size, respectively.

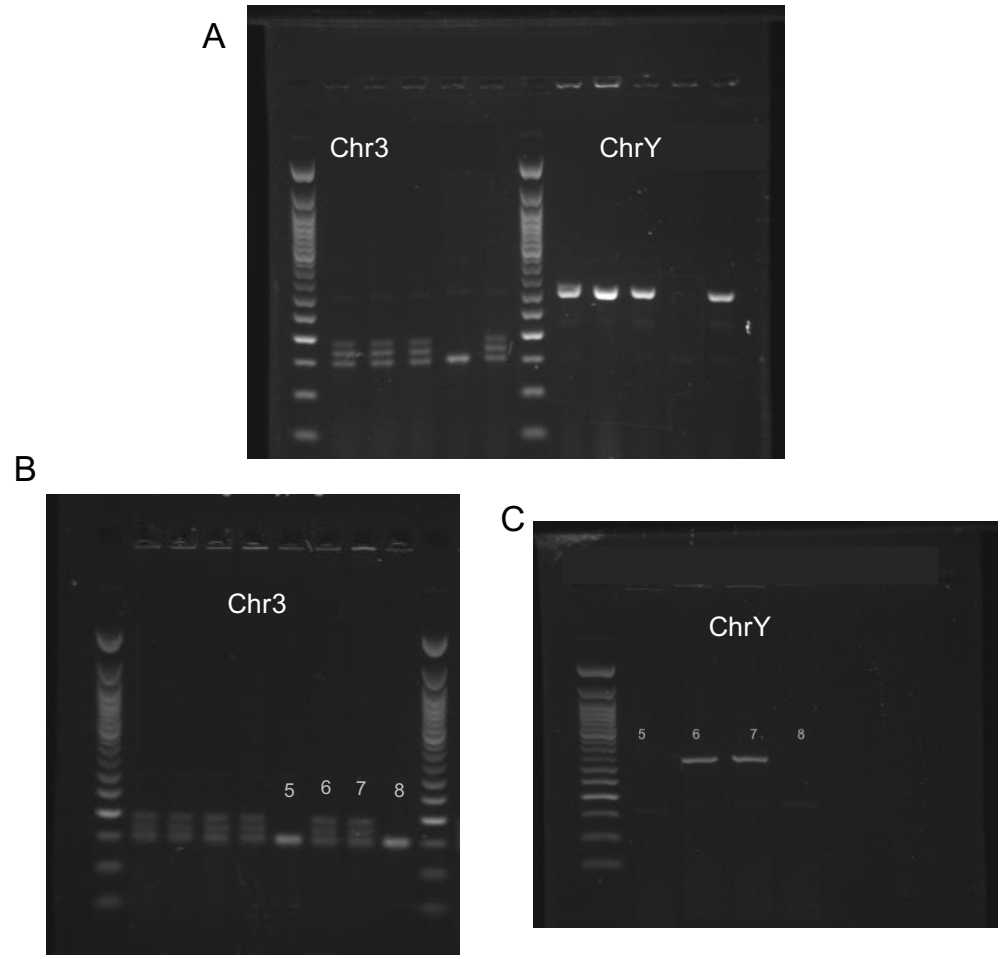

Figure S6.3. Nondisjunction occurred in lines that only had the D.nov Chr3 allele. A) side by side PCR of 5 samples for the Chr3 microsatellite locus (left panel) and ChrY (right panel). Sample 4 has only the D.nov allele and has no Y chromosome, indicating a nondisjunction event. B) PCR of 8 samples for the Chr3 microsatellite locus. Samples 5 and 8 have only the D.nov Chr3 allele. C) PCR of samples 5-8 for ChrY. Samples 5 and 8 have no Y chromosome, indicating a nondisjunction event.

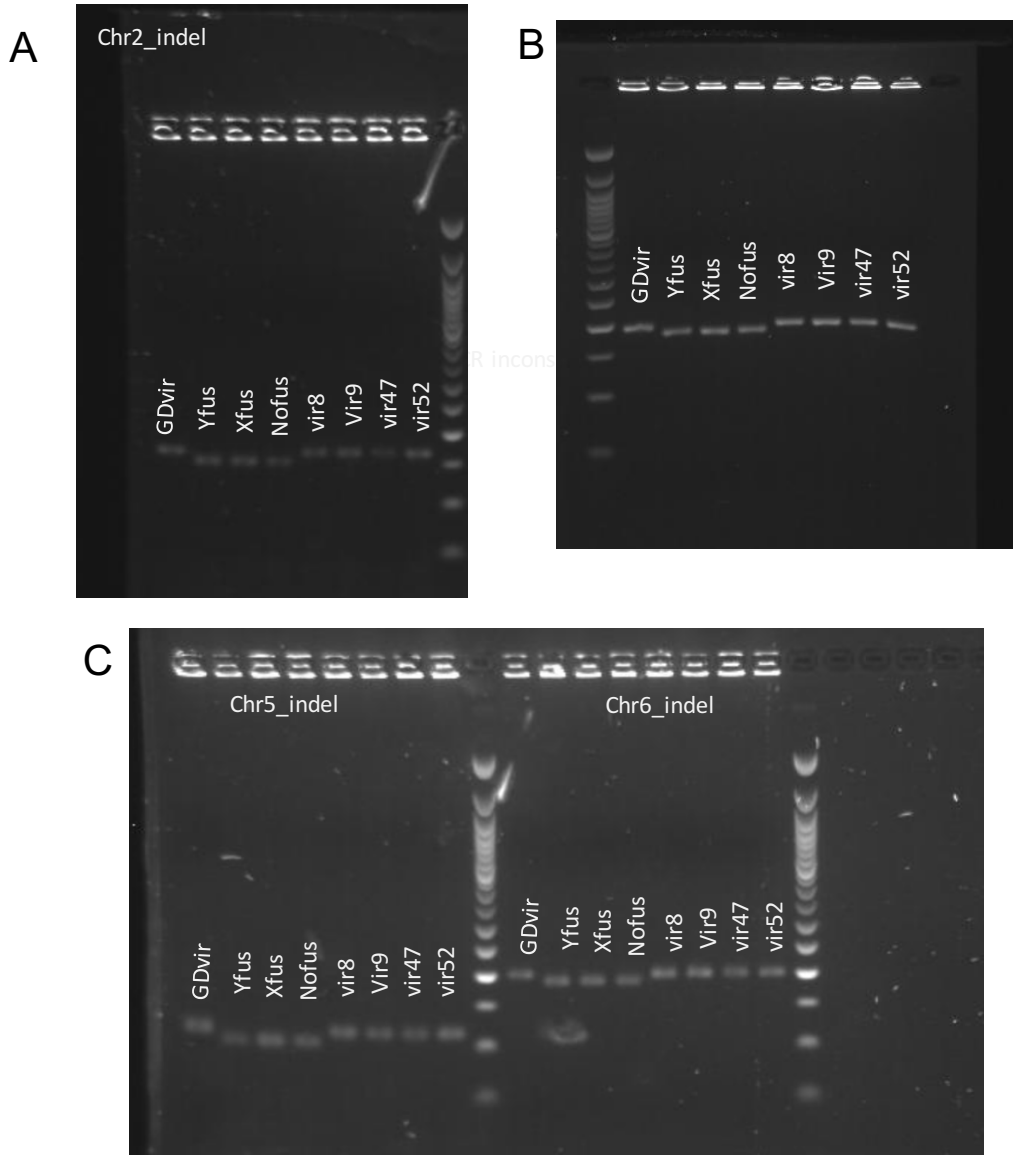

Figure S6.4. Indel validation of vir00 substrains. Gels demonstrating amplification of loci on four chromosomes that were found to contain a small deletion unique to vir00. The three substrains of vir00 (Yfus, Xfus, Nofus) all contain the deletions, shown by a slight band shift, on each of the loci.

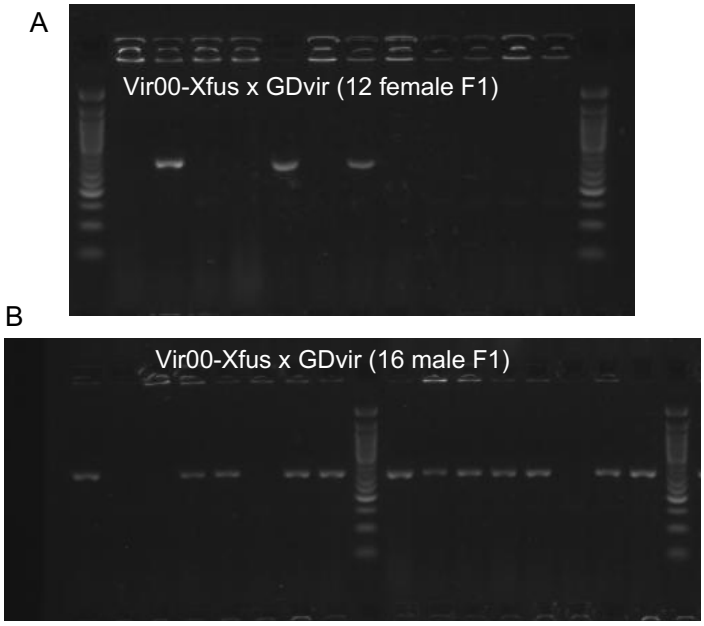

Figure S6.5. X-Y nondisjunction in a vir00-Xfus male. The focal male was crossed to two virgin Gdvir females. A) 3/12 female progeny contain a Y chromosome (XXY karyotype), indicating XY sperm from the father. B) 4/16 male progeny do not contain a Y chromosome (XO karyotype), indicating nullisomic sperm from the father.

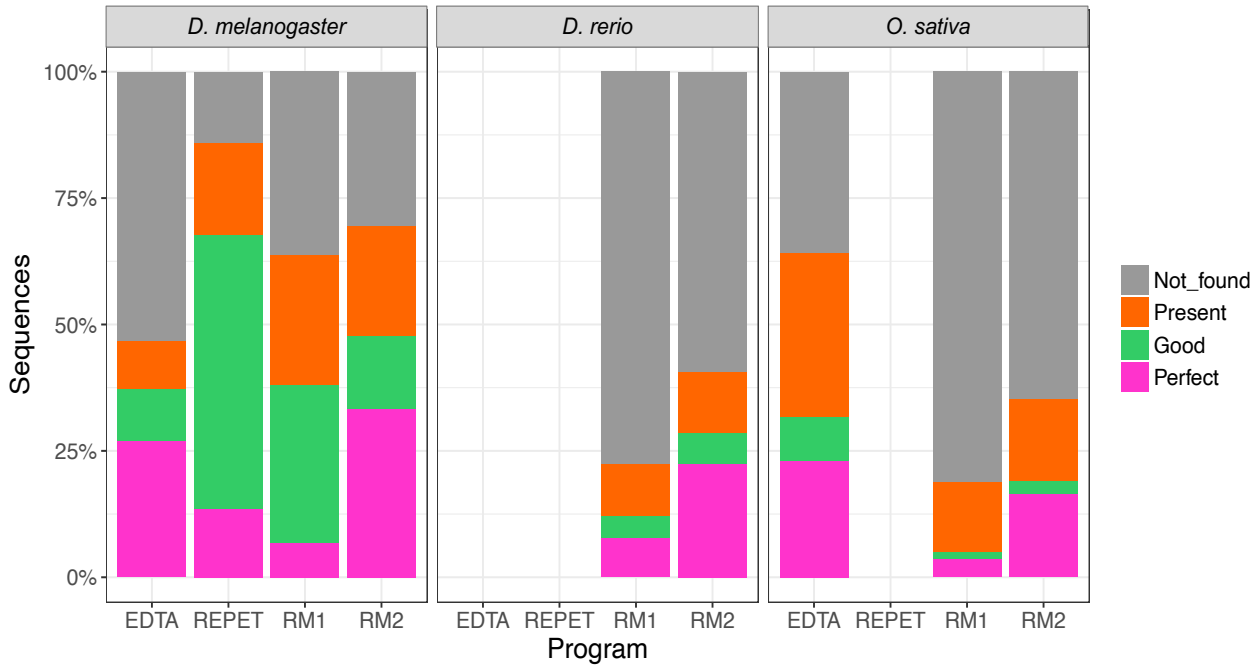

Figure S7.1. Comparison of family quality including EDTA and REPET for *D. melanogaster* and EDTA for *O. sativa.*

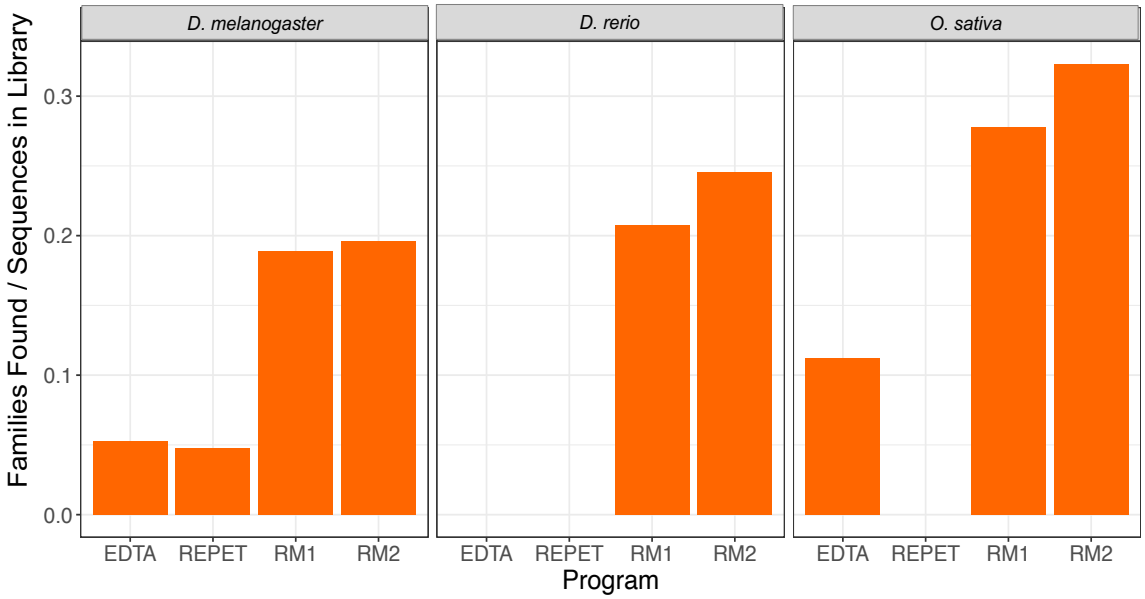

Figure S7.2. RepeatModeler2 has the highest families found: sequences in library ratio.