SUPPORTING MATERIAL

File S1. ***Demographic simulation methods.*** The following command lines were used to generate simulated data under demographic models. "msafe" refers to a version of ms (HUDSON 2002) modified to implement founder events with multiply mated females (courtesy of R. Hudson) and instantaneous admixture, available from A. Kern. Pairs of command lines are given for (1) the simulations used for diversity ratio analyses, and (2) the simulations used for LD analyses. Command lines are given separately for X-linked and autosomal loci. "+admix" refers to a modified version of each published model that includes 10% African admixture into the U.S. sample 1250 generations ago. Arguments after the admixture flag (-eA) refer to the timing (in coalescent units), destination population for admixture, source population for admixture, and admixture proportion. Arguments after the founder event flag (-eF) refer to timing, population affected, inheritance (0 for autosomal, 1 for X-linked), multiple mating parameter (number of males that each female in the founder party had mated with), the number of founder events, and then a series of numbers indicating the number of females present in each of those founder events. Further details regarding the founder event model can be found in (POOL and NIELSEN 2008). Additional information about these simulations, including the assumptions needed to implement each model, is given in the Materials and Methods section.

Simulations using the model of (THORNTON and ANDOLFATTO 2006).

X:

```
./msafe 39 10000 -t 8.4 -r 57.6 1000 -I 2 32 7 0 -en 0.0042 1 0.029 -en 0.0192 1 1
-ej 0.0192000001 1 2 ./msafe 32 1000 -t 194 -r 1328 23040 -eN 0.0042 0.029 -eN
0.0192 1
```

A:

```
./msafe 38 10000 -t 8.4 -r 38.1 1000 -I 2 32 6 0 -en 0.00315 1 0.029 -en 0.0144 1 1
-ej 0.0144000001 1 2 ./msafe 32 1000 -t 194 -r 879 23040 -eN 0.00315 0.029 -eN
0.0144 1
```

X+admix:

```
./msafe 39 10000 -t 8.4 -r 57.6 1000 -I 2 32 7 0 -eA 0.00067 1 2 0.1 -en 0.0042 1
0.029 -en 0.0192 1 1 -ej 0.0192000001 1 2 ./msafe 32 1000 -t 194 -r 1328 23040 -I 2
32 0 0 -eA 0.00067 1 2 0.1 -en 0.0042 1 0.029 -en 0.0192 1 1 -ej 0.0192000001 1 2
```

A+admix:

```
./msafe 38 10000 -t 8.4 -r 38.1 1000 -I 2 32 6 0 -eA 0.0005 1 2 0.1 -en 0.00315 1
0.029 -en 0.0144 1 1 -ej 0.0144000001 1 2 ./msafe 32 1000 -t 194 -r 879 23040 -I 2
32 0 0 -eA 0.0005 1 2 0.1 -en 0.00315 1 0.029 -en 0.0144 1 1 -ej 0.0144000001 1 2
```

Simulations using the models of (LI and STEPHAN 2006) and (HUTTER *et al.* 2007).

X:

```
./msafe 39 10000 -t 37 -r 254 1000 -c 5 86.5 -I 2 32 7 0 -en 0 1 0.124 -en
0.0044912 1 0.000256 -en 0.00459 1 1 -ej 0.00459000001 1 2 -eN 0.0174 0.2 ./msafe
32 1000 -t 852 -r 5842 23040 -c 5 86.5 -eN 0 0.124 -eN 0.0044912 0.000256 -eN
0.00459 1 -eN 0.0174 0.2
```

A:

```
./msafe 38 10000 -t 37.6 -r 171 1000 -c 5 86.5 -I 2 32 6 0 -en 0 1 0.183 -en
0.0037281 1 0.000377 -en 0.00381 1 1 -ej 0.00381000001 1 2 -eN 0.0145 0.2 ./msafe
32 200 -t 866 -r 3933 23040 -c 5 86.5 -eN 0 0.183 -eN 0.0037281 0.000377 -eN
0.00381 1 -eN 0.0145 0.2
```

X+admix:

```
./msafe 39 10000 -t 37 -t 852 -r 5842 -c 5 86.5 -I 2 32 7 0 -en 0 1 0.124 -eA
0.0000363 1 2 0.1 -en 0.0044912 1 0.000256 -en 0.00459 1 1 -ej 0.00459000001 1 2
-eN 0.0174 0.2 ./msafe 32 1000 -t 1150 -r 7880 23040 -c 5 86.5 -I 2 32 0 0 -en 0 1
0.124 -eA 0.0000363 1 2 0.1 -en 0.0044912 1 0.000256 -en 0.00459 1 1 -ej
0.00459000001 1 2 -eN 0.0174 0.2
```

A+admix:

```
./msafe 38 10000 -t 60.2 -r 171 1000 -c 5 86.5 -I 2 32 6 0 -en 0 1 0.183 -eA
0.0000301 1 2 0.1 -en 0.0037281 1 0.000377 -en 0.00381 1 1 -ej 0.00381000001 1 2
-eN 0.0145 0.2 ./msafe 32 1000 -t 866 -r 3933 23040 -c 5 86.5 -I 2 32 0 0 -en 0 1
0.183 -eA 0.0000301 1 2 0.1 -en 0.0037281 1 0.000377 -en 0.00381 1 1 -ej
0.00381000001 1 2 -eN 0.0145 0.2
```

Simulations using a model from (POOL and NIELSEN 2008).

X:

```
./msafe 39 10000 -t 8.4 -r 57.6 1000 -c 5 86.5 -I 2 32 7 0 -en 0 1 0.746 -eF
0.00597 1 1 5 4 4 1 1 1 -ej 0.0059700001 1 2 ./msafe 32 1000 -t 194 -r 1328 23040
-c 5 86.5 -eN 0 0.746 -eF 0.00597 1 1 5 4 4 1 1 1 -eN 0.0059700001 1
```

A:

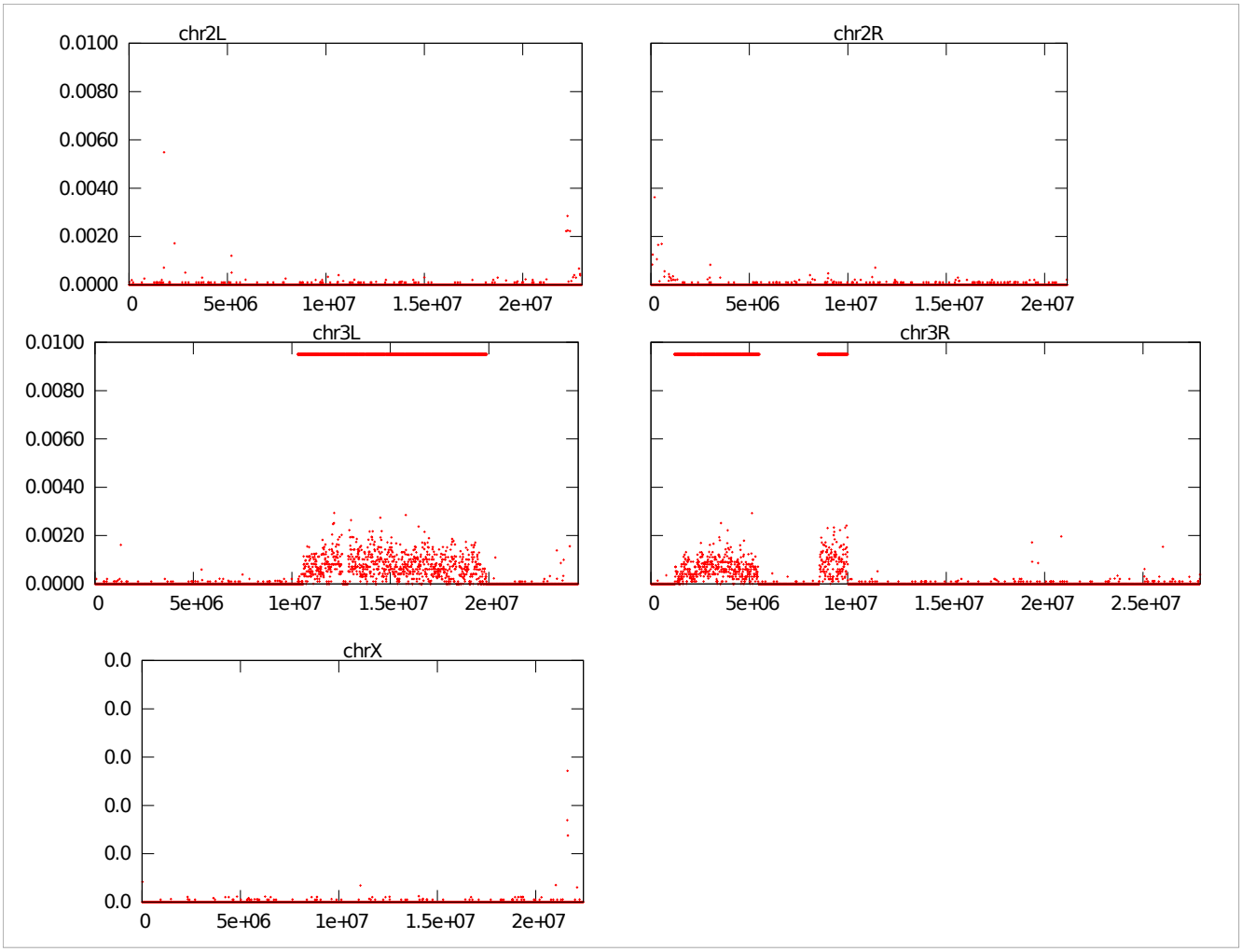
```
./msafe 38 10000 -t 8.4 -r 38.1 1000 -c 5 86.5 -I 2 32 6 0 -eF 0.005 1 0 5 4 4 1 1
1 -ej 0.0050000001 1 2 ./msafe 32 1000 -t 194 -r 879 23040 -c 5 86.5 -eF 0.005 1 0
5 4 4 1 1 1 -eN 0.005 1
```

X+admix:

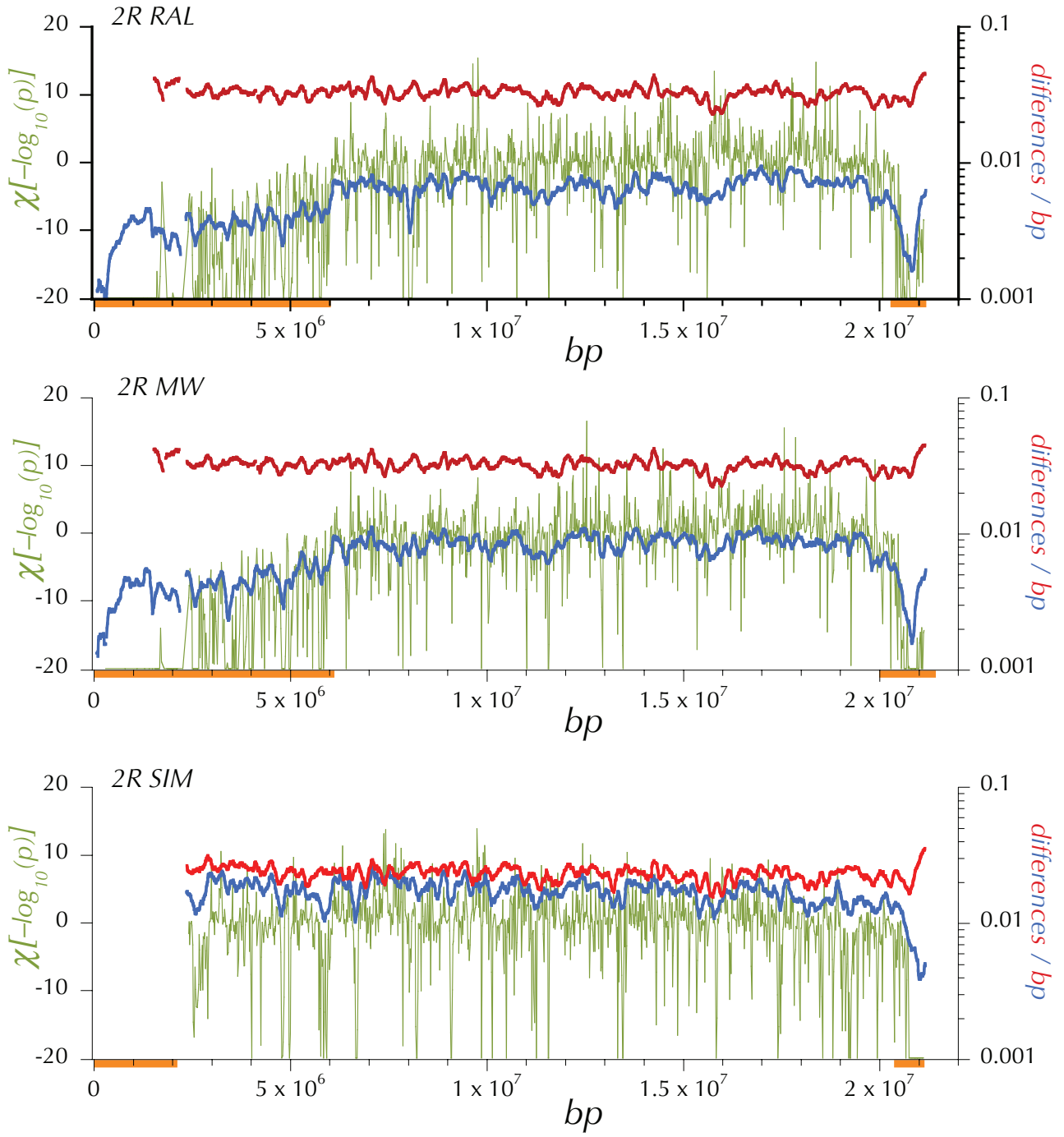
```
./msafe 39 10000 -t 8.4 -r 57.6 1000 -c 5 86.5 -I 2 32 7 0 -en 0 1 0.746 -eA 0.0008
1 2 0.1 -eF 0.00597 1 1 5 4 4 1 1 1 -ej 0.0059700001 1 2 ./msafe 32 1000 -t 194 -r
1328 23040 -c 5 86.5 -I 2 32 0 0 -en 0 1 0.746 -eA 0.0008 1 2 0.1 -eF 0.00597 1 1 5
4 4 1 1 1 -ej 0.0059700001 1 2
```

A+admix:

```
./msafe 38 10000 -t 8.4 -r 38.1 1000 -c 5 86.5 -I 2 32 6 0 -eA 0.0005 1 2 0.1 -eF
0.005 1 0 5 4 4 1 1 1 -ej 0.0050000001 1 2 ./msafe 32 1000 -t 194 -r 879 23040 -c 5
86.5 -I 2 32 0 0 -eA 0.0005 1 2 0.1 -eF 0.005 1 0 5 4 4 1 1 1 -ej 0.0050000001 1 2
```



Figure_S1. — Regions of residual heterozygosity in *RAL-335_1*. The proportion of sites called as heterozygotes in the "diploid" assembly *RAL-335_1* in 100 kbp windows plotted every 5 kbp on the major euchromatic chromosome arms. The bars of at the top indicate the segments designated as "residually heterozygous" and thus filtered from further analyses.



Figure_S2. — Expected heterozygosity, divergence and *HKAI* on the chr2R for the North American (RAL), African (MW) and simulans (SIM) samples. The (blue) expected heterozygosity, π at the midpoint of 150 kbp windows (incremented every 10 kbp, minimun coverage = 0.25 and Q30 sequence). The (red) lineage specific, average Q30 divergence in 150 kbp windows (incremented every 10 kbp and minimum coverage of 0.25). A preliminary application of *HKAI* on the Q30 data in windows of 4096 contiguous polymorphic or divergent sites identified centromere- and telomere-proximal regions (orange bars) in which the each window exhibited a deficiency of polymorphic sites relative to the chromosome-arm arm average. Then *HKAI* was applied again on the Q30 data in windows of 512 contiguous polymorphic or divergent sites (excluding these centromere-and telomere-proximal regions from calculation of the chromosome-arm-wide expected proportions, p_c and d_c. The (olive) $\chi[\log(p_{HKAI})]$ is the log of the p-value associated with *HKAI* plotted with the sign of the difference between the observed number and the expected number of polymorphic sites in the window.

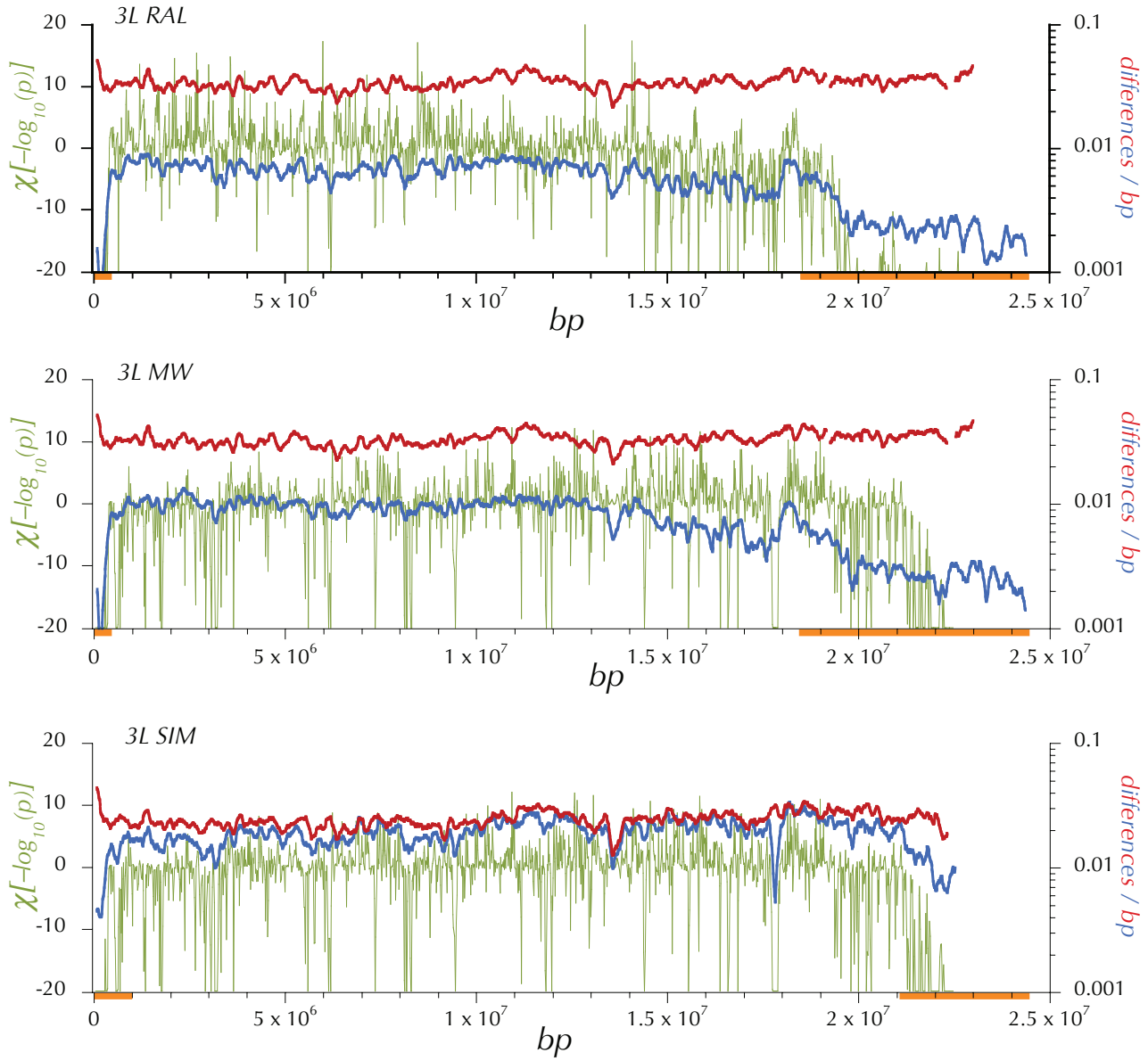


Figure S3. — Expected heterozygosity, divergence and *HKAI* on the chr3L for the North American (RAL), African (MW) and simulans (SIM) samples. The (blue) expected heterozygosity, π at the midpoint of 150 kbp windows (incremented every 10 kbp, minimun coverage = 0.25 and Q30 sequence). The (red) lineage specific, average Q30 divergence in 150 kbp windows (incremented every 10 kbp and minimum coverage of 0.25). A preliminary application of *HKAI* on the Q30 data in windows of 4096 contiguous polymorphic or divergent sites identified centromere- and telomere-proximal regions (orange bars) in which the each window exhibited a deficiency of polymorphic sites relative to the chromosome-arm arm average. Then *HKAI* was applied again on the Q30 data in windows of 512 contiguous polymorphic or divergent sites (excluding these centromere-and telomere-proximal regions from calculation of the chromosome-arm-wide expected proportions, p_c and d_c. The (olive) $\chi[\log(p_{HKAI})]$ is the log of the p-value associated with *HKAI* plotted with the sign of the difference between the observed number and the expected number of polymorphic sites in the window.

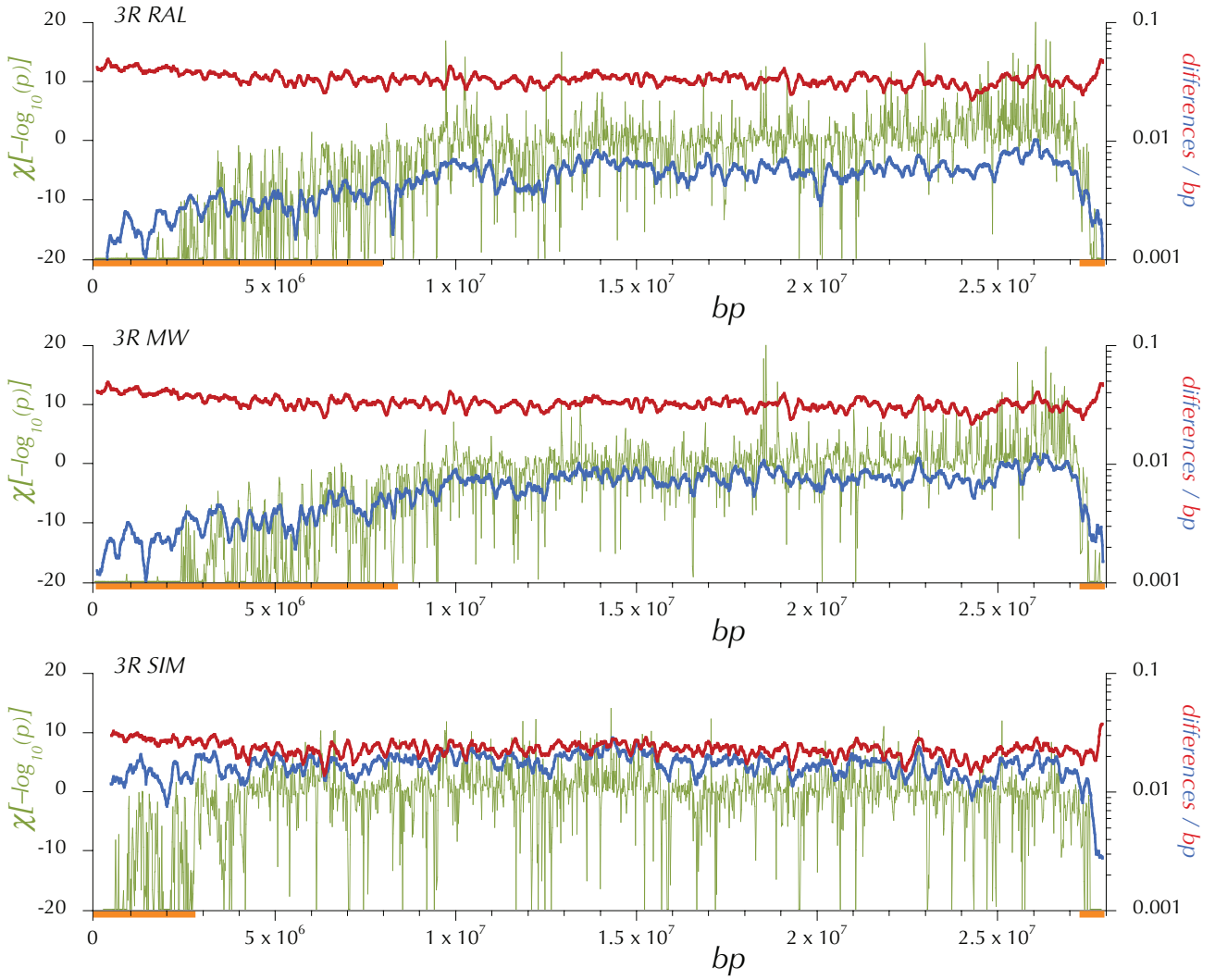


Figure S4. — Expected heterozygosity, divergence and *HKAI* on the chr3R for the North American (RAL), African (MW) and simulans (SIM) samples. The (blue) expected heterozygosity, π at the midpoint of 150 kbp windows (incremented every 10 kbp, minimun coverage = 0.25 and Q30 sequence). The (red) lineage specific, average Q30 divergence in 150 kbp windows (incremented every 10 kbp and minimum coverage of 0.25). A preliminary application of *HKAI* on the Q30 data in windows of 4096 contiguous polymorphic or divergent sites identified centromere- and telomere-proximal regions (orange bars) in which the each window exhibited a deficiency of polymorphic sites relative to the chromosome-arm arm average. Then *HKAI* was applied again on the Q30 data in windows of 512 contiguous polymorphic or divergent sites (excluding these centromere-and telomere-proximal regions from calculation of the chromosome-arm-wide expected proportions, p_c and d_c. The (olive) $\chi[\log(p_{HKAI})]$ is the log of the p-value associated with *HKAI* plotted with the sign of the difference between the observed number and the expected number of polymorphic sites in the window.

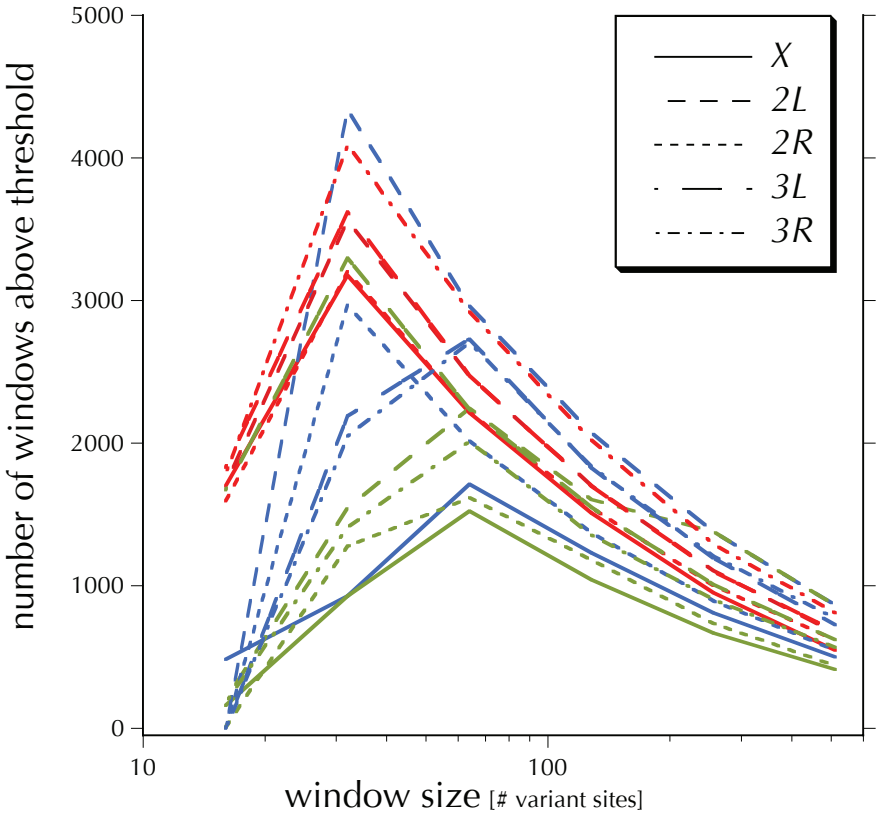


Figure S5. — False discovery rates for different *HKAI* window sizes, chromsome arm and samples of genomes. The numbers of windows, k with nominal p-values < k*0.05/n, (where n is the total number of windows on the chromosome arm) is plotted against window sizes: 16, 32, 64, 128, 256, and 512 bp for the African sample (MW, olive), North American (RAL, blue) and *simulans* (SIM, red). The different chromosome arms are plotted separately.

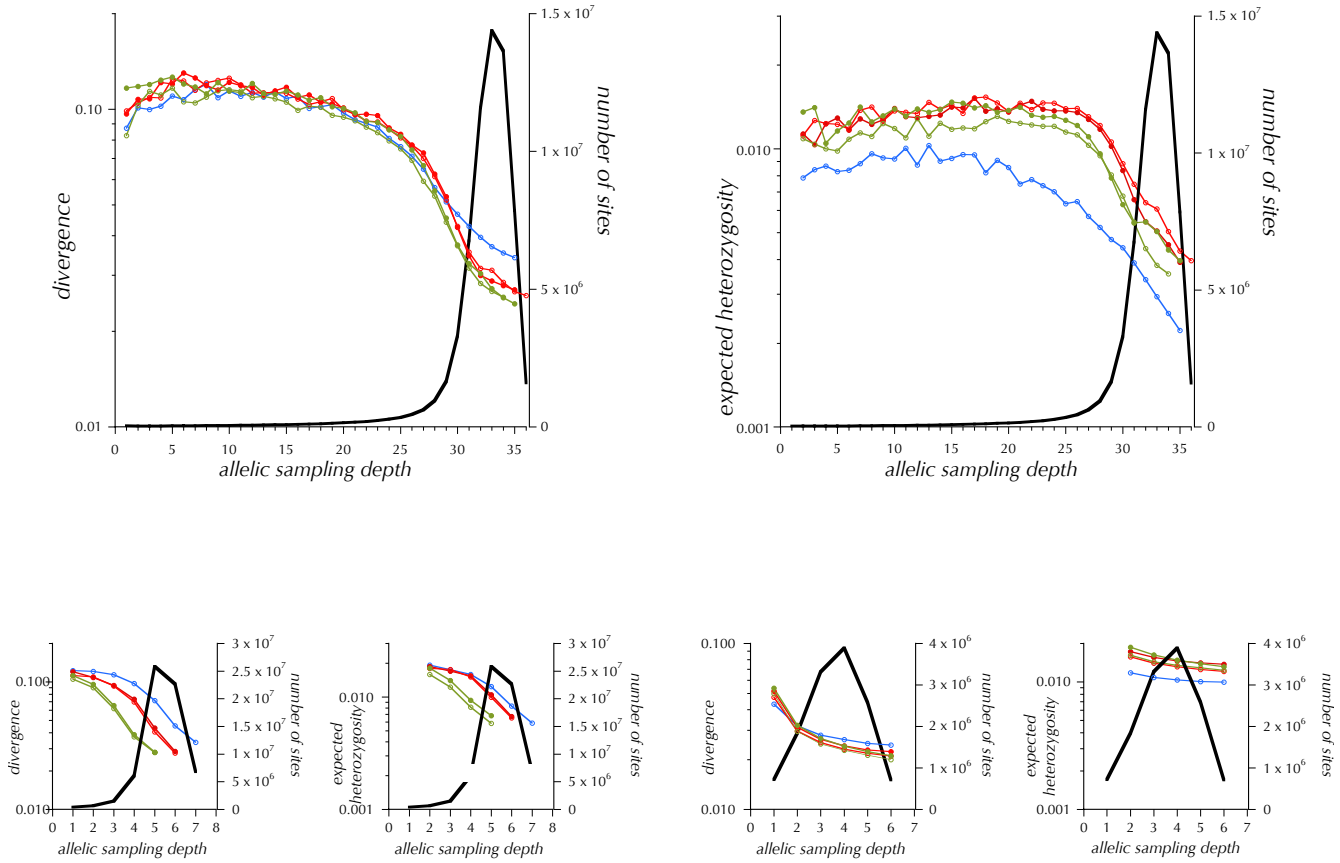


Figure S6. — The distributions of expected heterozygosity, divergence and number of sites at various (allele) sampling depths for Q30 data. Chromosome arms: X – blue, chr2 – red and chr3 – olive. Number of sites – black.

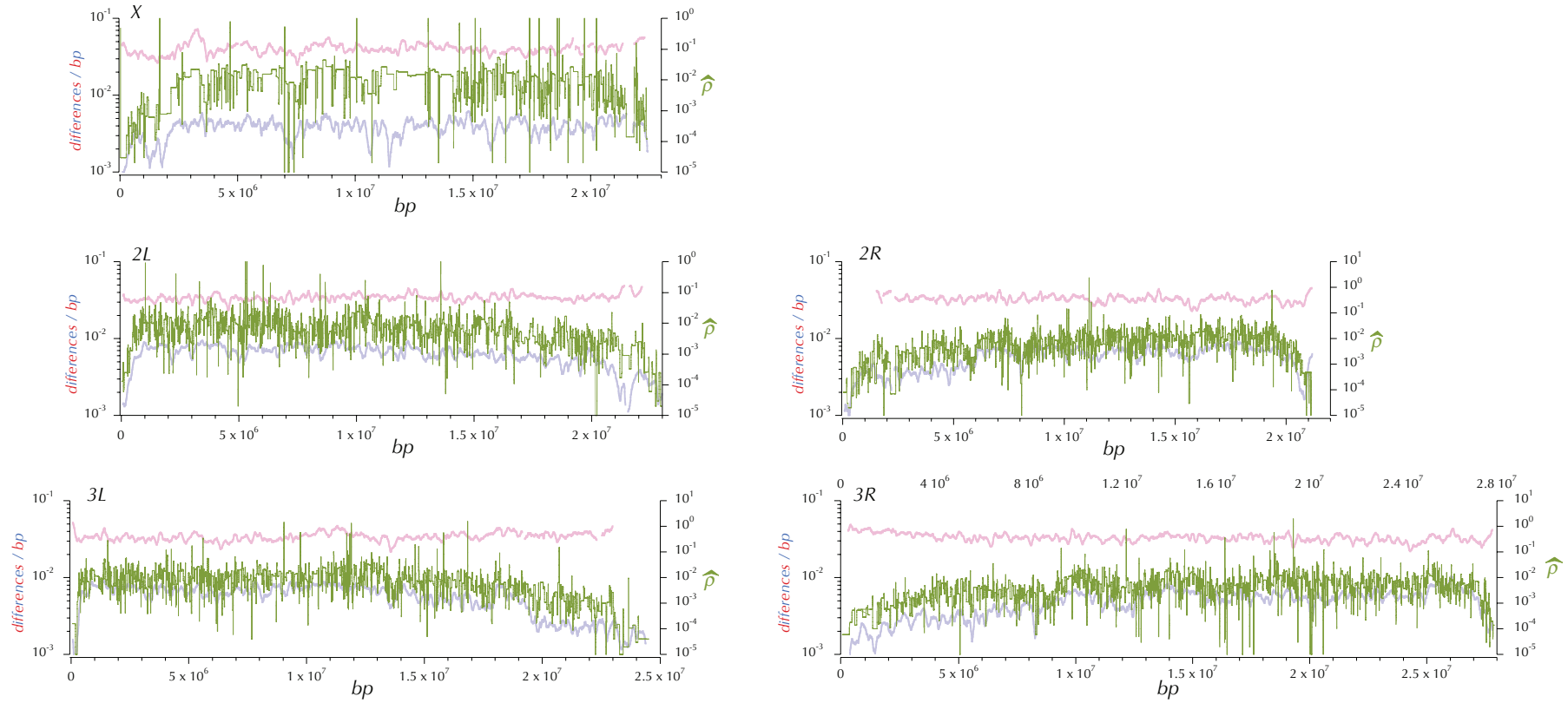


Figure S7. — The genomic distribution of $\hat{\rho}$ (olive), an estimate of the population recombination parameter *2Nr* along with expected heterozygosity (blue, π_w) and lineage specific divergence (red, δ_w).

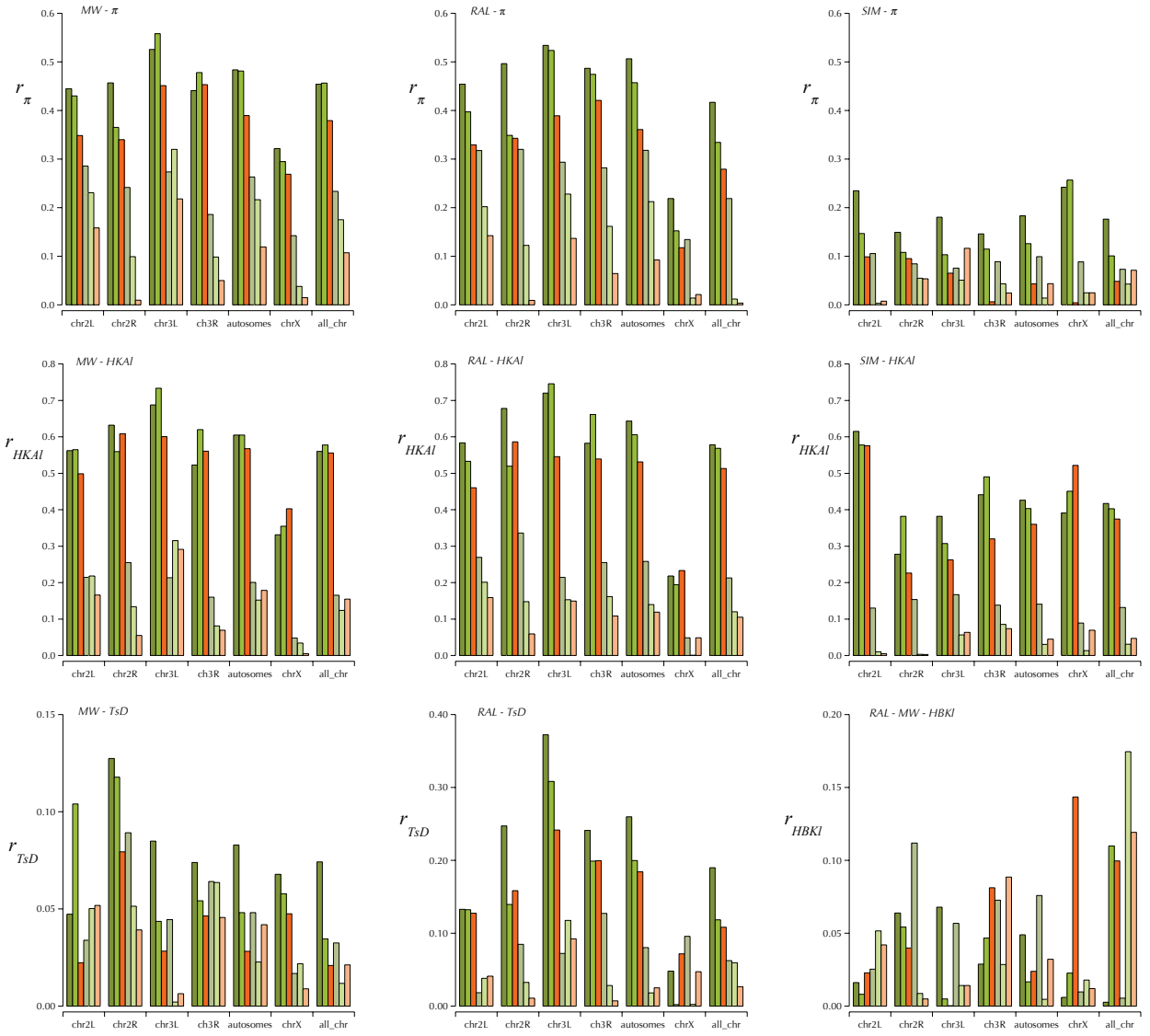


Figure S8. — Correlations between recombination rates and *HKAl*, π_w & *TsD*. *r* is the bp-weighted Pearsons correlation coefficient between *TsD*, π_w & *HKAl* and the logarithms of $\hat{\rho}_0$ (olive), $\hat{\rho}_{15}$ (light olive) and $\hat{r}_{15}$ (orange) across the autosomes and X chromosome. The three lower columns to the right (lighter shades) are the corresponding correlations for the "trimmed" euchromatic regions.

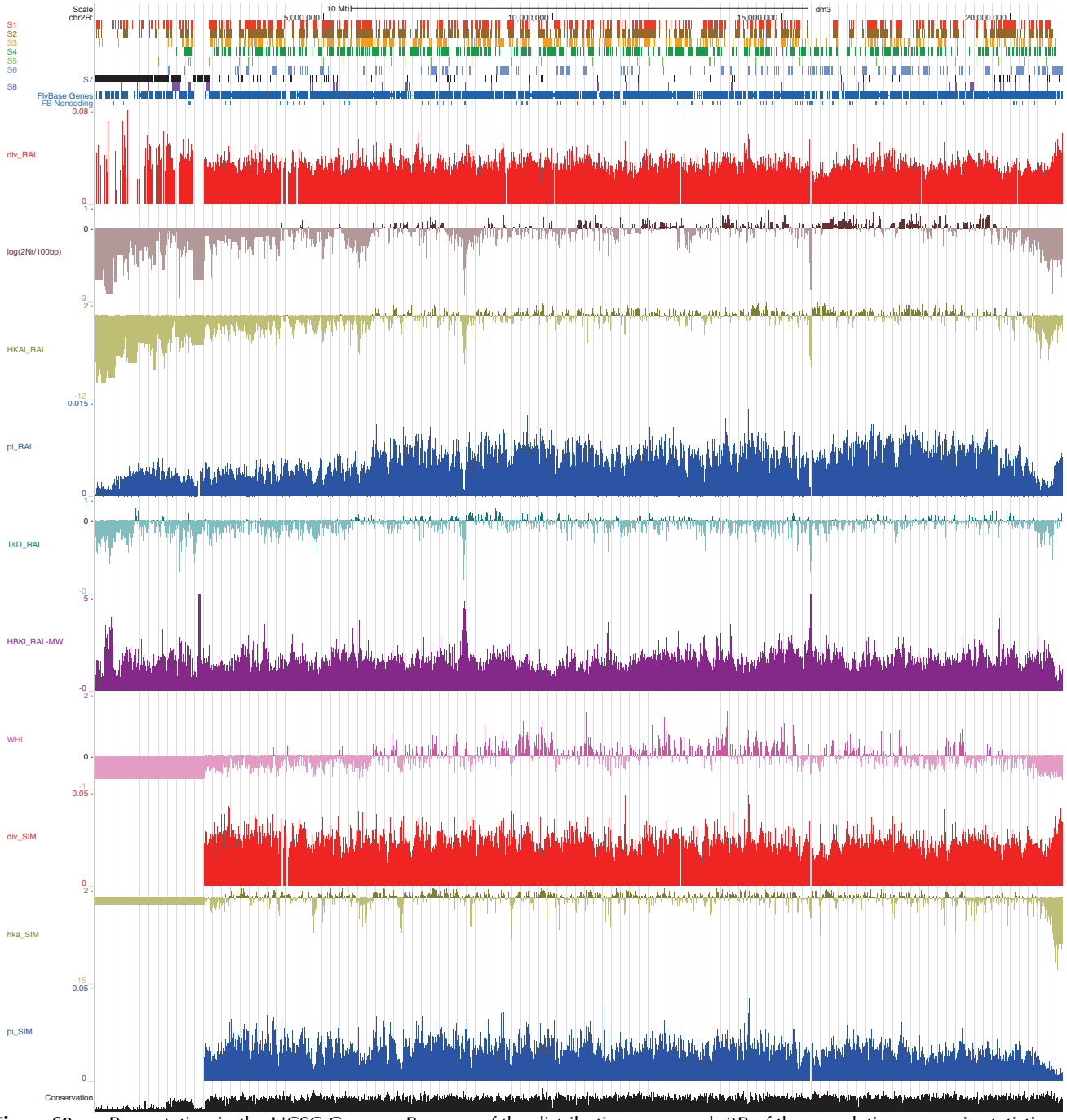


Figure S9. — Presentation in the UCSC Genome Browser of the distributions across chr2R of the population genomic statistics, π_w, δ_w, $\hat{\rho}$, *HKAl* and *TsD* for the RAL sample as well as *HBKl*, and *WHl*. At the top are *chromatin states 1* through *8*, followed by the Flybase annotated protein coding genes and then by noncoding genes. The eleventh track (down) begins the "custom" tracks from this study with $\log(\hat{\rho}/100)$. At the bottom are three standard UCSC Genome Browser annotation tracks, phylogenetic "Conservation," "RepeatMasker" and "Simple Repeats." Note the large reduction in $\log(\hat{\rho}/100)$, estimated $2Nr$ in the centromere-proximal 6 Mbp and the distal, telomere-proximal 1 Mbp. π_w (but not δ_w and thus), *HKAl* , *TsD* and *WHl* (not *HBKl*) all follow this pattern. In the intervening 14 Mbp of the euchromatic chr2R the patterns are on smaller scales. The region starting at 8 Mbp is an example local patterns that likely reflect the recent evolution of 31 genes in the cluster including strong geographic differentiation. This area is expanded in Figure S10. Access to these tracks over the entire genome of the MW, RAL and SIM samples is available through this *track data hub* containing these fine-scale statistics (this figure is obtained by expanding the view to the entire chromosome arm). Gap arise primarily for two reasons: a large repetitive region in the reference sequence(s) or gaps in the multispecies alignment.

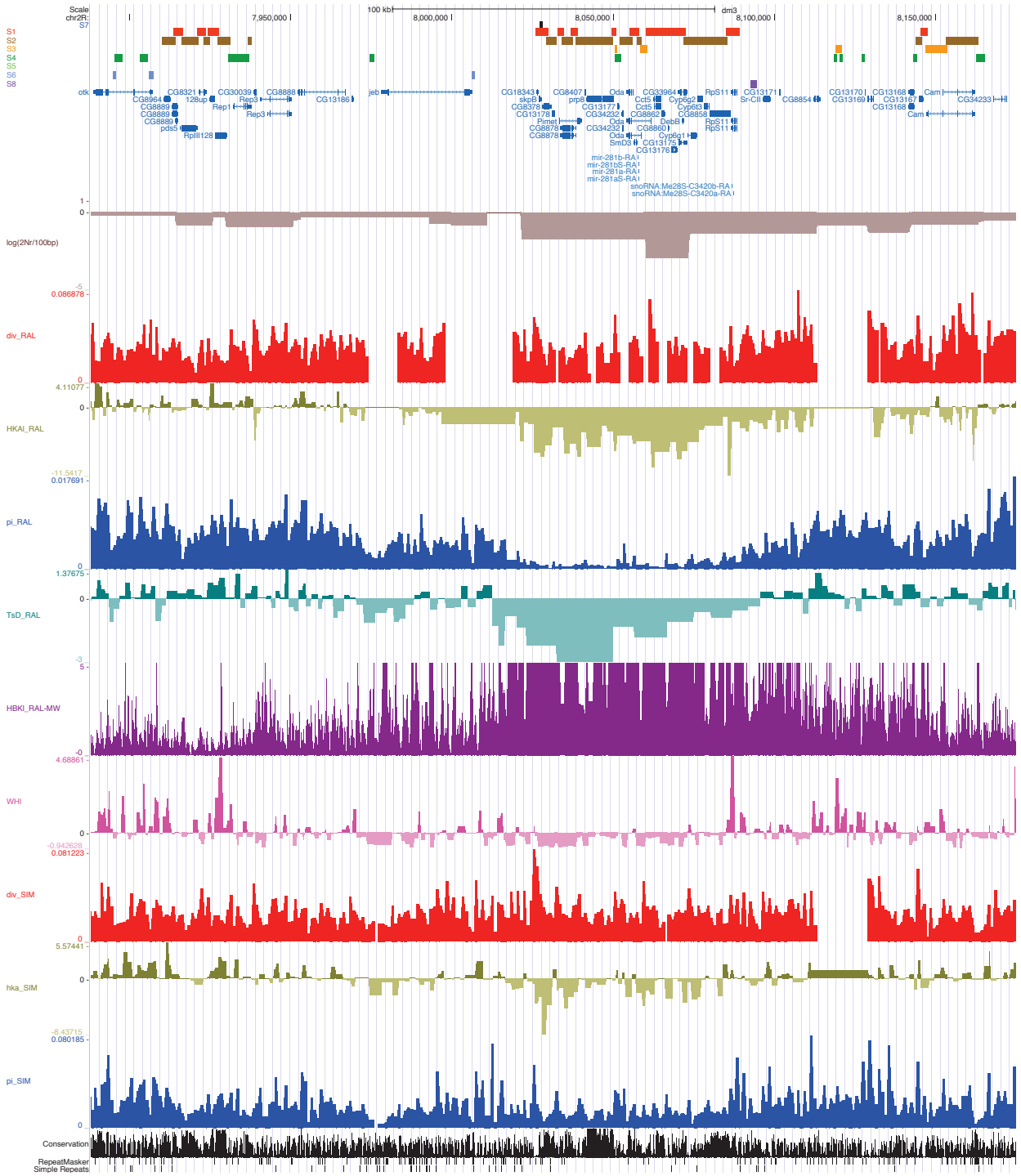


Figure S10. — Presentation in the UCSC Genome Browser of the distributions in the 260 kbp *Hen1* (*Pimet*) and *Cyp6g1* region of the population genomic statistics, π_w, δ_w, $\hat{\rho}$, *HKAI* and *TsD* for the RAL sample as well as *HBKI*, and *WHI*. At the top are *chromatin states 1* through *8*, followed by the Flybase annotated protein coding genes and then by noncoding genes. The eleventh track (down) begins the "custom" tracks from this study with $\log(\hat{\rho}/100)$, where $\hat{\rho}$ is an estimate of *2Nr*. At the bottom are three standard UCSC Genome Browser annotation tracks, phylogenetic "Conservation," "RepeatMasker" and "Simple Repeats." The browser page from which this figure was extracted can be access via this *track data hub*. Gaps arise primarily for two reasons: a large repetitive region in the reference sequence(s) or gaps in the multispecies alignment. In the region around *Cyp6g1* gaps are also attributable to the known structural polymorphisms (Schmidt *et al.* 2010). Note that the local scales of the various statistics or "tracks" are adaptive and thus variable depending on the range of the variation in the particular window.

Figure S11.— Boxplots of the distributions of π_w, δ_w, *HKAl*, *TsD*, *HBKl* and $\log(\hat{\rho})$ in genomic exonic, intronic and intergenic regions annotated as chromatin states 1 through 9.
Boxplots of the distributions of windows (weighted by bp) of π_w (RAL, MW and SIM), δ_w (RAL, MW and SIM), *HKAl* (RAL, MW and SIM), *TsD* (RAL), *HBKl* (RAL ↔ MW) and $\hat{\rho}$ (RAL) partitioned by chromatin state and gene structure (coding, intron and intergenic). The boxes are the central two quartiles, the whiskers are 1.5 time those, while the transparent light blue dots represent the outliers beyond the whiskers. The axes are labeled: "div" for δ_w ; "pi" for π_w ; "HKAl" for *HKAl* ; "D_w" for *TsD* ; "HBKl" for *HBKl* ; and "2Nr" for $\hat{\rho}$.

Figure S12.— ECDFs (epirical cumulative distribution functions) of π_w, δ_w, *HKAl*, *TsD*, *HKAl* and $\log(\hat{\rho})$ in genomic exonic, intronic and intergenic regions annotated as chromatin states 1 through 9. Empirical cumulative distribution functins of windows (weighted by bp) of π_w (RAL, MW and SIM), δ_w (RAL, MW and SIM), *HKAl* (RAL, MW and SIM), *TsD* (RAL), *HBKl* (RAL ↔ MW) and $\hat{\rho}$ (RAL) partitioned by chromatin state (see legend in each central panel) and gene structure (coding, intron and intergenic). The axes are labeled: "div" for δ_w ; "pi" for π_w ; "HKAl" for *HKAl* ; "D_w" for *TsD* ; "HBKl" for *HBKl* ; and "2Nr" for $\hat{\rho}$.

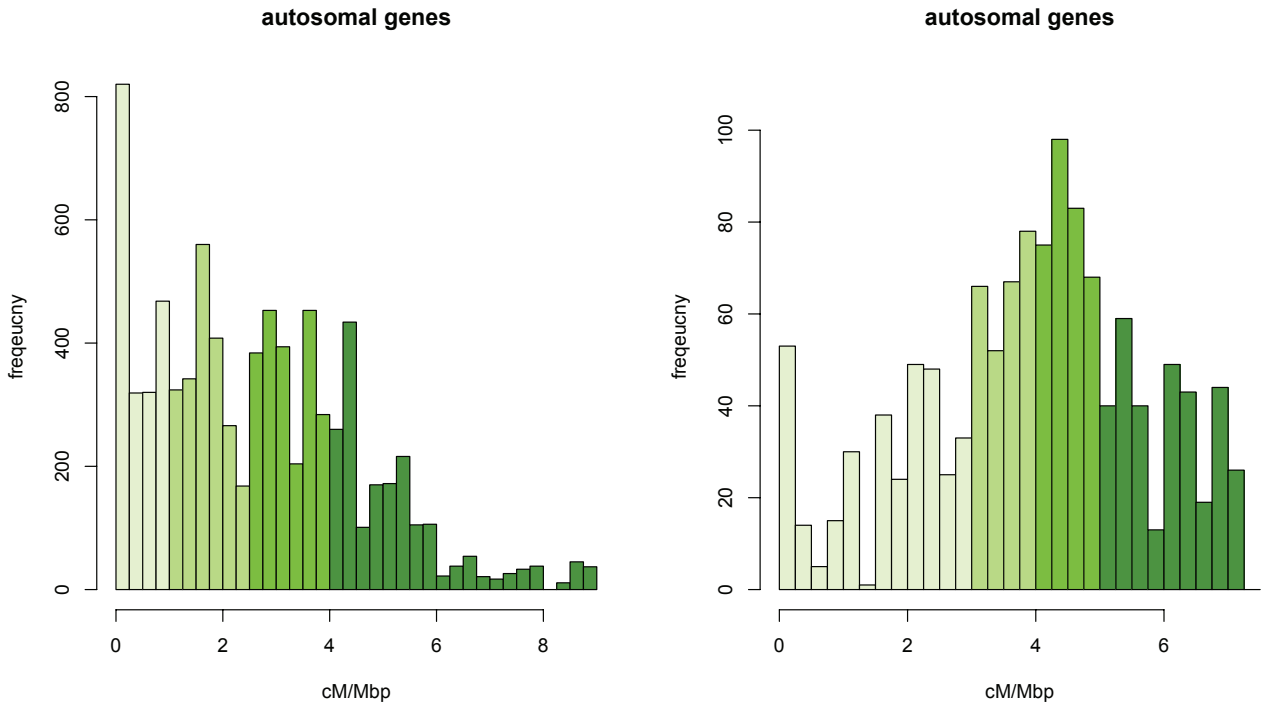


Figure S13. — The distribution of recombination rates of genes estimated by *loess* smoothed genetic maps, $\hat{r}_{15}$ in bins of 0.25 cM/Mbp. Autosomal and X-linked genes are classified into four recombination categories, which are represented with different colors (see text and methods). Darker the color, higher the recombination rate. The bins are for the purpose of showing the variations of recombination rates within each recombination categories.

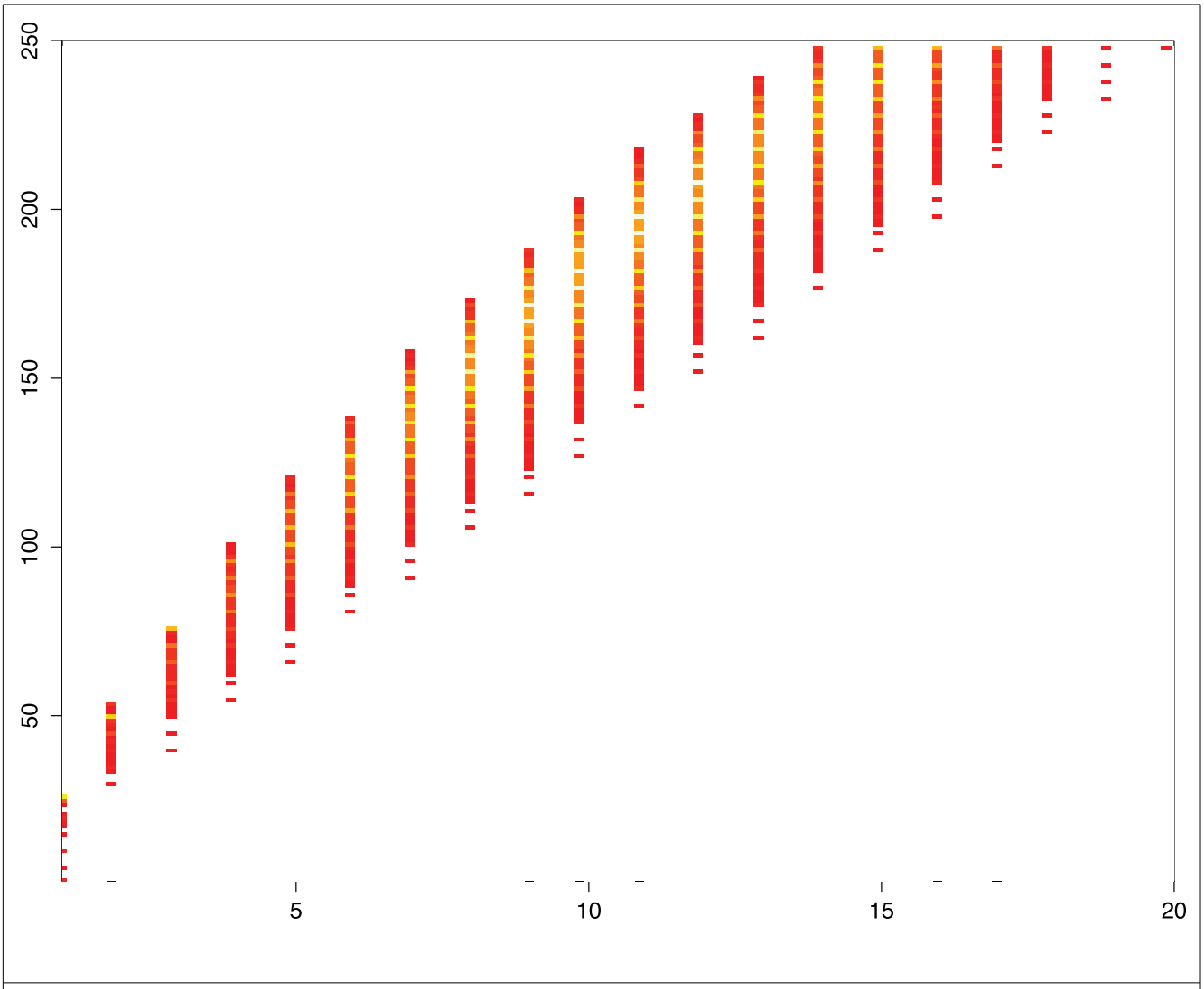


Figure SA1. — A two dimensional histogram showing the relationship between consensus quality score and read depth. The MAQ consensus quality score Q is plotted versus the MAQ consensus sequence depth (number of aligned reads contributing to the consensus). The more nucleotides that fall into a bin, the hotter (red –> yellow).

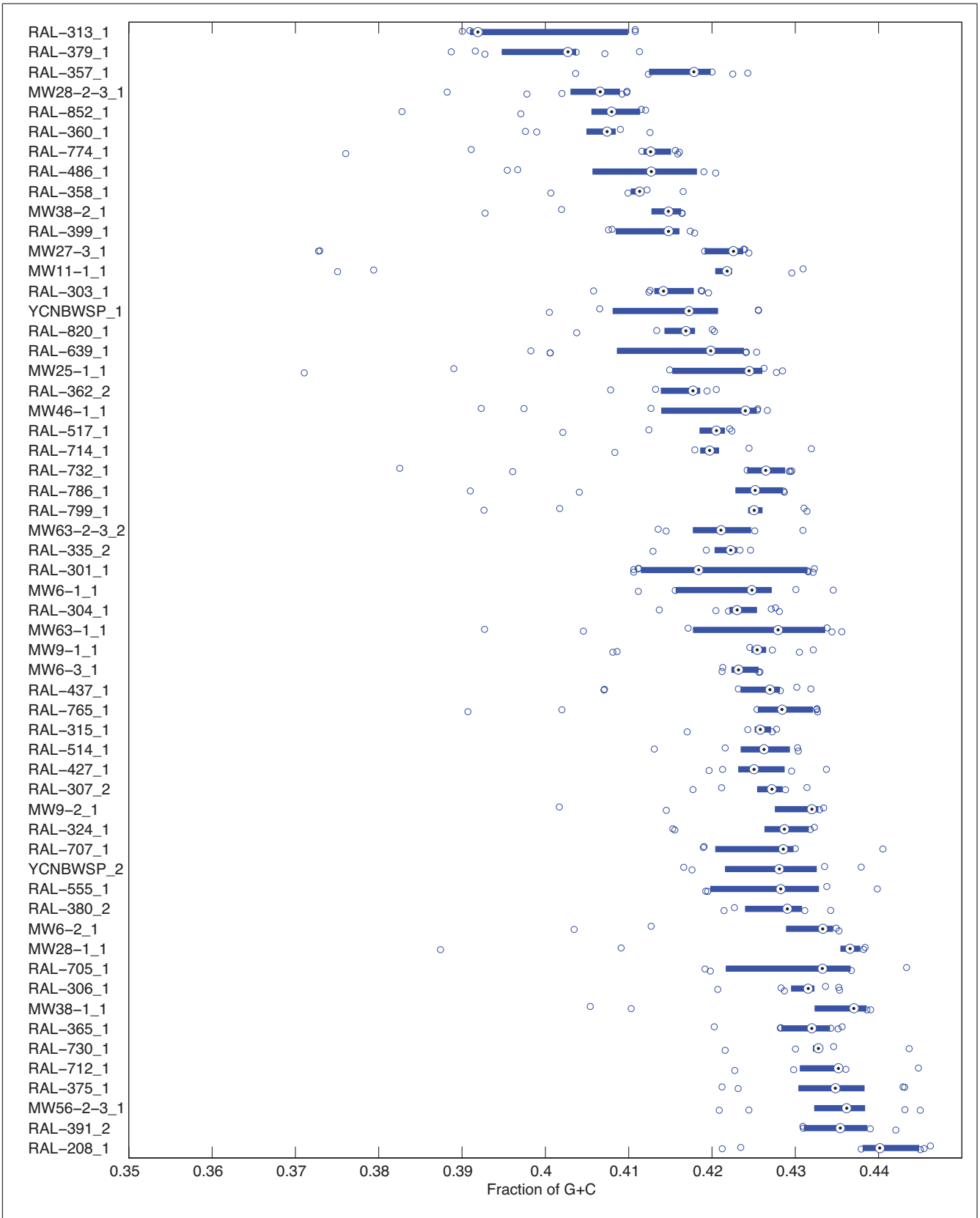


Figure SB1.— Box plots of G+C content by flow cell lane for all stocks.

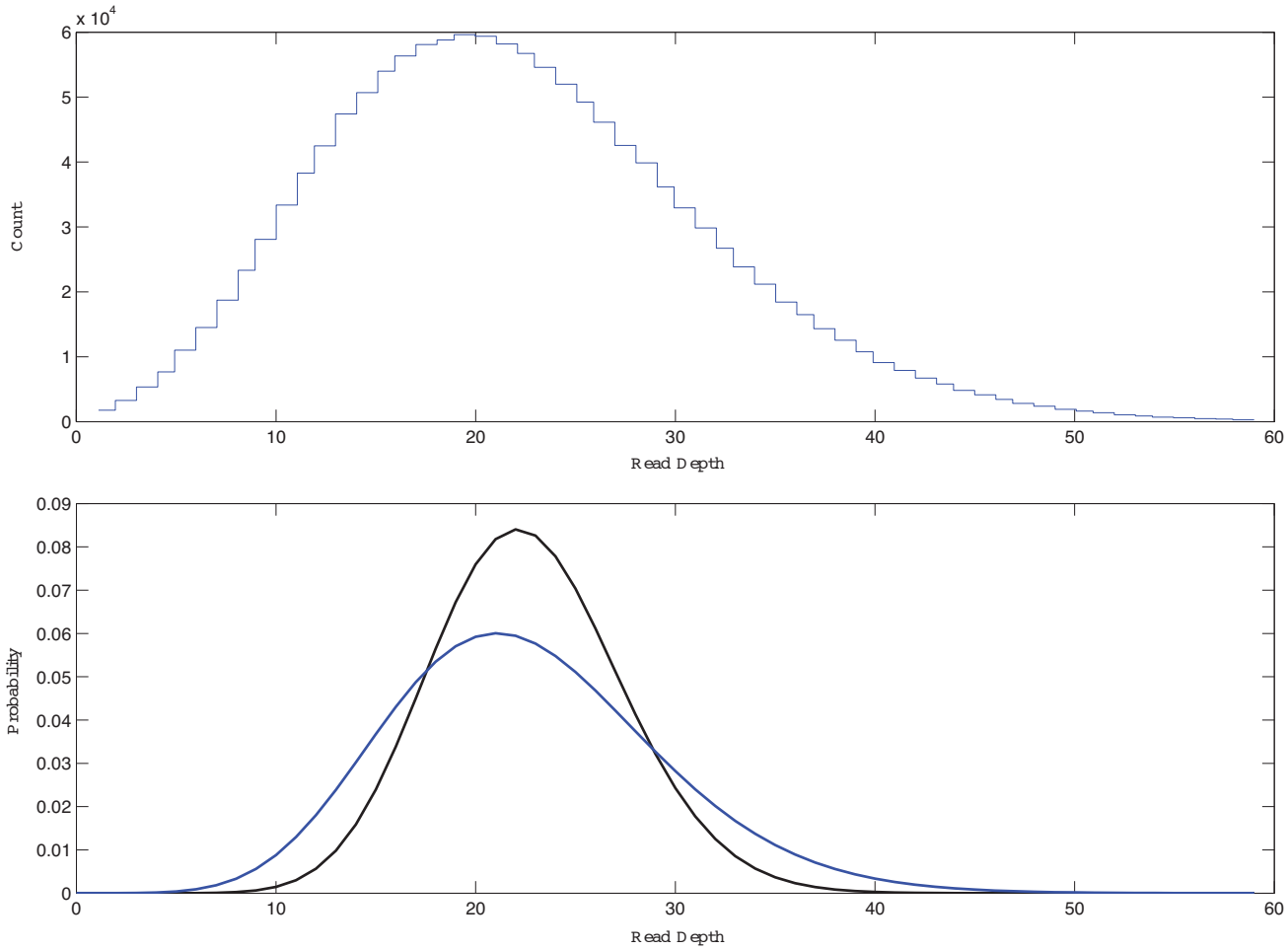


Figure SB2.— Empirical and theoretical distributions of read depth illustrating the utility of the negative binomial distribution to model read depth. The histogram for observed read counts (top) for the library `ywcnbwsp1` compared to two probability mass functions (bottom). The Poisson distribution with λ = 21.6 is shown in black. The negative binomial distribution with λ = 21.6 and η = 4.29 is shown in blue.

Table S1.
Chromosome segments identified as residually heterozygous in the indicated stock/assembly.

DPGP stock	*chromosome arm*	*begin*	*end*
RAL-301_1	chr2L	1	23011544
RAL-301_1	chr2R	1	7300000
RAL-301_1	chr3L	4800001	7600000
RAL-304_1	chr3L	1	1500000
RAL-304_1	chr3L	21200001	24543557
RAL-304_1	chr3R	1	8200000
RAL-306_1	chr3R	16000001	27905053
RAL-306_1	chr3R	7000001	9800000
RAL-307_2	chr2R	15100001	21146708
RAL-307_2	chr3R	24400001	26300000
RAL-307_2	chrX	20700001	22422827
RAL-315_1	chr2L	1	3800000
RAL-335_2	chr3L	10300001	19900000
RAL-335_2	chr3R	1200001	5500000
RAL-335_2	chr3R	8500001	10000000
RAL-357_1	chr2L	11600001	13000000
RAL-357_1	chr2L	15100001	16600000
RAL-357_1	chr2R	6600001	8500000
RAL-357_1	chr2R	16900001	21146708
RAL-358_1	chr2R	16200001	18400000
RAL-360_1	chr2L	2000001	3200000
RAL-360_1	chr3L	12800001	15900000
RAL-375_1	chr3L	15000001	19300000
RAL-375_1	chr3L	8100001	11400000
RAL-375_1	chr3R	10900001	16700000
RAL-380_2	chr2R	16700001	18600000
RAL-391_2	chr2L	5000001	6000000
RAL-391_2	chr2R	11300001	17500000
RAL-399_1	chr3R	20200001	21300000
RAL-486_1	chr2L	1	3800000
RAL-639_1	chr2L	14000001	14900000

RAL-705_1	chr3L	1	3200000
RAL-705_1	chr3R	19300001	23600000
RAL-707_1	chr3L	1	1800000
RAL-707_1	chr3R	10700001	11800000
RAL-714_1	chr2L	12500001	17500000
RAL-714_1	chr2L	7600001	9800000
RAL-732_1	chr3L	17400001	24543557
RAL-732_1	chr3R	1	27905053
RAL-774_1	chr2L	4000001	5900000
RAL-774_1	chr2L	14400001	23011544
RAL-774_1	chr2R	1	12600000
RAL-774_1	chr3L	14600001	21400000
RAL-774_1	chr3R	12600001	15500000

Table S2.
Regions filtered from the genomes *RAL-303_1, RAL-304_1 and RAL-306_1* because of apparent IBD with one another.

filtered genome	*chromosome arm*	*begin*	*end*	*IBD with genome*
RAL-303_1	chr2R	17500000	18550000	RAL-306_1
RAL-303_1	chrX	0	22422827	RAL-304_1 & RAL-306_1
RAL-304_1	chr2L	0	6400000	RAL-303_1
RAL-304_1	chr2L	0	16000000	RAL-306_1
RAL-304_1	chr3L	0	24543557	RAL-303_1 & RAL-306_1
RAL-304_1	chr3R	0	27905053	RAL-303_1 & RAL-306_1
RAL-304_1	chrX	0	22422827	RAL-303_1 & RAL-306_1
RAL-306_1	chr3L	0	end	RAL-303_1 & RAL-304_1
RAL-306_1	chr3R	0	end	RAL-303_1 & RAL-304_1

Table S3.
RAL sampling depth at Q30 plus the total numbers of assembled bp.

	X	*2L*	*2R*	*3L*	*3R*	*total*
mean	32.42	32.14	33.14	31.69	31.53	32.11
median	34	33	34	33	32	33
bp	333,706,603	478,702,933	362,718,910	440,887,238	482,662,618	2,098,678,302

Table S4.

MW sampling depth at Q30 plus the total numbers of assembled bp.

	X	*2L*	*2R*	*3L*	*3R*	*total*
mean	6.54	5.71	5.71	4.65	4.70	5.37
median	7	6	6	5	5	6
bp	61,826,328	85,405,440	61,141,436	64,766,238	70,178,732	343,318,174

Table S5.

SIM sampling depth at Q30 plus the total numbers of assembled bp.

	X	*2L*	*2R*	*3L*	*3R*	*total*
mean	3.61	4.09	4.18	4.12	4.20	4.06
median	4	5	5	5	5	5
bp	47,048,192	65,940,790	59,978,450	68,152,379	84,707,080	325,826,891

Table S6:
Potential sampling depth after filtering of residually heterozygous regions and those involved in obvious identity by descent.

	RAL		MW	SIM
	mean	max		
X	34.92	35	7	6
2L	33.90	35	6	6
2R	34.97	36	6	6
3L	33.29	35	5	6
3R	32.95	34	5	6
Total	33.95	35	5.76	6

Table S7:
RAL allelic depth at Q40 plus the total numbers of assembled bp.

	X	*2L*	*2R*	*3L*	*3R*	*total*
mean	27.73	29.38	30.21	31.69	28.86	29.02
median	30	31	32	33	30	31
bp	285,009,253	437,320,403	330,471,088	401,221,253	440,908,643	1,894,930,640

Table S8:
MW sampling depth at Q40 plus the total numbers of assembled bp.

	X	*2L*	*2R*	*3L*	*3R*	*total*
mean	5.84	5.22	5.21	4.14	4.20	4.84
median	6	6	6	5	5	5
bp	54,819,051	77,720,775	55,441,956	57,121,007	62,369,042	307,471,831

Table S9.
SIM sampling depth at Q40 plus the total numbers of assembled bp.

	X	*2L*	*2R*	*3L*	*3R*	*total*
mean	3.42	3.92	4.01	3.94	4.20	3.88
median	4	4	4	4	4	4
bp	44,488,065	62,987,720	57,347,630	65,118,060	81,020,800	310,962,275

Table S10.
Average sampling depth of coding regions on the X and on the atusomes.

	MW		RAL	
	X	autosomes	X	autosomes
Q30	6.66	5.19	31.61	32.05
Q40	5.81	4.55	24.95	28.15

Table S 11.
The correlation of *divergence* or *expected heterozygosity* in 1000 bp windows across each of the five major chromosome arms for the indicated pair of samples.

	Divergence			**Expected Heterozygosity**		
	RAL-MW	**RAL-SIM**	**MW-SIM**	**RAL-MW**	**RAL-SIM**	**MW-SIM**
X	0.98677	0.56509	0.56513	0.55434	0.29756	0.46922
2L	0.98273	0.61971	0.61706	0.73564	0.39305	0.45000
2R	0.98369	0.64671	0.64276	0.74692	0.36125	0.38505
3L	0.97959	0.63912	0.63543	0.75501	0.33818	0.32805
3R	0.98272	0.63757	0.63603	0.70080	0.32807	0.33591

Table S12.
Genetic-map-based estimates of the rate of recombination per bp, in "lettered" cytogenetic intervals of the five major chromosome arms. The assigned map position in cM is given in column two of each of the five sets (one for each of the chromosome arms).

Table S13
Distribution of missing data and statistics of partitions used in the *Ldhat*-based estimation of recombination.

Chromosome	# Blocks	# non-missing haplotypes			Average bps between consecutive SNPs
		Min	Max	Ave	
2L	20	32	35	33.6	130
2R	14	32	36	34.2	145
3L	18	31	35	33.2	146
3R	21	32	34	33.3	151
X	2	34	35	34.5	406

Table S14
The centromere proximal and telomere-proximal regions of the 5 major chromosome arms filtered or "trimmed" because of the preponderance of repetitive sequences and strong systematic effects associated with centromeres and telomeres (see text).

2L:

RAL: 0 to 844225 and 19946732 to the end.
MW: 0 to 698949 and 19954780 to the end.
SIM: 0 to 650976 and 20052632 to the end.

2R:

RAL: 0 to 6063980 and 20322335 to the end.
MW: 0 to 6090470 and 20020890 to the end.
SIM: 0 to 2935239 and 20321706 to the end.

3L:

RAL: 0 to 447386 and 18392988 to the end.
MW: 0 to 356604 and 18408033 to the end.
(14656580 to the end, except for two 4096 SNP windows in which $\chi[\log(p_{HKAl})] > 0$)
SIM: 0 to 897221 and 21109190 to the end.

3R:

RAL: 0 to 7940899 and 27237549 to the end.
MW: 0 to 8349278 and 27248244 to the end.
SIM: 0 to 2765860 and 26741546 to the end.

X:

RAL: 0 to 1036552 and 20902578 to the end.
MW: 0 to 2460008 and 20665672 to the end.

SIM: 0 to 2200059 and 19271518 to the end.

2L:

RAL: 0 to 844225 and 19946732 to the end.
MW: 0 to 698949 and 19954780 to the end.
SIM: 0 to 650976 and 20052632 to the end.

2R:

RAL: 0 to 6063980 and 20322335 to the end.
MW: 0 to 6090470 and 20020890 to the end.
SIM: 0 to 2935239 and 20321706 to the end.

3L:

RAL: 0 to 447386 and 18392988 to the end.
MW: 0 to 356604 and 18408033 to the end.
(14656580 to the end, except for two 4096 SNP windows in which $\chi[\log(p_{HKAl})] > 0$)
SIM: 0 to 897221 and 21109190 to the end.

3R:

RAL: 0 to 7940899 and 27237549 to the end.
MW: 0 to 8349278 and 27248244 to the end.
SIM: 0 to 2765860 and 26741546 to the end.

X:

RAL: 0 to 1036552 and 20902578 to the end.
MW: 0 to 2460008 and 20665672 to the end.
SIM: 0 to 2200059 and 19271518 to the end.

Table S15

The correlations between the logarithm of the estimated rates of recombination ($\log(\hat{r}_{15})$, $\log(\hat{\rho}_{15})$, $\log(\hat{\rho})$) with π_w, *HKAl*, *TsD*, or *HBKl* for each chromosome arm in the RAL, MW and SIM samples (see text).

		statistic		chromosome arms						
sample	**U | T**	**1**	**2**	**chr2L**	**chr2R**	**chr3L**	**chr3R**	**autosomes**	**chrX**	**all**
RAL-MW	U	$\log(\hat{\rho})$	*HBKl*	-0.0162	-0.0639	-0.0678	-0.0289	-0.0489	-0.0060	-0.0026
RAL-MW	T	$\log(\hat{\rho})$	*HBKl*	-0.0254	-0.1119	-0.0568	-0.0727	-0.0759	-0.0098	-0.0054
MW	U	$\log(\hat{\rho})$	*HKAl*	0.5624	0.6320	0.6875	0.5230	0.6049	0.3315	0.5603
MW	T	$\log(\hat{\rho})$	*HKAl*	0.2143	0.2549	0.2135	0.1602	0.2006	0.0483	0.1655
RAL	U	$\log(\hat{\rho})$	*HKAl*	0.5838	0.6781	0.7198	0.5827	0.6436	0.2181	0.5786
RAL	T	$\log(\hat{\rho})$	*HKAl*	0.2692	0.3363	0.2145	0.2548	0.2584	0.0486	0.2131
SIM	U	$\log(\hat{\rho})$	*HKAl*	0.6153	0.2782	0.3820	0.4413	0.4264	0.3910	0.4173
SIM	T	$\log(\hat{\rho})$	*HKAl*	0.1304	0.1539	0.1671	0.1382	0.1406	0.0892	0.1319
MW	U	$\log(\hat{\rho})$	π_w	0.4451	0.4568	0.5257	0.4407	0.4838	0.3217	0.4542
MW	T	$\log(\hat{\rho})$	π_w	0.2858	0.2417	0.2737	0.1861	0.2632	0.1425	0.2337
RAL	U	$\log(\hat{\rho})$	π_w	0.4541	0.4964	0.5342	0.4870	0.5064	0.2187	0.4169
RAL	T	$\log(\hat{\rho})$	π_w	0.3175	0.3199	0.2939	0.2819	0.3181	0.1345	0.2186
SIM	U	$\log(\hat{\rho})$	π_w	0.2351	0.1493	0.1803	0.1460	0.1834	0.2422	0.1762
SIM	T	$\log(\hat{\rho})$	π_w	0.1056	0.0845	0.0757	0.0888	0.0994	0.0884	0.0733
MW	U	$\log(\hat{\rho})$	*TsD*	0.0473	0.1275	0.0850	0.0739	0.0829	0.0678	0.0742
MW	T	$\log(\hat{\rho})$	*TsD*	0.0340	0.0892	0.0446	0.0641	0.0481	0.0168	0.0325
RAL	U	$\log(\hat{\rho})$	*TsD*	0.1328	0.2471	0.3724	0.2410	0.2598	0.0480	0.1898
RAL	T	$\log(\hat{\rho})$	*TsD*	0.0185	0.0848	0.0724	0.1272	0.0802	0.0959	0.0625
RAL-MW	U	$\log(\hat{\rho}_{15})$	*HBKl*	-0.0082	0.0543	0.0050	0.0467	0.0167	0.0226	0.1099
RAL-MW	T	$\log(\hat{\rho}_{15})$	*HBKl*	0.0517	0.0087	0.0142	0.0286	-0.0047	0.0179	0.1746
MW	U	$\log(\hat{\rho}_{15})$	*HKAl*	0.5648	0.5596	0.7337	0.6196	0.6052	0.3547	0.5775
MW	T	$\log(\hat{\rho}_{15})$	*HKAl*	0.2184	0.1338	0.3154	0.0813	0.1521	0.0341	0.1237
RAL	U	$\log(\hat{\rho}_{15})$	*HKAl*	0.5328	0.5196	0.7458	0.6612	0.6060	0.1945	0.5687
RAL	T	$\log(\hat{\rho}_{15})$	*HKAl*	0.2012	0.1476	0.1530	0.1618	0.1400	-0.0007	0.1204
SIM	U	$\log(\hat{\rho}_{15})$	*HKAl*	0.5775	0.3819	0.3073	0.4900	0.4032	0.4508	0.4027
SIM	T	$\log(\hat{\rho}_{15})$	*HKAl*	0.0100	0.0033	0.0564	0.0853	0.0307	-0.0136	0.0309
MW	U	$\log(\hat{\rho}_{15})$	π_w	0.4302	0.3654	0.5581	0.4783	0.4812	0.2948	0.4562
MW	T	$\log(\hat{\rho}_{15})$	π_w	0.2308	0.0992	0.3200	0.0984	0.2166	0.0381	0.1752
RAL	U	$\log(\hat{\rho}_{15})$	π_w	0.3974	0.3488	0.5236	0.4746	0.4574	0.1525	0.3343
RAL	T	$\log(\hat{\rho}_{15})$	π_w	0.2023	0.1226	0.2282	0.1618	0.2125	0.0140	0.0122
SIM	U	$\log(\hat{\rho}_{15})$	π_w	0.1467	0.1081	0.1033	0.1149	0.1261	0.2568	0.1009
SIM	T	$\log(\hat{\rho}_{15})$	π_w	0.0034	-0.0549	-0.0508	0.0432	0.0141	0.0247	-0.0428
MW	U	$\log(\hat{\rho}_{15})$	*TsD*	0.1040	0.1179	-0.0436	0.0543	0.0482	0.0578	0.0346
MW	T	$\log(\hat{\rho}_{15})$	*TsD*	0.0502	0.0514	0.0021	0.0636	0.0227	0.0219	-0.0117
RAL	U	$\log(\hat{\rho}_{15})$	*TsD*	0.1324	0.1396	0.3084	0.1989	0.1997	-0.0021	0.1185
RAL	T	$\log(\hat{\rho}_{15})$	*TsD*	0.0382	0.0324	0.1177	0.0282	0.0182	0.0024	-0.0595
RAL-MW	U	$\log(\hat{r}_{15})$	*HBKl*	-0.0228	0.0399	-0.0002	0.0676	0.0206	0.1435	0.0975
RAL-MW	T	$\log(\hat{r}_{15})$	*HBKl*	0.0420	0.0049	0.0142	0.0631	0.0321	0.0121	0.1173
MW	U	$\log(\hat{r}_{15})$	*HKAl*	0.4989	0.6084	0.6005	0.5607	0.5677	0.4026	0.5556
MW	T	$\log(\hat{r}_{15})$	*HKAl*	0.1661	0.0548	0.2912	0.0695	0.1789	-0.0049	0.1549
RAL	U	$\log(\hat{r}_{15})$	*HKAl*	0.4605	0.5863	0.5460	0.5392	0.5315	0.2333	0.5132
RAL	T	$\log(\hat{r}_{15})$	*HKAl*	0.1592	0.0592	0.1492	0.1084	0.1192	-0.0485	0.1052
SIM	U	$\log(\hat{r}_{15})$	*HKAl*	0.5756	0.2264	0.2624	0.3203	0.3603	0.5216	0.3747
SIM	T	$\log(\hat{r}_{15})$	*HKAl*	0.0047	0.0030	0.0635	0.0737	0.0450	0.0695	0.0472
MW	U	$\log(\hat{r}_{15})$	π_w	0.3486	0.3402	0.4515	0.4531	0.3898	0.2688	0.3791
MW	T	$\log(\hat{r}_{15})$	π_w	0.1584	0.0098	0.2183	0.0501	0.1194	0.0148	0.1072
RAL	U	$\log(\hat{r}_{15})$	π_w	0.3295	0.3430	0.3895	0.4210	0.3610	0.1175	0.2795
RAL	T	$\log(\hat{r}_{15})$	π_w	0.1427	0.0093	0.1370	0.0643	0.0927	-0.0214	-0.0037
SIM	U	$\log(\hat{r}_{15})$	π_w	-0.0988	-0.0953	-0.0654	0.0063	-0.0432	-0.0046	-0.0485
SIM	T	$\log(\hat{r}_{15})$	π_w	-0.0077	-0.0536	-0.1165	0.0243	-0.0438	0.0248	-0.0711
MW	U	$\log(\hat{r}_{15})$	*TsD*	0.0224	0.0794	-0.0283	0.0465	0.0282	0.0476	0.0210
MW	T	$\log(\hat{r}_{15})$	*TsD*	0.0518	0.0392	0.0064	0.0456	0.0419	0.0090	0.0213
RAL	U	$\log(\hat{r}_{15})$	*TsD*	0.1276	0.1586	0.2417	0.1996	0.1844	-0.0717	0.1083
RAL	T	$\log(\hat{r}_{15})$	*TsD*	-0.0411	-0.0110	0.0924	0.0074	0.0252	-0.0471	-0.0266

Table S16.
Amino acid replacement F_{ST} GO enrichment analysis – *biological* categories.

GO category	proportion significant genes	p-values	GOslim description	GO description
GO:0042060	0.833	0.0001	NA	The series of events that restore integrity to a damaged tissue, following an injury.
GO:0006807	0.667	0.0002	nitrogen compound metabolic process	The chemical reactions and pathways involving various organic and inorganic nitrogenous compounds; includes nitrogen fixation, nitrification, denitrification, assimilatory/dissimilatory nitrate reduction and the interconversion of nitrogenous organic matter and ammonium.
GO:0006887	0.667	0.0002	exocytosis	A process of secretion by a cell that results in the release of intracellular molecules (e.g. hormones, matrix proteins) contained within a membrane-bounded vesicle by fusion of the vesicle with the plasma membrane of a cell. This is the process whereby most molecules are secreted from eukaryotic cells.
GO:0008152	0.239	0.0003	metabolic process	The chemical reactions and pathways, including anabolism and catabolism, by which living organisms transform chemical substances. Metabolic processes typically transform small molecules, but also include macromolecular processes such as DNA repair and replication, and protein synthesis and degradation.
GO:0050909	0.364	0.0003	NA	The series of events required for an organism to receive a gustatory stimulus, convert it to a molecular signal, and recognize and characterize the signal. Gustation involves the direct detection of chemical composition, usually through contact with chemoreceptor cells. This is a neurological process.
GO:0046529	0.556	0.0005	NA	The joining of the parts of the wing imaginal discs, giving rise to the adult thorax.

GO:0007561	0.571	0.0006	imaginal disc eversion	The eversion (turning inside out) of imaginal discs from their peripodial sacs, resulting in movement of the epithelium to the outside of the larval epidermis.
GO:0007296	0.600	0.0017	vitellogenesis	The production of yolk. Yolk is a mixture of materials used for embryonic nutrition.
GO:0006310	0.500	0.0025	DNA recombination	Any process by which a new genotype is formed by reassortment of genes resulting in gene combinations different from those that were present in the parents. In eukaryotes genetic recombination can occur by chromosome assortment, intrachromosomal recombination, or nonreciprocal interchromosomal recombination. Intrachromosomal recombination occurs by crossing over. In bacteria it may occur by genetic transformation, conjugation, transduction, or F-duction.
GO:0007362	0.417	0.004	terminal region determination	Specification of the terminal regions (the two non-segmented ends) of the embryo by the gap genes; exemplified in insects by the actions of huckebein and tailless gene products.
GO:0009620	0.500	0.0042	NA	A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus from a fungus.
GO:0008298	0.375	0.0046	intracellular mRNA localization	Any process by which mRNA is transported to, or maintained in, a specific location within the cell.
GO:0046907	0.500	0.0048	NA	The directed movement of substances within a cell.
GO:0007030	0.385	0.0057	Golgi organization	A process that is carried out at the cellular level which results in the assembly, arrangement of constituent parts, or disassembly of the Golgi apparatus.
GO:0007426	0.385	0.0067	tracheal outgrowth, open tracheal system	The projection of branches of an open tracheal system towards their target tissues. An example of this is found in *Drosophila melanogaster*.
GO:0045610	0.429	0.0088	NA	Any process that modulates the frequency, rate or extent of hemocyte differentiation.

GO:0007157	0.429	0.0111	heterophilic cell adhesion	The attachment of an adhesion molecule in one cell to a nonidentical adhesion molecule in an adjacent cell.
GO:0008285	0.364	0.0126	negative regulation of cell proliferation	Any process that stops, prevents or reduces the rate or extent of cell proliferation.
GO:0007131	0.364	0.0138	reciprocal meiotic recombination	The cell cycle process whereby double strand breaks are formed and repaired through a double Holliday junction intermediate. This results in the equal exchange of genetic material between non-sister chromatids in a pair of homologous chromosomes. These reciprocal recombinant products ensure the proper segregation of homologous chromosomes during meiosis I and create genetic diversity.
GO:0007265	0.364	0.0144	Ras protein signal transduction	A series of molecular signals within the cell that are mediated by a member of the Ras superfamily of proteins switching to a GTP-bound active state.
GO:0046843	0.316	0.0148	NA	Establishment of the dorsal filaments, elaborate specializations of the chorion that protrude from the anterior end of the egg and facilitate embryonic respiration.
GO:0000165	0.375	0.0172	MAPKKK cascade	A cascade of at least three protein kinase activities culminating in the phosphorylation and activation of a MAP kinase. MAPKKK cascades lie downstream of numerous signaling pathways.
GO:0006259	0.211	0.019	DNA metabolic process	The chemical reactions and pathways involving DNA, deoxyribonucleic acid, one of the two main types of nucleic acid, consisting of a long, unbranched macromolecule formed from one, or more commonly, two, strands of linked deoxyribonucleotides.
GO:0006726	0.375	0.0203	eye pigment biosynthetic process	The chemical reactions and pathways resulting in the formation of eye pigments, any general or particular coloring matter in living organisms, found or utilized in the eye.
GO:0006727	0.333	0.021	ommochrome biosynthetic process	The chemical reactions and pathways resulting in the formation of ommochromes, any of a large group of natural polycyclic pigments commonly found in the Arthropoda, particularly in the ommatidia of the compound eye.

GO:0033227	0.313	0.021	NA	The directed movement of dsRNA, double-stranded ribonucleic acid, into, out of, within or between cells by means of some external agent such as a transporter or pore.
GO:0007298	0.256	0.0214	border follicle cell migration	The directed movement of the border cells through the nurse cells to reach the oocyte. An example of this is found in *Drosophila melanogaster*.
GO:0009617	0.375	0.0223	NA	A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus from a bacterium.
GO:0006856	0.400	0.0229	eye pigment precursor transport	The directed movement of eye pigment precursors, the inactive forms of visual pigments, into, out of, within or between cells by means of some external agent such as a transporter or pore.
GO:0007605	0.400	0.0229	sensory perception of sound	The series of events required for an organism to receive an auditory stimulus, convert it to a molecular signal, and recognize and characterize the signal. Sonic stimuli are detected in the form of vibrations and are processed to form a sound.
GO:0006869	0.400	0.0234	lipid transport	The directed movement of lipids into, out of, within or between cells by means of some external agent such as a transporter or pore. Lipids are compounds soluble in an organic solvent but not, or sparingly, in an aqueous solvent.
GO:0006378	0.333	0.0242	mRNA polyadenylation	The enzymatic addition of a sequence of 40-200 adenylyl residues at the 3' end of a eukaryotic mRNA primary transcript.
GO:0006352	0.400	0.0244	transcription initiation	Any process involved in the assembly of the RNA polymerase complex at the promoter region of a DNA template, resulting in the subsequent synthesis of RNA from that promoter.
GO:0006904	0.400	0.0245	vesicle docking during exocytosis	The initial attachment of a vesicle membrane to a target membrane, mediated by proteins protruding from the membrane of the vesicle and the target membrane, during exocytosis.

GO:0007307	0.333	0.0245	eggshell chorion gene amplification	Amplification by up to 60-fold of the loci containing the chorion gene clusters. Amplification is necessary for the rapid synthesis of chorion proteins by the follicle cells, and occurs by repeated firing of one or more origins located within each gene cluster.
GO:0006967	0.400	0.0247	positive regulation of antifungal peptide biosynthetic process	Any process that activates or increases the frequency, rate, or extent of antifungal peptide biosynthesis.
GO:0006777	0.400	0.0248	Mo-molybdopterin cofactor biosynthetic process	The chemical reactions and pathways resulting in the formation of the Mo-molybdopterin cofactor, essential for the catalytic activity of some enzymes. The cofactor consists of a mononuclear molybdenum (Mo) ion coordinated by one or two molybdopterin ligands.
GO:0042386	0.400	0.0255	NA	The process whereby a relatively unspecialized cell acquires the characteristics of a mature hemocyte. Hemocytes are blood cells associated with a hemocoel (the cavity containing most of the major organs of the arthropod body) which are involved in defense and clotting of hemolymph, but not involved in transport of oxygen.
GO:0016202	0.400	0.0256	NA	Any process that modulates the frequency, rate or extent of striated muscle development.
GO:0019732	0.400	0.0256	NA	An immune response against a fungus mediated through a body fluid. An example of this process is the antifungal humoral response in *Drosophila melanogaster*.
GO:0008333	0.400	0.0257	endosome to lysosome transport	The directed movement of substances from endosomes to lysosomes.
GO:0042810	0.400	0.0259	NA	The chemical reactions and pathways involving pheromones, a substance that is secreted and released by an organism and detected by a second organism of the same or a closely related species, in which it causes a specific reaction, such as a definite behavioral reaction or a developmental process.

GO:0007613	0.400	0.026	memory	The activities involved in the mental information processing system that receives (registers), modifies, stores, and retrieves informational stimuli. The main stages involved in the formation and retrieval of memory are encoding (processing of received information by acquisition), storage (building a permanent record of received information as a result of consolidation) and retrieval (calling back the stored information and use it in a suitable way to execute a given task).
GO:0030111	0.400	0.0268	NA	Any process that modulates the frequency, rate or extent of the activity of the Wnt receptor mediated signal transduction pathway.
GO:0050830	0.308	0.0274	NA	Reactions triggered in response to the presence of a Gram-positive bacterium that act to protect the cell or organism.
GO:0008293	0.294	0.0285	torso signaling pathway	The series of molecular signals generated as a consequence of the torso transmembrane receptor tyrosine kinase binding to its physiological ligand.
GO:0016477	0.308	0.0286	NA	The orderly movement of cells from one site to another, often during the development of a multicellular organism or multicellular structure.
GO:0008586	0.273	0.0294	imaginal disc-derived wing vein morphogenesis	The process by which anatomical structures of the veins on an imaginal disc-derived wing are generated and organized. Morphogenesis pertains to the creation of form.
GO:0048675	0.333	0.0294	NA	Long distance growth of a single process.
GO:0035017	0.333	0.0299	NA	The regionalization process that gives rise to the patterns of cell differentiation in the cuticle.
GO:0050832	0.308	0.0303	NA	Reactions triggered in response to the presence of a fungus that act to protect the cell or organism.
GO:0006417	0.333	0.0319	regulation of translation	Any process that modulates the frequency, rate or extent of the chemical reactions and pathways resulting in the formation of proteins by the translation of mRNA.
GO:0042048	0.273	0.0323	NA	The actions or reactions of an organism in response to an odor.

GO:0051017	0.333	0.0327	NA	The assembly of actin filament bundles; actin filaments are on the same axis but may be oriented with the same or opposite polarities and may be packed with different levels of tightness.
GO:0030154	0.333	0.0329	cell differentiation	The process whereby relatively unspecialized cells, e.g. embryonic or regenerative cells, acquire specialized structural and/or functional features that characterize the cells, tissues, or organs of the mature organism or some other relatively stable phase of the organism's life history. Differentiation includes the processes involved in commitment of a cell to a specific fate.
GO:0046667	0.333	0.0391	NA	Programmed cell death that occurs in the retina to remove excess cells between ommatidia, thus resulting in a hexagonal lattice, precise with respect to cell number and position surrounding each ommatidium.
GO:0045087	0.261	0.0402	NA	Innate immune responses are defense responses mediated by germline encoded components that directly recognize components of potential pathogens.
GO:0009408	0.237	0.0414	NA	A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a heat stimulus, a temperature stimulus above the optimal temperature for that organism.
GO:0007076	0.286	0.0417	mitotic chromosome condensation	The cell cycle process whereby chromatin structure is compacted prior to and during mitosis in eukaryotic cells.
GO:0007628	0.333	0.0419	adult walking behavior	The actions or reactions of an adult relating to the progression of that organism along the ground by the process of lifting and setting down each leg.
GO:0016081	0.286	0.0424	NA	The initial attachment of a synaptic vesicle membrane to the presynaptic membrane, mediated by proteins protruding from the membrane of the synaptic vesicle and the target membrane.

GO:0035172	0.333	0.0427	NA	The multiplication or reproduction of hemocytes, resulting in the expansion of the cell population. Hemocytes are blood cells associated with a hemocoel (the cavity containing most of the major organs of the arthropod body) which are involved in defense and clotting of hemolymph, but not involved in transport of oxygen.
GO:0006281	0.250	0.043	DNA repair	The process of restoring DNA after damage. Genomes are subject to damage by chemical and physical agents in the environment (e.g. UV and ionizing radiations, chemical mutagens, fungal and bacterial toxins, etc.) and by free radicals or alkylating agents endogenously generated in metabolism. DNA is also damaged because of errors during its replication. A variety of different DNA repair pathways have been reported that include direct reversal, base excision repair, nucleotide excision repair, photoreactivation, bypass, double-strand break repair pathway, and mismatch repair pathway.
GO:0000070	0.286	0.0431	mitotic sister chromatid segregation	The cell cycle process whereby replicated homologous chromosomes are organized and then physically separated and apportioned to two sets during the mitotic cell cycle. Each replicated chromosome, composed of two sister chromatids, aligns at the cell equator, paired with its homologous partner. One homolog of each morphologic type goes into each of the resulting chromosome sets.
GO:0007020	0.333	0.0434	microtubule nucleation	The 'de novo' formation of a microtubule, in which tubulin heterodimers form metastable oligomeric aggregates, some of which go on to support formation of a complete microtubule. Microtubule nucleation usually occurs from a specific site within a cell.
GO:0016334	0.333	0.0434	NA	Any cellular process that results in the specification, formation or maintenance of a polarized follicular epithelial sheet.

GO:0035160	0.333	0.0435	NA	Ensuring that tracheal tubes in an open tracheal system maintain their epithelial structure during the cell shape changes and movements that occur during the branching process.
GO:0019731	0.286	0.0438	NA	An immune response against bacteria mediated through a body fluid. Examples of this process are the antibacterial humoral responses in *Mus musculus* and *Drosophila melanogaster*.
GO:0006302	0.333	0.0443	double-strand break repair	The repair of double-strand breaks in DNA via homologous and nonhomologous mechanisms to reform a continuous DNA helix.
GO:0007617	0.333	0.0449	mating behavior	The behavioral interactions between organisms for the purpose of mating, or sexual reproduction resulting in the formation of zygotes.
GO:0006825	0.333	0.045	copper ion transport	The directed movement of copper (Cu) ions into, out of, within or between cells by means of some external agent such as a transporter or pore.
GO:0016339	0.300	0.0456	NA	The attachment of one cell to another cell via adhesion molecules that require the presence of calcium for the interaction.
GO:0007160	0.333	0.046	cell-matrix adhesion	The binding of a cell to the extracellular matrix via adhesion molecules.
GO:0015914	0.333	0.0466	NA	The directed movement of phospholipids into, out of, within or between cells by means of some external agent such as a transporter or pore. Phospholipids are any lipids containing phosphoric acid as a mono- or diester.
GO:0035277	0.300	0.0467	NA	The process by which the anatomical structures of a spiracle are generated and organized. Spiracles are the openings in the insect open tracheal system; externally they connect to the epidermis and internally they connect to the tracheal trunk.

GO:0008535	0.333	0.0468	respiratory chain complex IV assembly	The aggregation, arrangement and bonding together of a set of components to form respiratory chain complex IV (also known as cytochrome c oxidase), the terminal member of the respiratory chain of the mitochondrion and some aerobic bacteria. Cytochrome c oxidases are multi-subunit enzymes containing from 13 subunits in the mammalian mitochondrial form to 3-4 subunits in the bacterial forms.
GO:0035220	0.300	0.0478	NA	Progression of the wing disc over time, from its initial formation through to its metamorphosis to form adult structures including the wing hinge, wing blade and pleura.
GO:0007528	0.300	0.0479	neuromuscular junction development	The process whose specific outcome is the progression of the neuromuscular junction over time, from its formation to the mature structure.
GO:0007494	0.263	0.0497	midgut development	The process whose specific outcome is the progression of the midgut over time, from its formation to the mature structure. The midgut is the middle part of the alimentary canal from the stomach, or entrance of the bile duct, to, or including, the large intestine.

Table S17

Amino acid replacement F_{ST} GO enrichment analysis – *cellular* categories.

GO	proportion significant genes	p-values	GOslim description	GO description
GO:0000796	0.571	0.001	condensin complex	A multisubunit protein complex that plays a central role in chromosome condensation.
GO:0008076	0.400	0.0088	voltage-gated potassium channel complex	A protein complex that forms a transmembrane channel through which potassium ions may cross a cell membrane in response to changes in membrane potential.
GO:0016222	0.400	0.0258	NA	A protein complex that catalyzes the formation of procollagen trans-4-hydroxy-L-proline and succinate from procollagen L-proline and 2-oxoglutarate, requiring Fe2+ and ascorbate. Contains two alpha subunits that contribute to most parts of the catalytic sites, and two beta subunits that are identical to protein-disulfide isomerase.
GO:0005682	0.400	0.0266	U5 snRNP	A ribonucleoprotein complex that contains small nuclear RNA U5.
GO:0016327	0.400	0.0266	NA	The apical end of the lateral plasma membrane of epithelial cells.
GO:0005694	0.308	0.0305	chromosome	A structure composed of a very long molecule of DNA and associated proteins (e.g. histones) that carries hereditary information.
GO:0005654	0.308	0.0332	nucleoplasm	That part of the nuclear content other than the chromosomes or the nucleolus.
GO:0005643	0.261	0.039	nuclear pore	Any of the numerous similar discrete openings in the nuclear envelope of a eukaryotic cell, where the inner and outer nuclear membranes are joined.
GO:0019898	0.300	0.0451	NA	Loosely bound to one surface of a membrane, but not integrated into the hydrophobic region.

Table S18

Amino acid replacement F_{ST} GO enrichment analysis – *molecular* categories.

GO	proportion significant genes	p-values	GOslim description	GO description
GO:0004656	0.556	0.0005	procollagen-proline 4-dioxygenase activity	Catalysis of the reaction: procollagen L-proline + 2-oxoglutarate + O2 = procollagen trans-4-hydroxy-L-proline + succinate + CO2.
GO:0031418	0.455	0.002	NA	Interacting selectively and non-covalently with L-ascorbic acid, (2R)-2-[(1S)-1,2-dihydroxyethyl]-4-hydroxy-5-oxo-2,5-dihydrofuran-3-olate; L-ascorbic acid is vitamin C and has co-factor and anti-oxidant activities in many species.
GO:0005506	0.290	0.010	iron ion binding	Interacting selectively and non-covalently with iron (Fe) ions.
GO:0008527	0.333	0.010	taste receptor activity	Combining with soluble compounds to initiate a change in cell activity. These receptors are responsible for the sense of taste.
GO:0008026	0.400	0.010	ATP-dependent helicase activity	Catalysis of the reaction: ATP + H2O = ADP + phosphate to drive the unwinding of a DNA or RNA helix.
GO:0005544	0.429	0.011	calcium-dependent phospholipid binding	Interacting selectively and non-covalently with phospholipids, a class of lipids containing phosphoric acid as a mono- or diester, in the presence of calcium.
GO:0050839	0.429	0.011	NA	Interacting selectively and non-covalently with a cell adhesion molecule.
GO:0016702	0.429	0.012	NA	Catalysis of an oxidation-reduction (redox) reaction in which hydrogen or electrons are transferred from one donor, and two oxygen atoms is incorporated into a donor.
GO:0003724	0.375	0.019	RNA helicase activity	Catalysis of the reaction: NTP + H2O = NDP + phosphate to drive the unwinding of a RNA helix.
GO:0003824	0.224	0.019	catalytic activity	Catalysis of a biochemical reaction at physiological temperatures. In biologically catalyzed reactions, the reactants are known as substrates, and the catalysts are naturally occurring macromolecular substances known as enzymes. Enzymes possess specific binding sites for substrates, and are usually composed wholly or largely of protein, but RNA that has catalytic activity (ribozyme) is often also regarded as enzymatic.
GO:0008528	0.400	0.022	peptide receptor activity, G-protein coupled	Combining with an extracellular or intracellular peptide to initiate a G-protein mediated change in cell activity. A G-protein is a signal transduction molecule that alternates between an inactive GDP-bound and an active GTP-bound state.
GO:0004012	0.400	0.025	phospholipid-translocating ATPase activity	Catalysis of the movement of phospholipids from one membrane face to the other (phospolipid 'flippase' activity), driven by the hydrolysis of ATP.

GO:0046030	0.400	0.025NA	Catalysis of the removal of one of the three phosphate groups of an inositol trisphosphate.
GO:0004000	0.400	0.025adenosine deaminase activity	Catalysis of the reaction: adenosine + H2O = inosine + NH3.
GO:0004165	0.400	0.025dodecenoyl-CoA delta-isomerase activity	Catalysis of the reaction: 3-cis-dodecenoyl-CoA = 2-trans-dodecenoyl-CoA.
GO:0004370	0.400	0.025glycerol kinase activity	Catalysis of the reaction: ATP + glycerol = ADP + glycerol 3-phosphate.
GO:0016757	0.400	0.025NA	Catalysis of the transfer of a glycosyl group from one compound (donor) to another (acceptor).
GO:0004652	0.400	0.026polynucleotide adenylyltransferase activity	Catalysis of the template-independent extension of the 3'- end of an RNA or DNA strand by addition of one adenosine molecule at a time. Cannot initiate a chain 'de novo'. The primer, depending on the source of the enzyme, may be an RNA or DNA fragment, or oligo(A) bearing a 3'-OH terminal group.
GO:0008188	0.267	0.027neuropeptide receptor activity	Combining with a neuropeptide to initiate a change in cell activity.
GO:0003705	0.400	0.027RNA polymerase II transcription factor activity, enhancer binding	Functions to initiate or regulate RNA polymerase II transcription by binding an enhancer region of DNA.
GO:0004386	0.258	0.027helicase activity	Catalysis of the reaction: NTP + H2O = NDP + phosphate to drive the unwinding of a DNA or RNA helix.
GO:0004872	0.286	0.028receptor activity	Combining with an extracellular or intracellular messenger to initiate a change in cell activity.
GO:0046983	0.273	0.033NA	The formation of a protein dimer, a macromolecular structure consists of two noncovalently associated identical or nonidentical subunits.
GO:0004004	0.273	0.036ATP-dependent RNA helicase activity	Catalysis of the reaction: ATP + H2O = ADP + phosphate, driving the unwinding of an RNA helix.
GO:0005234	0.278	0.039extracellular-glutamate-gated ion channel activity	Catalysis of the transmembrane transfer of an ion by a channel that opens when extracellular glutamate has been bound by the channel complex or one of its constituent parts.
GO:0004970	0.286	0.042ionotropic glutamate receptor activity	Combining with glutamate to initiate a change in cell activity through the regulation of ion channels.
GO:0005154	0.333	0.043epidermal growth factor receptor binding	Interacting selectively and non-covalently with the epidermal growth factor receptor.
GO:0004016	0.333	0.044adenylate cyclase activity	Catalysis of the reaction: ATP = 3',5'-cyclic AMP + diphosphate.
GO:0005249	0.300	0.045voltage-gated potassium channel activity	Catalysis of the transmembrane transfer of a potassium ion by a voltage-gated channel.
GO:0051015	0.333	0.045NA	Interacting selectively and non-covalently with an actin filament, also known as F-actin, a helical filamentous polymer of globular G-actin subunits.
GO:0046982	0.263	0.047NA	Interacting selectively and non-covalently with a nonidentical protein to form a heterodimer.

Table S19.

MW HKA valleys (2.5% lowest quantile, merged if within 10 kbp), and position relative to nearest gene. Regions analyzed: 2L:659000-15653000, 2R:6508000-19932000, 3L:411000-14556000, 3R:12601000-27224000, X:3701000-19185000 .

Table S20.

Gene ontology analysis of MW HKA low outliers

Table S21.

MW HKA peaks (2.5% highest quantile, merged if within 10 kbp), and position relative to nearest gene. Regions analyzed: 2L:659000-15653000, 2R:6508000-19932000, 3L:411000-14556000, 3R:12601000-27224000, X:3701000-19185000.

Table S22.

Diversity ratio valleys (see text) shared between MW *D. melanogaster* and the *D. simulans* data of Begun et al. (2007). Regions analyzed: 2L:3047000-15560000, 2R:6508000-19002000, 3L:3339000-14556000, 3R:12909000-26707000, X:3701000-18963000

Table S23.

Gene ontology analysis of MW HKA high outliers.

Table S24.

Diversity ratio valleys (2.5% lowest quantile, merged if within 10 kbp), and position relative to nearest gene. Regions analyzed: 2L:705000-15560000, 2R:6508000-19680000, 3L:411000-14556000, 3R:12909000-27218000, X:3701000-19185000.

Table S25.

Gene ontology analysis of heterozygosity ratio outliers.

Table S26.

Diversity valleys (see text) shared between MW *D. melanogaster* and the *D. simulans* data of Begun et al. (2007). Regions analyzed: 2L:3047000-15560000, 2R:6508000-19002000, 3L:3339000-14556000, 3R:12909000-26707000, X:3701000-18963000.