Supplemental Information

ChIP-seq

Strains were maintained at room temperature. Whole animal white pre-pupa material for each species was collected within a strict 20-minute time period, enabling comparison of stages across species. Duplicate ChIP-seq experiments were carried out using the standard modENCODE protocols (www.modencode.org). Sequencing data was generated by the High-Throughput Genome Analysis Core (HGAC) at the Institute for Genomics and Systems Biology.

We sequenced single-end ChIP libraries on the Illumina GAII platform. 36 base pair reads in FASTQ format were mapped to the appropriate reference genome using Bowtie (v.0.12.7; Langmead et al. 2009) with default parameters, outputting to the SAM format (-k 1-n 2 –e 70 –l 28 -maxbts 125 -S) unless otherwise noted. We used the following reference genomes: for *melanogaster*, release 5; for *simulans*, r1.3; for *yakuba*, r1.3; and for *pseudoobscura*, r2.15. A full list of replicates and sequencing statistics thereof can be found in Supplementary Table 1.

RNA-seq

RNA was isolated using TRIzol; integrity was checked with an Agilent Bioanalyzer. One replicate was performed for each species. Reads in fastq format were trimmed (fastx_clipper -a GATCGGAAGAGC -M 12 -Q 33) and mapped to the appropriate genome using Tophat (-p 8 -n 2 -g 1 -M; Trapnell et al. 2012). PCR

duplicates were removed using SAMtools' rmdup (Li et al. 2009). Normalized FPKM values were called using Cufflinks (-b -uw).

Read simulation

In order to determine the extent to which gene paralogs are mappable, we created simulated reads using the software ART (Huang et al. 2012). We made 36 bp reads at 20-fold coverage in each genome, using the Illumina sequencer error model. We then mapped the resulting reads using our default procedure and tabulated simulated read coverage as per our standard pipeline (see below).

In parallel, we also mapped the true locations of the reads in order to compute the "true" signal for each gene. These true locations are the exact spots in the genome from which the simulated reads were generated. Therefore, each gene was assigned a true alignment signal and a mapped read signal, and the difference between these was indicative of the degree to which 36nt Illumina reads mismapped to each gene.

Genes were excluded if the difference in mean signal per exon between true and simulated signal exceeded 20 reads, but results were qualitatively similar with thresholds of 1 and 10 reads as well.

Quantitative analysis

To prepare the quantitative analysis, we used BEDTools' coverageBed tool to count the number of reads on either strand occurring in specified regions (Figure S1;

Quinlan and Hall 2010). We used two types of regions: protein-coding sequences, as annotated in FlyBase release 5.32 gff files; and regular sliding windows, obtained via dividing the genome into 1kb bins using BEDTools' windowMaker command (see below).

When tabulating the amount of H3K27me3 signal per gene, we used the mean number of reads per exon. We chose to use a per-exon measure because we noticed that different genomes differ in terms of the mean number of exons annotated per gene, leading to systematic differences in H3K27me3 signal between genomes driven by annotation issues. For example, *D. melanogaster* has more than 8x as many annotated exons as *D. pseudoobscura* (138,474 in Mel; 16,771 in Pse); as a result, a summation of H3K27me3 signal over all exons would lead to *D. melanogaster* showing systematic enrichment in H3K27me3 signal. Further complicating matters, the additional, *melanogaster*-specific exons are not evenly distributed across all genes and show enrichment for particular gene ontology categories (data not shown), which would confound our analyses. We note that because we chose to average ChIP-seq read depth over the number of exons, DESeq is not appropriate as an analytical framework. We did not calculate the mean H3K27me3 signal per base pair because ChIP-seq produces read count data. As in RNA-seq approaches, per-base data has reduced reliability and statistical power (Anders and Huber 2010).

For non-melanogaster species, reads were mapped to the appropriate genomes using the latest reference sequences. Definition of protein-coding sequence was obtained via FlyBase's precomputed gffs from Release 5.32 (2010_09). To map non-melanogaster species reads into the melanogaster genome, we used UCSC's liftOver

(Fujita et al. 2011), employing the precomputed chain files and a minMatch parameter of 0.5 for all species (results were similar for two other minMatch parameter values). In general, we found that liftOver mapped reads accurately (Fig. S12), as long as regions which could not be mapped were excluded (such regions fall along the y-axis in Fig. S12).

Window-based comparison

As a further control against biasing our results by relying on single-copy orthologous genes, we also sought to compare H3K27me3 signal between species without reference to genic sequence. To this end, we matched ChIP-seq reads from each non- *D. melanogaster* species to the *D. melanogaster* genome (see Methods, Fig. S1; Hinrichs et al. 2006), then divided the genome into 1kb bins. We then compared signal between Mel and each other species using DESeq (Anders and Huber 2010).

With this procedure, we found at an FDR of .01 that <1% of *D. simulans*, 29% of *D. yakuba*, and 13% of *D. pseudoobscura* bins had undergone significant changes in H3K27me3 levels. Similar to our genic comparison (above), we do not observe evidence of a linearly decreasing trend in H3K27me3 with divergence time. The same qualitative trend was observed even after limiting the analysis to regions with shared orthology in all species, indicating that genome assembly quality does not drive the observed differences.

To obtain correction values for use in the windowed analysis, we created mock read files consisting of a 36bp read mapping to every single base pair. We then

mapped this mock file to the melanogaster genome using liftOver. From the resultant file, we excluded bins which failed to map reads to each base pair in the window (i.e. for which the coverage was less than 100%). This yielded 27.3 million bp of comparable sequence for *simulans*, 40.4 million bp for *yakuba*, and 14.3 million bp for *pseudoobscura*. For all mappable windows, we assigned a correction factor to the observed read number equal to (read number in the lifted mock species file / read number in the mock *D. melanogaster* file). Resulting read counts were then rounded to the nearest whole number.

Duplicated gene analysis

To assign locations and distances to duplicated genes, we first segregated genes which remained on the same chromosome arm from those which did not. Following Schaeffer et al. 2008, we considered gene translocations to be intrachromosomal if both copies of a duplicated gene were in the same Muller element (chromosome arm), or interchromosomal otherwise. For intrachromosomal duplicates, distance was defined as the absolute difference in map positions of a gene in one species and its ortholog in the other species. Location-based analysis was only performed in comparisons of *D. melanogaster-D. simulans* and *D. melanogaster-D. yakuba*, due to the abundance of chromosomal translocations and rearrangements between *D. melanogaster* and *D. pseudoobscura* (Schaeffer et al. 2008).

We assigned ages to gene duplicates by considering the oldest species in which the same duplication was present.

Pseudogene location analysis

In order to examine the H3K27me3 signal occurring in the locations of recently duplicated genes/pseudogenes, we extracted the two flanking regions (both 500bp) around each recently duplicated gene or pseudogene (as annotated in the ortholog list and the FlyBase 5.32 gff file, respectively). These two flanking regions were mapped into the *D. simulans* genome using liftOver (minMatch parameter = .5, as above). To limit our analysis to very recent duplication events, we only analyzed the comparison of *D. melanogaster-D. simulans*. If both flanking regions mapped, and the distance between the two was less than 5kb, the intervening region was considered the new gene or pseudogene insertion region. H3K27me3 signal in this region was then tabulated as described above.

We note that this method of analysis is formally identical to ancestral state reconstruction under a Brownian motion model of trait evolution, in which only one extant taxon is available (we cannot consider *D. melanogaster*'s state because of the influence of the duplication itself, and applying the same mapping approach to *D. yakuba* added very few additional insertion regions; Schluter et al. 1997). In this case, the Maximum Likelihood Estimate of the ancestor's trait value (H3K27me3 signal) is equal to the single taxon's H3K27me3 signal (where the single taxon is *D. simulans*). While this method is likely to be inaccurate for any single location's ancestral H3K27me3 signal, there is no reason to believe it is systematically biased, especially biased differently between new genes and pseudogenes.

Statistical analysis

To call individual intervals as significantly diverged in different species, we used DESeq (Anders and Huber 2010). In rough outline, it asks whether variation in read counts between treatments exceeds variation in read counts between replicates. More technically, DESeq fits observed data to a generalized linear model using the negative binomial distribution as a link function. Though originally designed for comparison of RNA-seq data under different experimental conditions, in principle it is appropriate for any read count data which follows a negative binomial distribution (which our normalized, filtered ChIP-seq read count data does; Figure S13; likelihood ratio test (Hoaglin, Mosteller, and Tukey 1985; Friendly 2000), p-values in Table S2).

In order to take into account the control samples, we subtracted the read counts in the control sample from each matched ChIP replicate. We excluded bins in which there was no enrichment in the ChIP samples (for which the maximum read count for all samples was 0 or less), and assigned a value of 0, corresponding to no enrichment, to bins in which the total number of reads was less than zero after input subtraction.

Following this normalization, we proceeded to use DESeq's standard analysis pipeline, which estimates library size correction factors automatically in order to normalize the mean fold change to zero. We used the estimateDispersions function with fitType="local". Otherwise, all parameters in DESeq were run at default settings.

To estimate a False Discovery Rate (FDR), we employed the q-value package in R (Storey 2003).

To estimate the significance of differences between two populations' sample means, we used permutation tests as described in Sokal and Rohlf (pg. 808; called therein "sampled randomization" tests). These results were confirmed with Wilcoxon tests, and p-values were similar in most cases. To estimate the significance of observed correlations, we randomly reordered the correlated observations 10000 times and counted how many random orderings had a correlation coefficient at least as extreme as the observed correlation coefficient. All statistical analysis was performed in R.

In order to call individual genes as significantly diverged, we calculated a z-statistic for each gene, defined as:

$$Z_i = \frac{X_i - \mu}{\sigma^2}$$

Where X_i is equal to the mean per exon H3K27me3 signal difference (between *D. melanogaster* and another species) of the ith gene, calculated from two biological replicates; μ is the mean signal difference (thus normalizing for differences in library size); and σ^2 is the standard deviation of all the signal differences. To obtain a null distribution for this statistic the matrix of replicates was randomly permuted 100 times, and the Z-statistic was calculated for each row. The reported p-value for each gene is equal to the number of permuted Z-statistics from the null distribution with values more extreme than the actual Z-statistic divided by the total number of permuted Z-statistics.

Sequence evolution

To determine the relative number of substitutions in our bin-based analysis, we used Galaxy to parse UCSC's pairwise alignment files (Blankenberg et al. 2011). We first determined the number of substitutions in each bin, discarding bins for which there was no alignment. We then repeatedly randomly selected a number of bins (equal to the number of diverged bins) and asked how often the randomly selected bins had a mean number of substitutions greater than that observed for diverged bins.

For genic d_N/d_S, we used the data from Begun et al. 2007. For orthologous gene identification, we used the modENCODE list of orthologous genes (Boyle et al., in preparation). Results were qualitatively consistent when using Flybase orthologs as well (data not shown).

Supplementary Figure Captions

Supplementary Figure 1: Experimental Analysis Scheme

Diagram of analysis procedure. Two replicates and one control were collected for each experiment. The resulting reads were aligned to the appropriate genome with Bowtie using default parameters.

For the genic analysis, reads were counted within the exons (as defined by Flybase's precomputed gffs). Each gene was assigned an H3K27me3 signal equal to the mean number of reads per exon. Genic H3K27me3 signal was compared using the modEncode list of orthologs.

For the binned analysis, if the experiment was non-*melanogaster*, reads were mapped to the melanogaster genome with liftOver using a minMatch parameter of 0.5. Reads were then counted within each bin, giving a bin signal value which could be compared between species.

Supplementary Figure 2: Genic replicate experiments are highly correlated

Each dot represents the exonic H3K27me3 signal in a gene from *Drosophila melanogaster*. The x-axis is the signal for one replicate, and the y-axis is the signal for the other. The replicates correlate at a Spearman's rho of .963.

Supplementary Figure 3: H3K27me3 is highly correlated with gene expression

Each dot is a single gene, and the graph depicts the correlation between log H3K27me3 signal (x-axis) and gene expression in log FPKM. H3K27me3 signal and gene expression are negatively correlated at an SCC = -.42.

Supplementary Figure 4: *Distal-less* H3K27me3 domain is conserved

A genome browser snapshot over the *Distal-less* locus shows strong conservation of the H3K27me3 domain. Top two rows are H3K27me3 signal from *D. melanogaster*; third row is the input (control) signal. In the middle area, in grey, is the *Distal-less* gene model, along with nearby genes on the left (both Watson and Crick gene models are shown). The next two rows are the H3K27me3 signal from the orthologous location in *D. yakuba*, and the corresponding control signal. Finally, the last row shows *D. yakuba*'s gene models. Both vertical lines in each pane are at the transcription start site

of each species' *Distal-less* gene. Note that the H3K27me3 domain is conserved across species with nearly identical boundaries.

Supplementary Figure 5: Melanogaster to elegans comparison at embryo time stage.

As in figure 2A, but using data from the embryo, 0-4 hours (E0-4H) stage in *melanogaster* and the Early Embryo (EE) data in *elegans*. Similar to Figure 2A, the Spearman's rho = .46, indicating a moderately strong correlation.

Supplementary Figure 6: Terminology

We consider three types of comparison within the paper. Branches indicate gene-specific phylogenies; the upper grey line denotes the speciation event, the lower grey line denotes gene duplication events in panels B and C. Finally, brackets denote the genes whose H3K27me3 signal is being compared for each term.

Panel A: In the first sections, we address change in H3K27me3 signal within single-copy orthologs, which are pairs of orthologous genes which are present in a single copy in both genomes. In this example, we consider H3K27me3 signal differences between A1 and B1.

Panel B: Later, we examine duplicated genes. We compare each duplicated genes' H3K27me3 signal to its single-copy ortholog's H3K27me3 signal in the other genome. Unlike single-copy orthologs, then, this compares multiple duplicated genes to single orthologs. In the schematic, two comparisons are therefore made: A1-B1, and A2-B1.

Panel C: Finally, we examine the relationship of duplicated genes' H3K27me3 signal to each other. Such genes are paralogous with respect to each other, since they have arisen by a gene duplication event. In order to accurately infer paralogy relationships, we restrict ourselves in these analyses to pairs of duplicated genes; in this schematic, we compare A1-A2.

Supplementary Figure 7: Comparison of simulated vs. actual read mapping to paralogs.

Reads were simulated using ART, and then genic coverage of both simulated, mapped reads was compared to the coverage computed based on the true locations of reads. Simulated, mapped reads' genic signal (x-axis) are plotted against the true signal (y-

axis). There is strong agreement between the two: a Pearson correlation coefficient of r = .99.

Supplementary Figure 8: Bootstrap confidence is not correlated with change in H3K27me3.

On the x-axis is bootstrap confidence; on the y-axis is the absolute change in H3K27me3 signal between any pair of orthologous genes (each dot represents one gene comparison). If poor paralog identification is at the root of the observed differences between ortholog sets (Fig. 3), then one would expect that bootstrap confidence would be correlated with absolute change in H3K27me3. In fact, these are not significantly correlated (SCC = .02).

Supplementary Figure 9: Differences in H3K27me3 correlation by ortholog set with large gene families excluded.

As in Figure 3C, but with all gene families of size 3 or more excluded. General trends still hold: duplicated gene orthologs are less correlated than single-copy orthologs.

Supplementary Figure 10: Comparison of H3K27me3 signal in promoters of orthologous genes yields the same trends.

As in Figure 3C, but considering the correlation between the promoters of single-copy and duplicated gene orthologs. We expect that promoter sequences suffer less from read mismapping. Note that the same trends hold for promoters' H3K27me3 signal as for genes (Figure 3C).

Supplementary Figure 11: Bin divergence is connected to sequence divergence.

Panel A: Bins which have undergone divergence experience significantly more substitutions than randomly selected bins. Each pair of bars is a species (left to right: *simulans*, *yakuba*, *pseudoobscura*): the left bar represents the actual data, while the right bar represents permuted data. Lines are the 95% bootstrapped confidence intervals. Differences between categories for *D. yakuba* and *D. pseudoobscura* are significant at $p < .01$.

Supplementary Figure 12: Signal in mapped and native genome correlates.

This figure compares genic H3K27me3 signal derived from two methods: native genome *D. simulans* H3K27me3 as computed in Supplementary Figure 1, and reads lifted over from *D. simulans* into the *D. melanogaster* genome. For the latter method, genic H3K27me3 was computed using melanogaster gene definitions. The resulting genic H3K27me3 signals were plotted against each other by using the modEncode ortholog list. The two methods are highly correlated (SCC = .888), but note that some genes fail to show any H3K27me3 signal because reads cannot be lifted over.

Supplementary Figure 13: Example negative binomial-ness plot

To check whether binned H3K27me3 data could be analyzed using a negative binomial GLM (as employed in DESeq), we used Friendly's concept of a negative binomial diagnostic plot and goodness of fit test for Negative Binomial data. The plot depicts the expected (red line) and actual (dots) number of occurrences, with confidence intervals around the actual numbers. If the data fits a negative binomial, then the actual dots should line up on and around the red line. The melanogaster data from the *D. melanogaster - D. simulans* comparison is shown here. In this case, the goodness of fit test p-value was $<10^{-250}$.

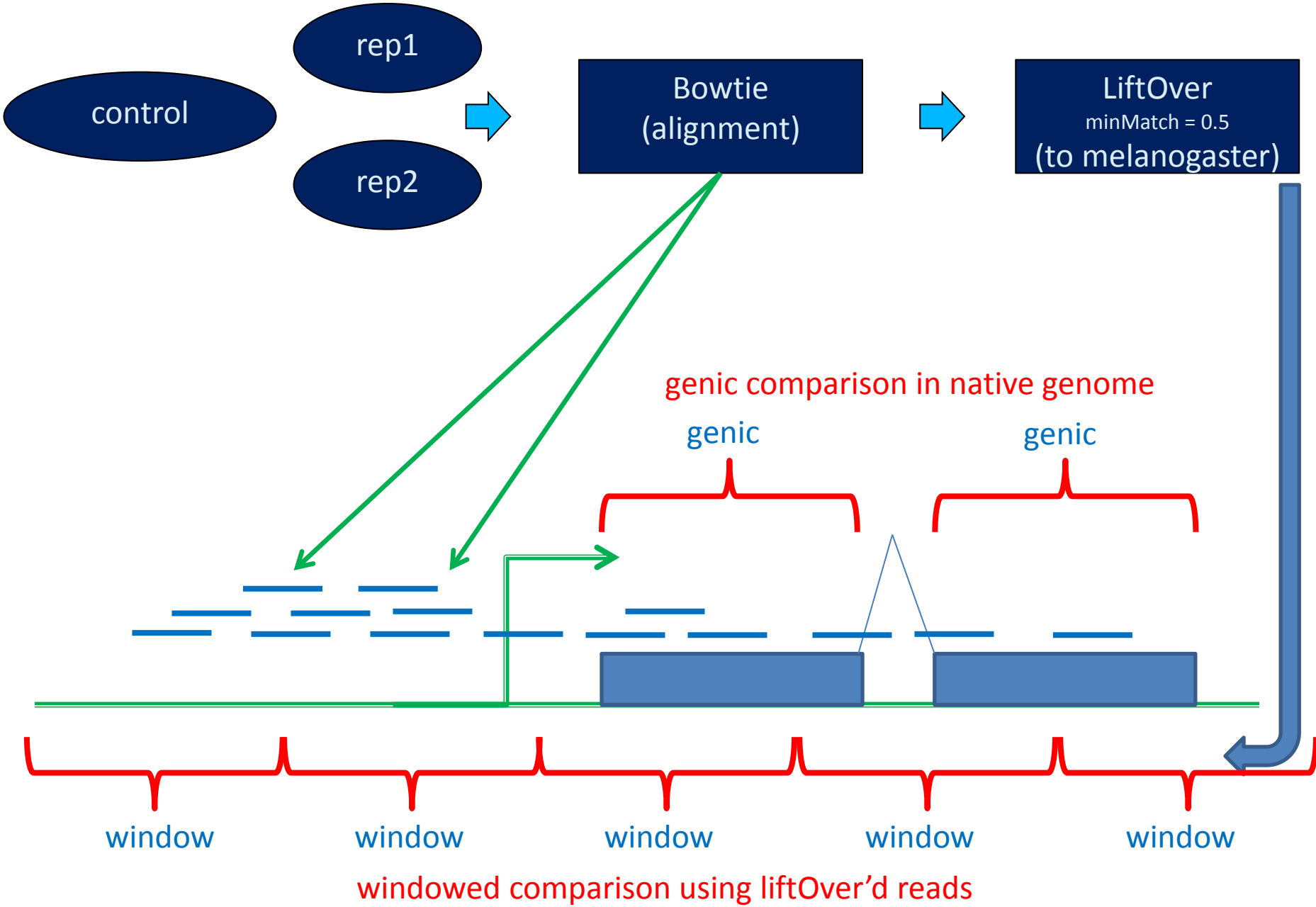
Supplementary Table 1: Sequencing statistics

Sequencing statistics for each experiment and replicate. The first column is an internal dataset control number; the second is the description of the experiment; the third is the

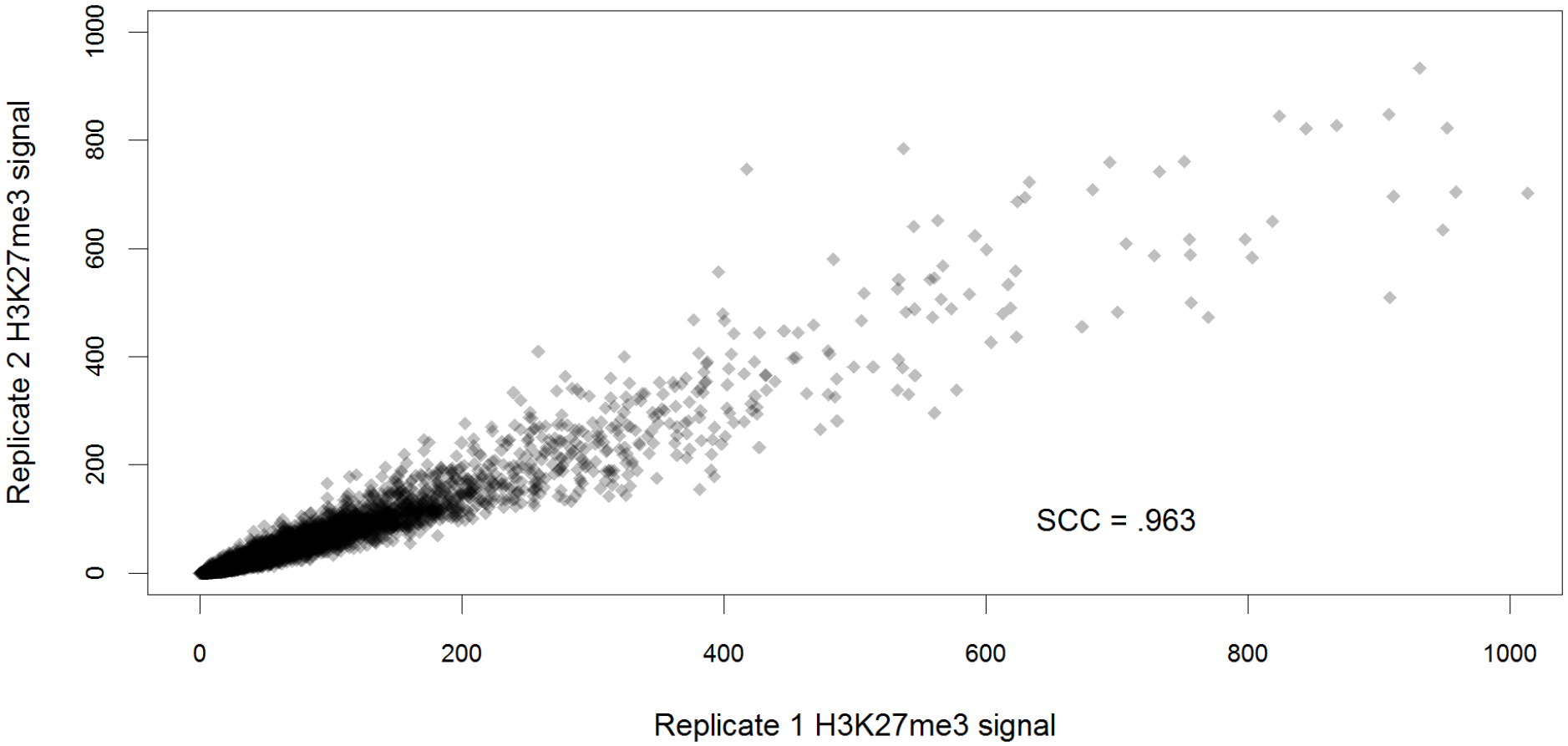
total number of reads, the fourth is the number of reads which map uniquely, and the fifth is the percent of the reads which mapped uniquely.

Supplementary Table 2: Results of negative binomial goodness of fit tests

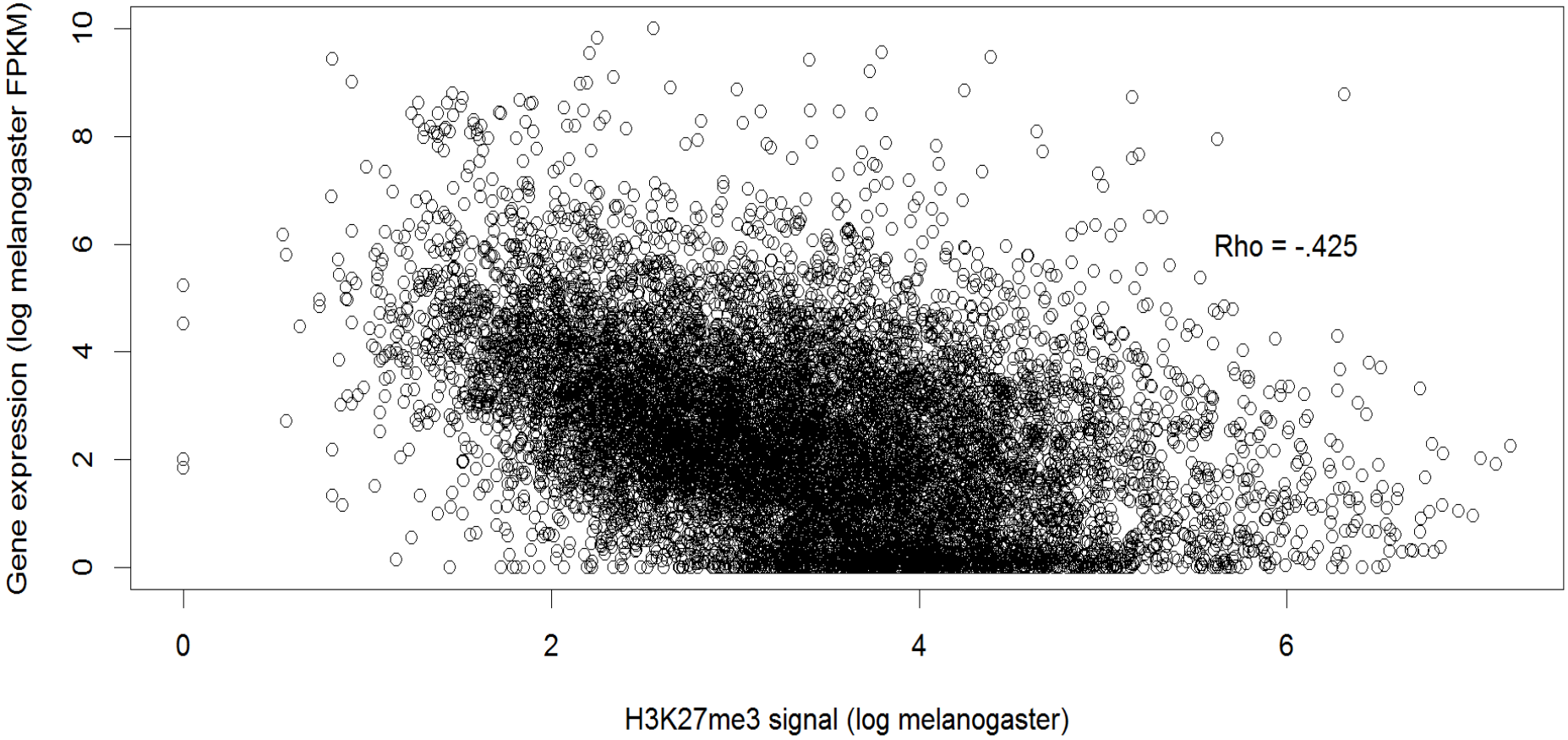
We performed Friendly's (2000) goodness of fit test for negative binomial data. Low p-values indicate agreement with a fitted negative binomial distribution. The first column indicates the pairwise comparison, the second is the particular replicate, and the third is the p-value.



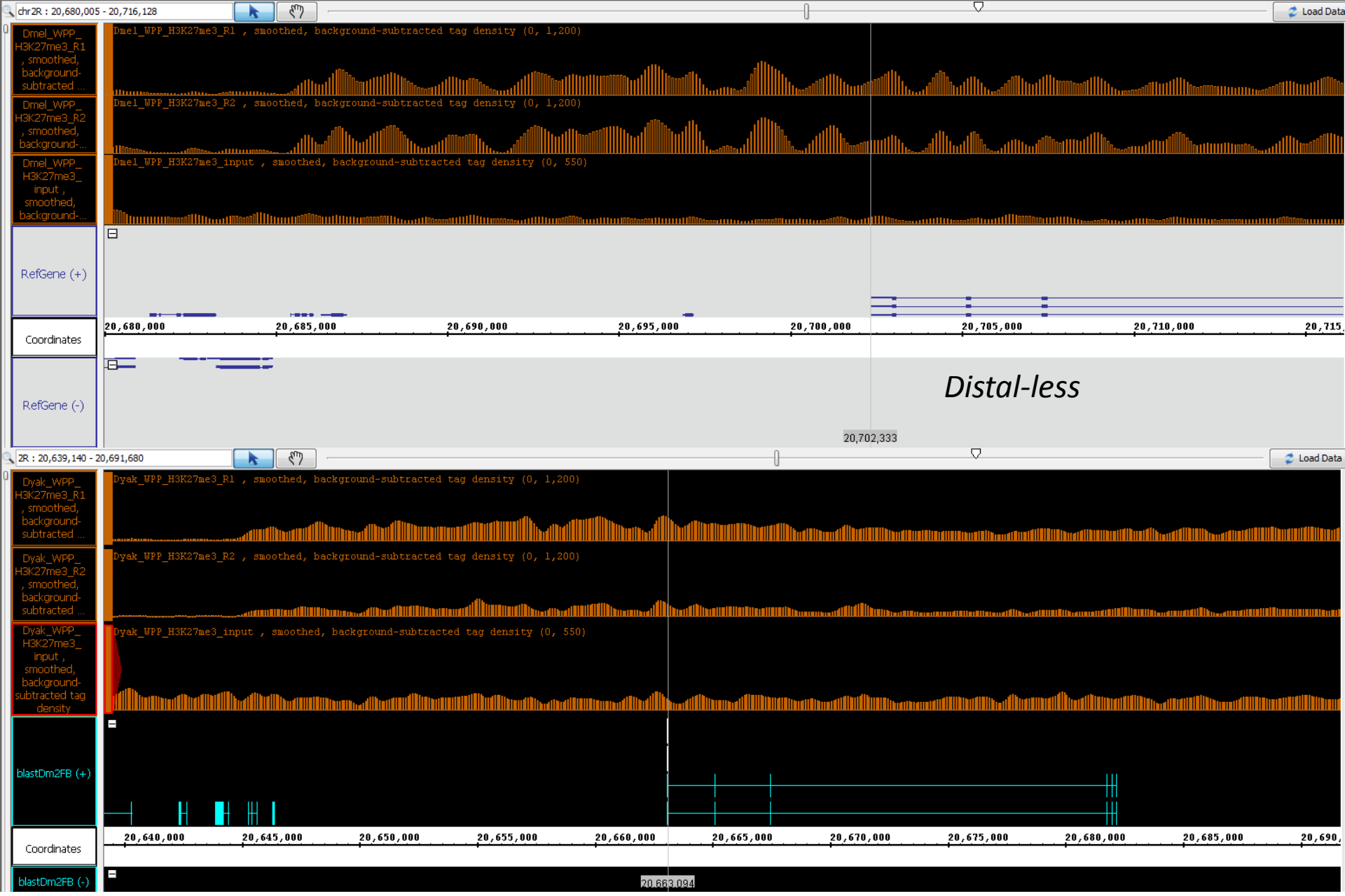
Supplementary Figure 1



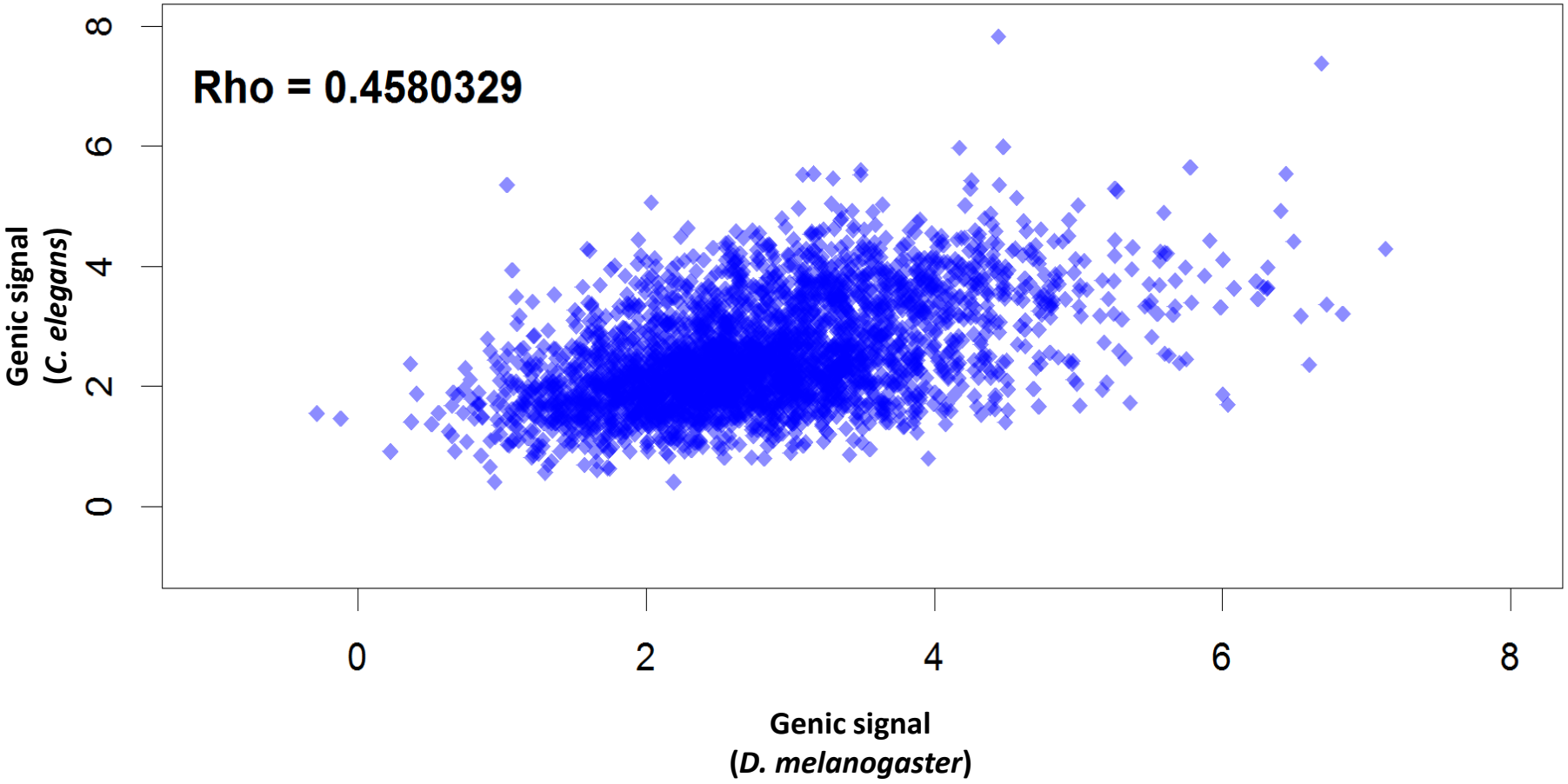
Supplementary Figure 2



Supplementary Figure 3

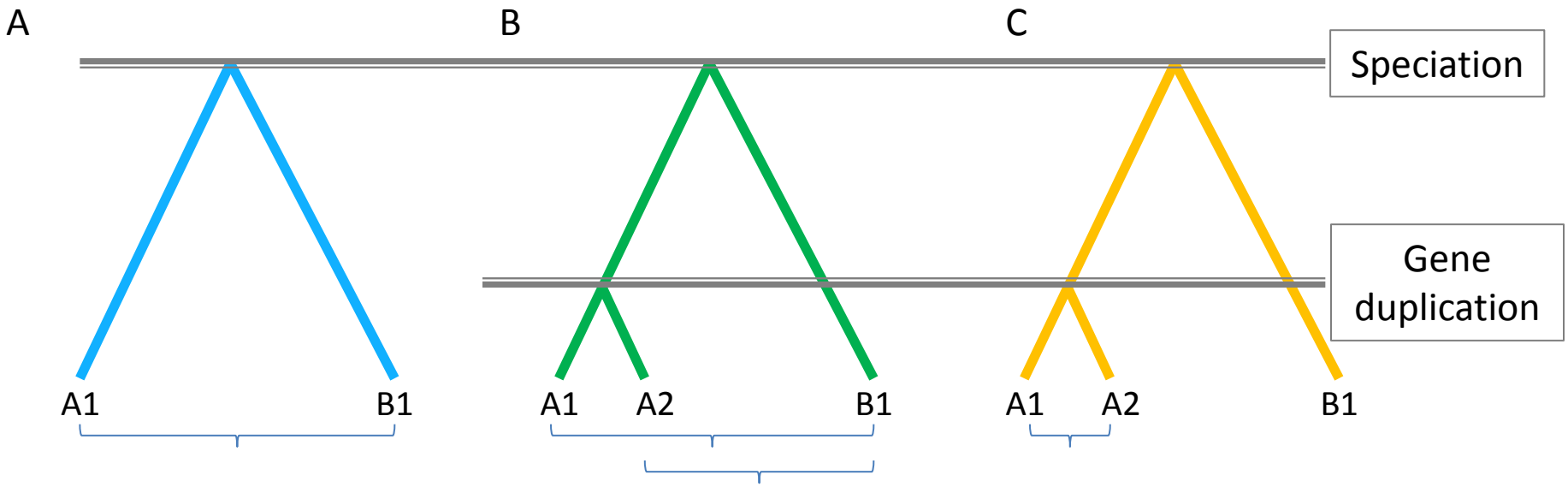


Supplementary Figure 4



Supplementary Figure 5

Terminology



Single-copy orthologs

Duplicated gene orthologs

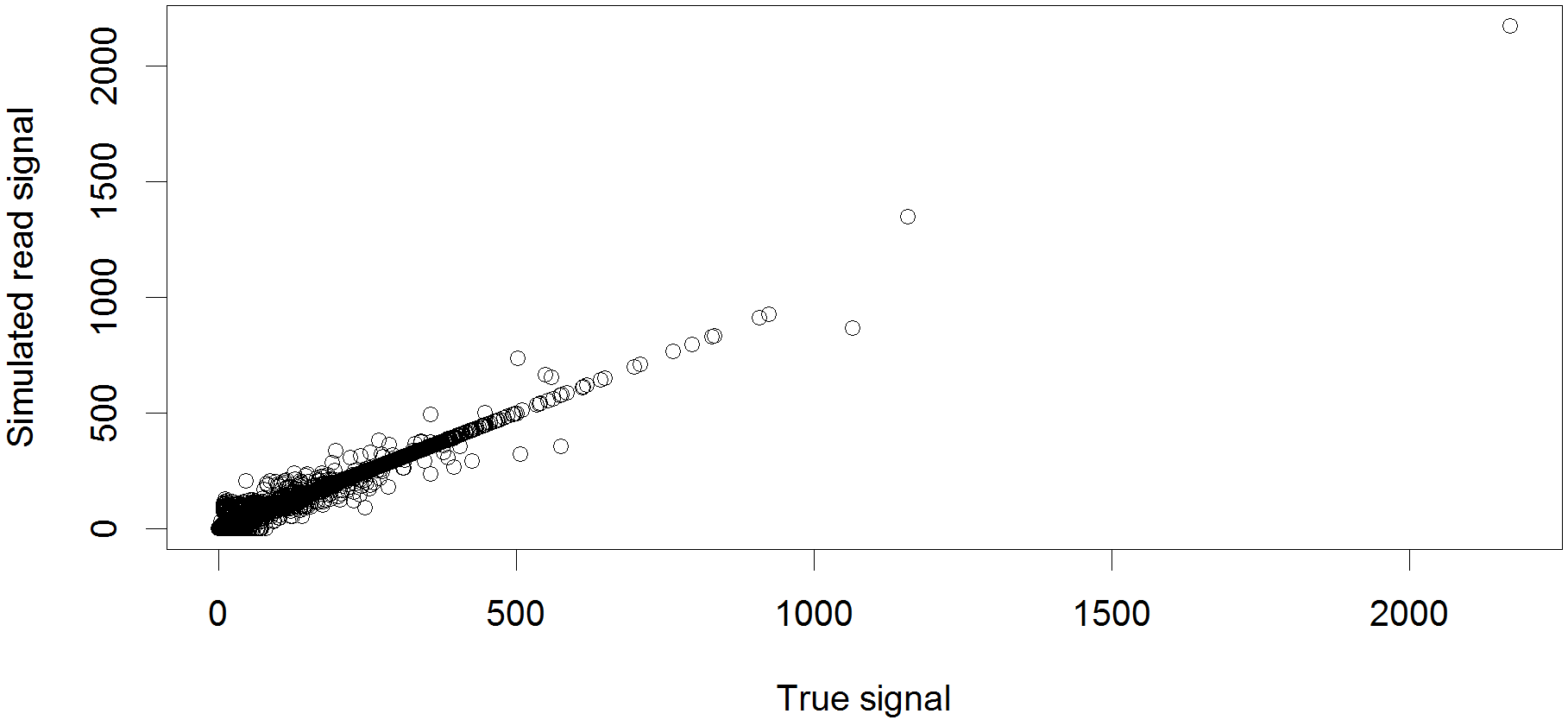
Duplicated gene paralogs

Single-copy orthologs are present in both genomes in one copy.

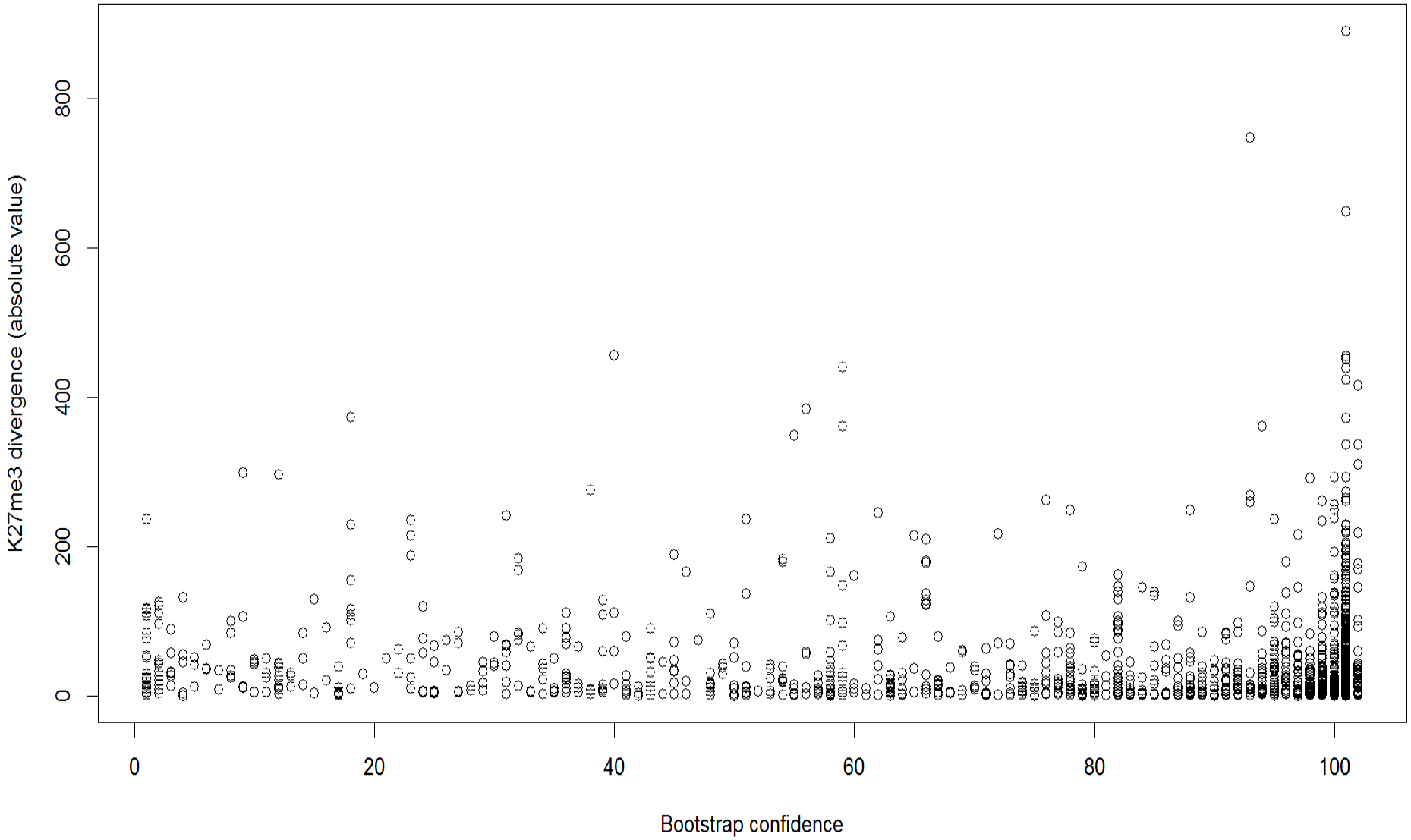
Duplicated gene orthologs are present in two or more copies in one genome (due to a gene duplication event), but only one copy in the other genome.

A set of duplicated genes in one genome are considered to be *paralogous* with respect to each other.

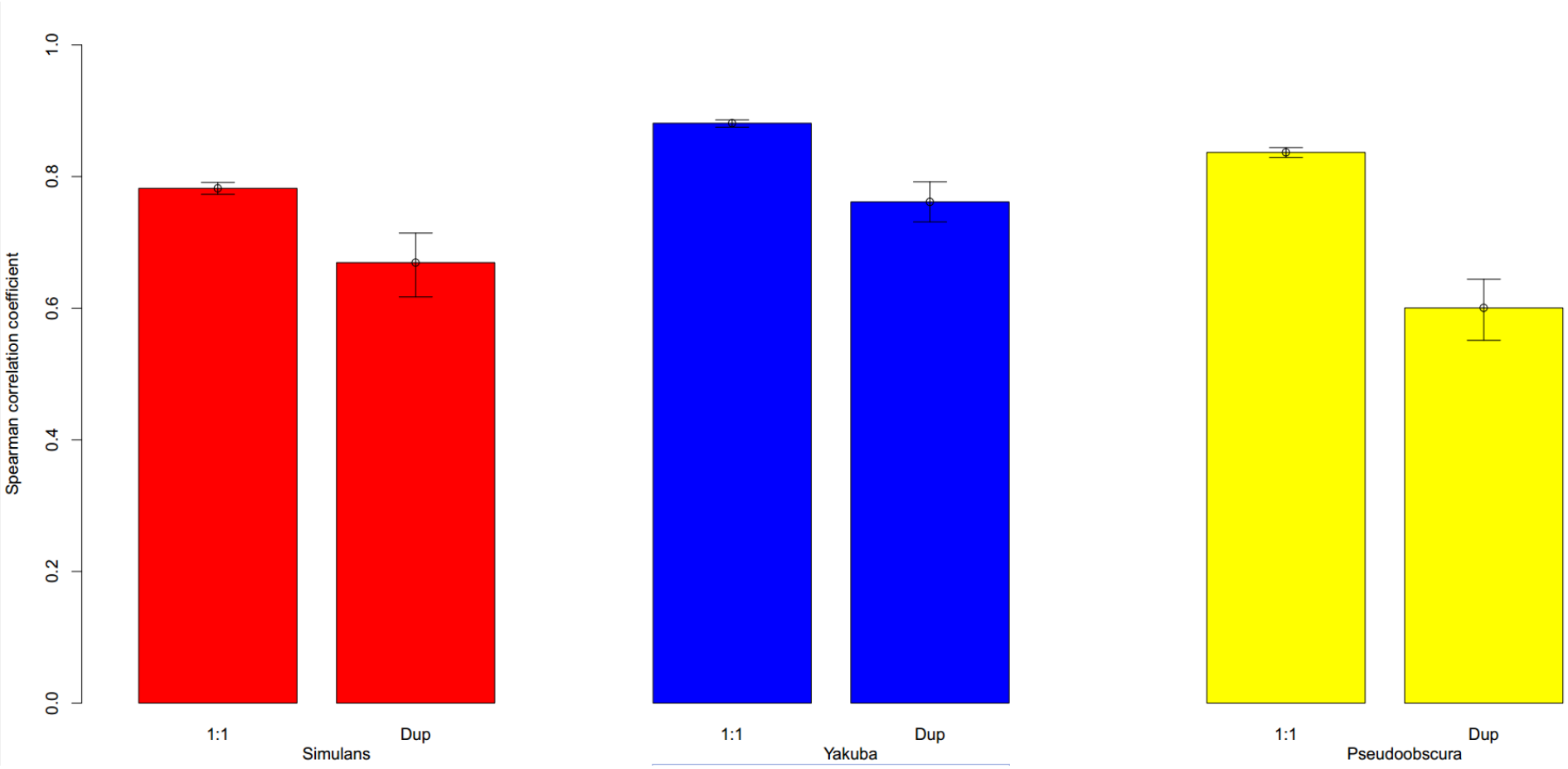
Supplementary Figure 6



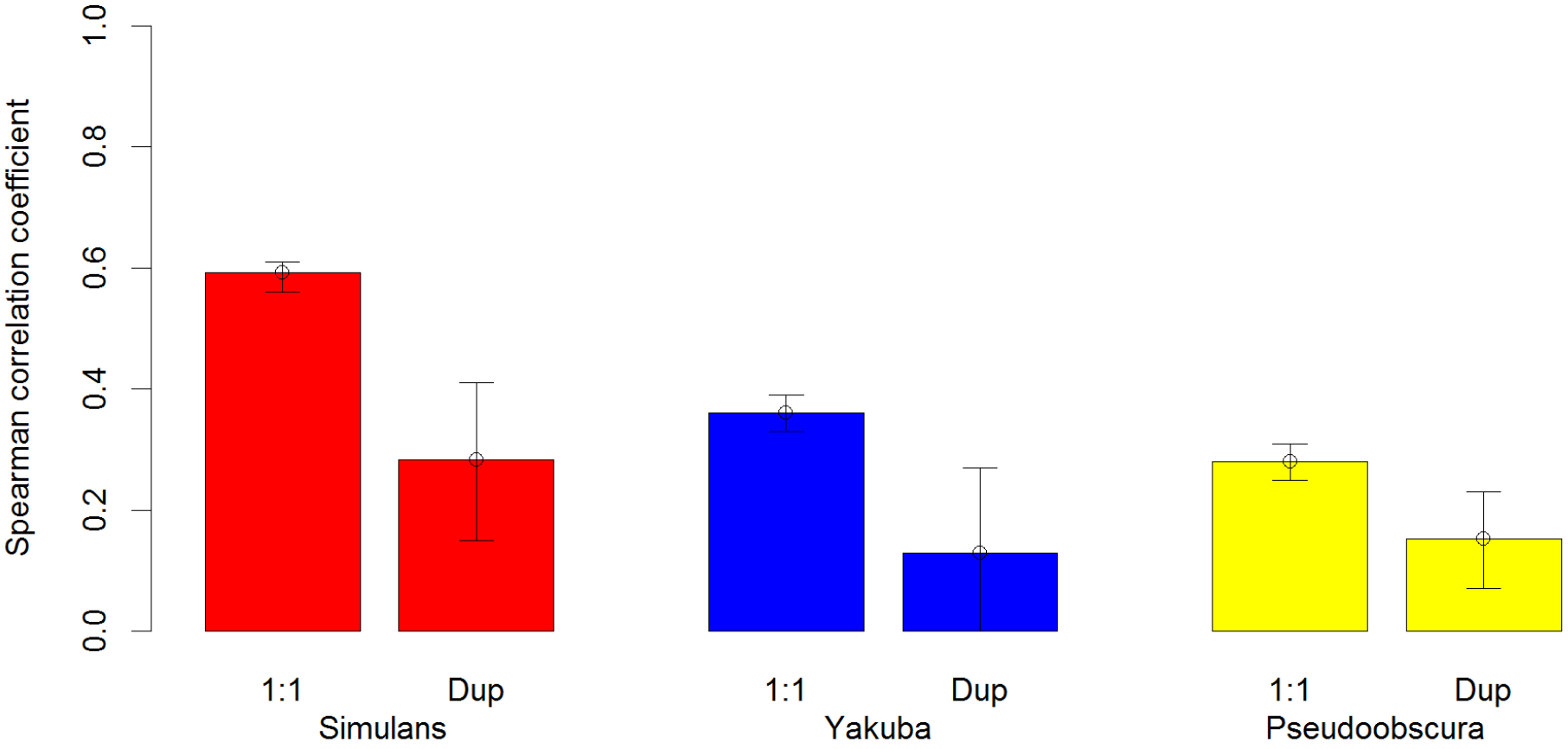
Supplementary Figure 7



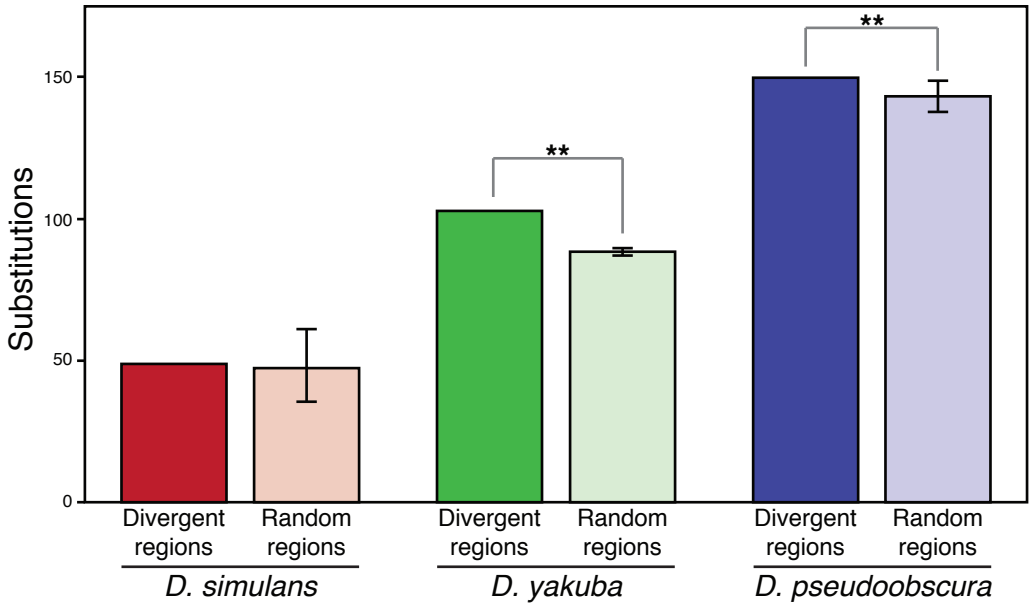
Supplementary Figure 8



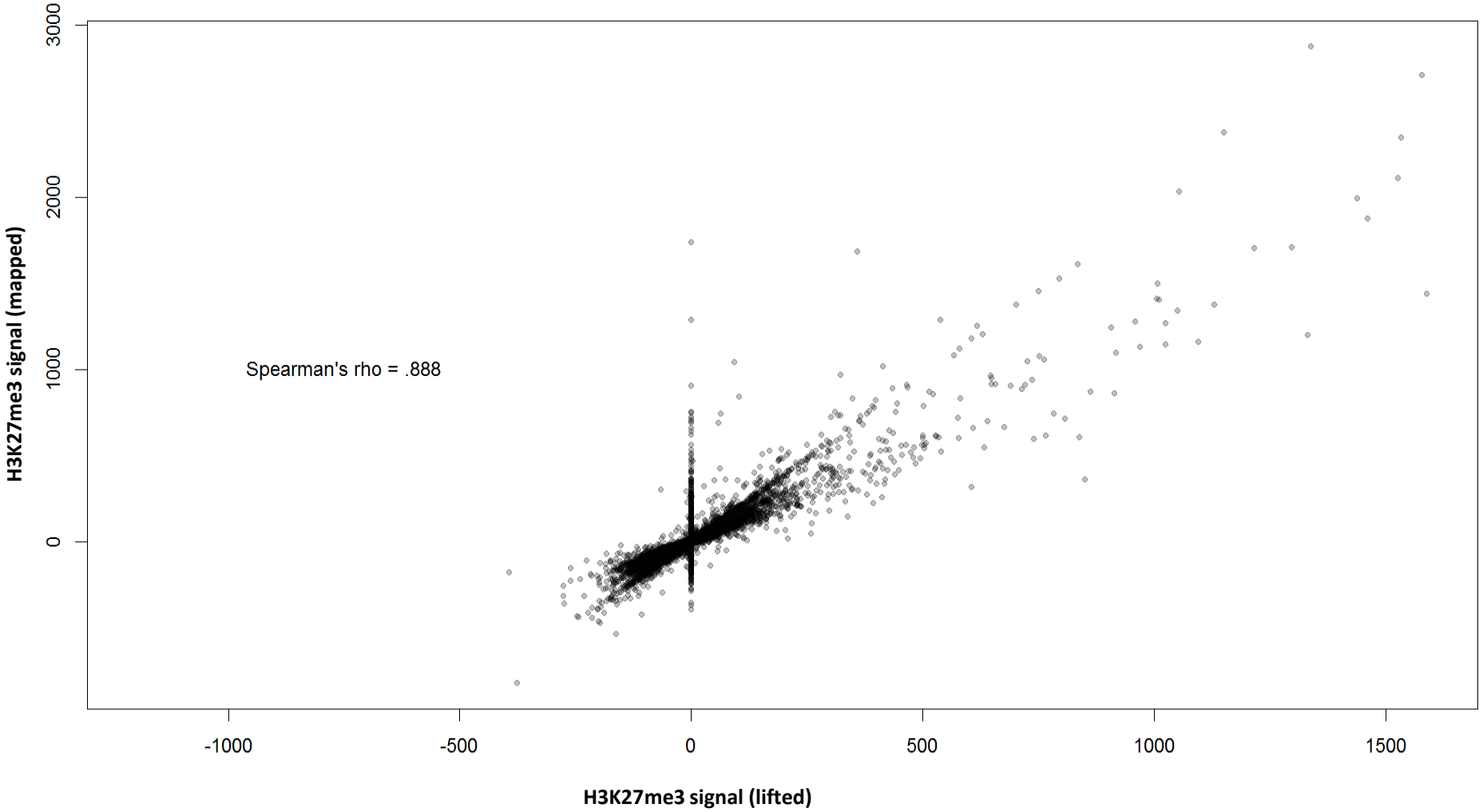
Supplementary Figure 9



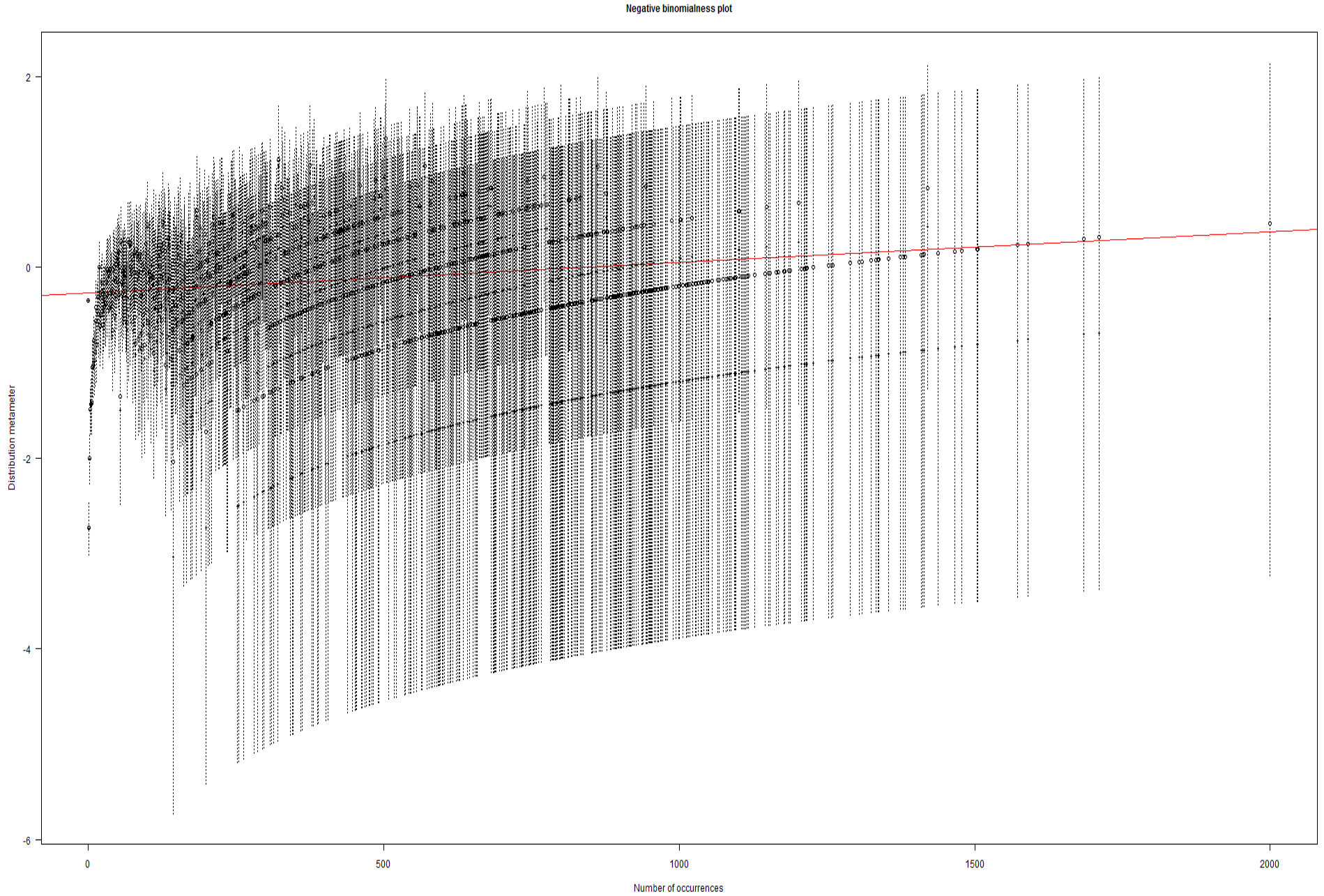
Supplementary Figure 10



Supplementary Figure 11



Supplementary Figure 12



Supplementary Figure 13

dataset	description	total reads	reads mapped	percent reads m	included?
2010-1790	mel WPP R1	19706651	16940628	0.859640129	yes
2010-1791	mel WPP R2	19932071	17416264	0.873780953	yes
2010-1793	mel WPP control	18604075	15877797	0.853458019	yes
2010-828	pse WPP R1	15589085	14837325	0.951776515	yes
2010-829	pse WPP R2	20118797	19542517	0.97135614	yes
2010-831	pse WPP control	13584227	12957515	0.953864729	yes
2010-1790	mel WPP R1	19706651	16940628	0.859640129	yes
2010-1791	mel WPP R2	19932071	17416264	0.873780953	yes
2010-1793	mel WPP control	18604075	15877797	0.853458019	yes
2010-824	sim WPP R1	20811449	19336143	0.929110847	yes
2010-825	sim WPP R2	17009025	16010692	0.941305689	yes
2010-827	sim WPP control	19562229	18073324	0.923888786	yes
2010-1790	mel WPP R1	19706651	16940628	0.859640129	yes
2010-1791	mel WPP R2	19932071	17416264	0.873780953	yes
2010-1793	mel WPP control	18604075	15877797	0.853458019	yes
2010-832	yak WPP R1	23441200	23085300	0.98481733	yes
2010-833	yak WPP R2	24247767	23841889	0.983261222	yes
2010-835	yak WPP control	26590770	25651377	0.964672215	yes

comparison	dataset	p-value for Friendly's goodness of fit test
Mel-Sim	Mel1	0
Mel-Sim	Mel2	0
Mel-Sim	Sim1	0
Mel-Sim	Sim2	1.16E-262
Mel-Yak	Mel1	0
Mel-Yak	Mel2	0
Mel-Yak	Yak1	0
Mel-Yak	Yak2	0
Mel-Pse	Mel1	1.13E-296
Mel-Pse	Mel2	0
Mel-Pse	Pse1	2.15E-268
Mel-Pse	Pse2	8.34E-236