SUPPLEMENTAL METHODS

Resequencing and genome assembly

Production of the *zhr* and *z30* genomes was previously described in Coolon et al. (2012). To construct the *D. simulans tsimbazaza* and *D. sechellia droSec1* genomes, gDNA sequence reads from *D. sechellia droSec1* and *D. simulans Tsimbazaza* were aligned to the *D. simulans* and *D. sechellia* genome assemblies respectively (Drosophila 12 Genomes Consortium 2007) using BWA (Li and Durbin 2010) (version 0.5.6). Each read was aligned separately using default parameters, and SAM format files were generated using the BWA *sampe* command. Alignment files were converted to BAM format, and VCF files describing single nucleotide polymorphisms (SNPs) and indels were created using the SAMtools package (Li et al. 2009) (version 0.1.7a; modules view, sort, and pileup). SNP and indel calls were filtered using the samtools.pl varFilter command (as described at http://samtools.sourceforge.net/cns0.shtml) to retain SNPs and indels with *phred* scale quality scores of 20 or higher.

The public reference genomes of *D. simulans* and *D. sechellia* were originally sequenced at a coverage depth of 3- and 5-fold, respectively. This low coverage left large genomic regions unfinished. To close these gaps, we realigned gDNA sequences to the SNP and indel corrected genomes. Unmapped read-pairs were assembled into contigs with Velvet (Zerbino and Birney 2008). Contigs whose ends both aligned to the genomes were considered "gap spanning", and extended 100 bp in each direction. Velvet assembled contigs (including gap-spanning) were aligned to the *D. melanogaster* reference genome (dm3) using LASTZ (Harris 2007). Contigs that aligned uniquely to *D. melanogaster* were retained as the "extra genome", and comprised 12.6 and 1.6 Mb of sequence from *D. simulans* and *D. sechellia*, respectively. liftOver coordinate files were assigned to extra-genome contigs using the axtChain, chainNet, and netChainSubset utilities from the UCSC Genome Browser (Kent et al. 2002). gDNA sequence reads were then remapped to identify SNPs and indels in the extra genome, and genome sequences were rewritten accordingly.

Despite our initial inbreeding, SAMtools identified residual heterozygosity at some sites in each genotype. This complicates allele-specific RNA-seq when one of the alleles segregating in strain 1 matches the allele that is invariant in strain 2. For example, consider a site polymorphic for A and C in *zhr,* but fixed for C in *Tsimbazaza.* RNA-seq reads originating from the *Tsimbazaza* allele would align to genome sequences for both strains, whereas reads originating from the *zhr* allele would be align only to the *zhr* genome. To eliminate the impact of such sites on measures of allele-specific expression, we changed such sites to the polymorphic genotype in both strains using a custom Perl script (snp_compare_filter.pl), effectively making these sites uninformative for allele assignment and producing comparison and strain specific genomes.

These comparison- and strain-specific genome sequences were produced using a custom Perl script (snp_adder.pl). This script sequentially rewrites the corresponding genome with corrected SNP calls and indels. The positions of insertions and deletions were recorded in custom liftOver chain files during the rewriting process. These chain files allowed the conversion of genomic features between strains and reference genomes using the UCSC Genome Browser

liftOver tool (http://genome.ucsc.edu) (Kent et al. 2002). All genome and chain files are available upon request.

Mapping sequencing reads to genes and alleles

We aligned each mate of the paired-end RNA-seq reads separately to the strain- or species-specific genomes specific to each comparison using the MOSAIK aligner (version 1.0.1384, http://bioinformatics.bc.edu/marthlab/Mosaik). For example, in the *mel-mel* comparison, reads derived from *zhr*, *z30*, and the F_1 hybrids from reciprocal crosses between *zhr* and *z30* were each aligned to both the *zhr* and *z30* genomes that had been created specifically for the *mel-mel* comparison. Aligning reads to both parental genomes prevents the biased mapping described in prior RNA-seq studies of allele-specific expression (Degner et al. 2009; Stevenson et al. 2013). The following command line options were used for these alignments: -hs 13 –mm 0 –p24 –mph 100 –act 20. The 13 base hash size (–hs 13) option allowed >99% of ambiguous base containing regions to be seeded for alignment by MOSAIK. Reads aligning uniquely with no mismatches to one or both genomes were considered "mapped reads" and retained for analysis. After the initial 76 bp reads were aligned to both reference genomes, reads that did not map to either genome were trimmed 13 bases from the 3' end using a custom Perl script (fastq_trimmer.pl) and again aligned using MOSAIK. This was repeated three times (sequence lengths 76bp, 63bp, 50bp, 37bp). Any sequences that did not uniquely align after the final iteration were discarded.

Using the chain files created in the genome assembly process, we converted the genome coordinates from the *zhr, z30, droSec1,* and *Tsimbazaza* genomes to the sequenced *D. melanogaster dm3* coordinates using the liftOver utility from the UCSC Genome Browser (Kent et al. 2002) (http://genome.ucsc.edu) and a custom Perl script (convert.pl). Sequence reads were then filtered based on their alignment to a previously identified set of constitutively expressed exons within the *D. melanogaster* genome (McManus et al. 2010) using the intersectBed module of BEDTools (Quinlan and Hall 2010), with reads not aligning to these regions discarded. Additionally, overlapping regions in the constitutive exon set were removed using intersectBed module of BEDTools and custom scripts. Gap files were produced for each comparison and combined using the mergeBed module of BEDTools to create one gap file used for all comparisons. Sequences containing gaps in one or more genotypes were excluded. Reads were assigned to alleles based on alignments to strain-specific genomes using a custom Perl script (classify.pl). Because paired-end reads are derived from a single transcript, each set of paired-end reads was treated as a single read count regardless of whether one or both reads were mapped successfully.

Pyrosequencing

To evaluate the reproducibility of expression measurements derived from our RNA-seq data, we used pyrosequencing (Qiagen) to independently measure differences in total and allelic expression in the *mel-mel* and *sim-sech* comparisons. We focused our validation efforts on the *mel-mel* and *sim-sech* comparisons because they contained fewer divergent sites than the *mel-sim* comparison, making allele assignments more challenging. For the *mel-mel* comparison, we analyzed new F_1 hybrid and mixed parental cDNA libraries synthesized from the same RNA samples used for Illumina sequencing, incorporating variation from technical replication. These

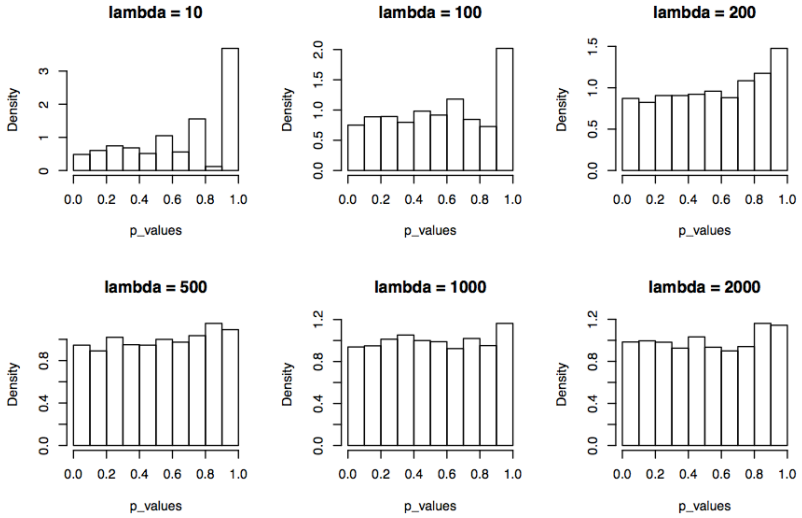
mixed parental libraries were constructed by pooling equal amounts of RNA prior to cDNA synthesis. For the *sim-sech* comparison, we used RNA extracted from 4 biological replicate mixed parental (each containing 10 *D. simulans* and 10 *D. sechellia* flies) and F_1 hybrid (each containing 20 F_1 hybrid flies) samples to synthesize cDNA pools, incorporating variation from both technical and biological replication.

Pyrosequencing assays were developed for 10 genes in the *mel-mel* comparison and 18 genes in the *sim-sech* comparison (Table S6). Custom dispensation orders were used for the pyrosequencing reactions and custom formulas were developed for calculating relative allelic abundance (Table S6). Both gDNA and cDNA were analyzed in mixed parental and F_1 hybrid samples in each case. cDNA was always synthesized from total RNA using T(18)VN primers and SuperScript II (Invitrogen) according to manufacturer recommendations. gDNA was extracted from an independent pool of F_1 flies for *mel-mel* and sequentially from the same homogenate of flies as the RNA for each biological replicate of *sim-sech* using the protocol described in Wittkopp (2011). For *mel-mel*, pyrosequencing reactions were performed in triplicate for both the cDNA and gDNA samples. For *sim-sech*, single pyrosequencing reactions were performed on the cDNA and gDNA samples from each biological replicate.

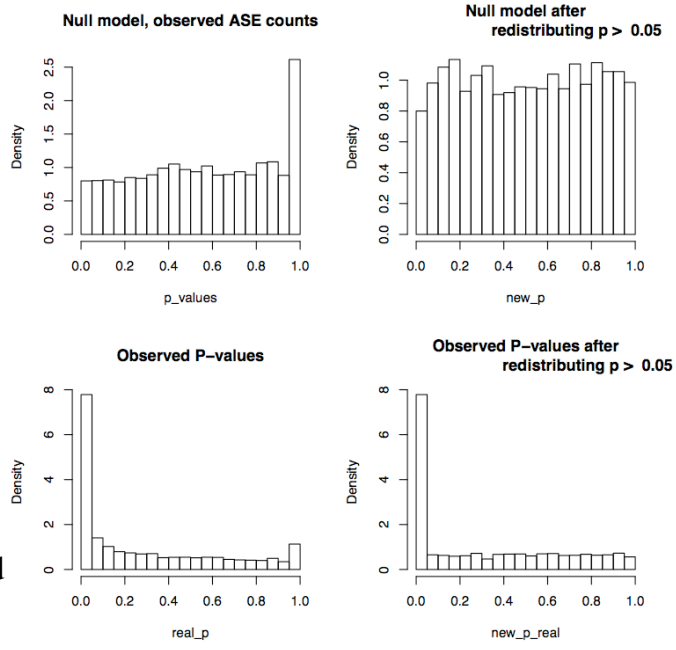
Relative allelic abundance observed in the gDNA samples was used to normalize measurements from the corresponding cDNA samples, as described in (Wittkopp 2011). Following normalization, the mean $\log_2$-transformed ratio of relative allelic expression reported by pyrosequencing for each gene was compared to the $\log_2$-transformed ratio of relative allelic expression derived from the RNA-seq data using a type 2 regression in R.

SUPPLEMENTAL NOTE

When using binomial exact tests to identify significant differences in relative allelic expression from RNA-seq data, p-values are not uniformly distributed when the null hypothesis (p = 0.5) is true, violating an assumption of the widely-used false discovery rate (FDR) correction for multiple testing (Skelly et al. 2011). To better understand the non-uniformity of p-values when the null hypothesis (p = 0.05) is true, we repeated and extended the simulations reported in Skelly et al. (2011). We followed the same simulation strategy except that we did not add a noise parameter. Specifically, we simulated allele-specific read count data by (i) sampling total allele-specific read counts (allele 1 + allele 2) for 4851 genes (the number of genes we examined for allele-specific expression) from a Poisson distribution with a mean (lambda) of 10, 100, 200, 500, 1000, or 2000 reads; (ii) determining the number of reads from one allele for each gene by drawing from a binomial distribution with p = 0.5, and (iii) determining the p-value from a binomial exact test comparing the reads from one allele to the total number of allele-specific reads in each case. As shown in the figure at right, the uniformity of the null p-value distribution increased with increasing total read counts.

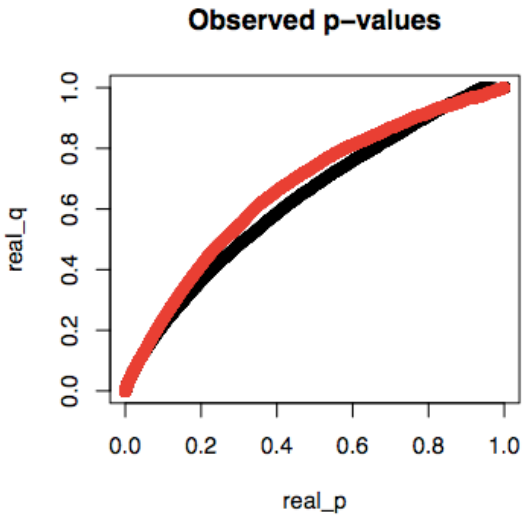


In our Dataset 2, the total number of read counts per gene ranged from 20 to 16165, with a median of 142. We repeated the simulation described above using the observed number of total allele-specific read counts for the 4851 genes and found that the distribution of p-values assuming the null hypothesis was true did indeed violate uniformity. This is important because the FDR correction uses the assumption of a uniform distribution when the null hypothesis is true to model the FDR based on the observed distribution of p-values. To determine the impact of this non-uniform distribution on the proportion of genes called significant using the q-value = 0.05 cutoff we used, we artificially eliminated the over-



representation of large, non-significant p-values from the null model in the observed p-value distribution by replacing all p-values > 0.05 with an equal number of values drawn from a uniform distribution (min = 0.05, max = 1). We did this for the dataset simulated assuming the null hypothesis was true (top panels) as well as for the p-values observed in the *mel-mel* comparison (bottom panels).

Next, we applied the FDR correction to the list of p-values from each dataset. As shown in the figure at right for the observed p-values, q-values before (black) and after (red) the redistribution of non-significant p-values are quite similar. The largest difference in q-value was for p-values between 0.2 and 0.8. With the q-value threshold (0.05) used in the paper, as well as for other q-value thresholds up to 0.1, the number of genes called significant using the FDR correction was unchanged by the redistribution of non-significant p-values (see Table below). This suggests that the non-uniformity of the BET null distribution has a negligible effect on the FDR in our study. Furthermore, even if the FDR determined by the correction model was slightly higher or lower than the true FDR, it is expected to deviate similarly in all comparisons because the same number of total allele-specific read counts were considered in each case, resulting in the same null distribution. Based on these results, we conclude that the non-uniformity of the BET null distribution does not affect the comparisons of regulatory evolution across divergence times that we report.



q-value:	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7
non-uniform null	1081	1476	1756	2162	2508	2767	3071	3347	3651
nearly uniform null	1081	1476	1756	1987	2153	2311	2541	2748	3094

SUPPLEMENTAL REFERENCES

Coolon JD, Stevenson KR, McManus CJ, Graveley BR, Wittkopp PJ. 2012. Genomic imprinting absent in *Drosophila melanogaster* adult females. *Cell Rep* **2**: 69–75.

Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, Pritchard JK. 2009. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* **25**: 3207–3212.

Drosophila 12 Genomes Consortium 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**: 203–218.

Harris R. 2007. Improved Pairwise Alignment of Genomic DNA. Thesis. The Pennsylvania State University.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996–1006.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM. 2011. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res* **21**: 1728–1737.

Stevenson KR, Coolon JD, Wittkopp PJ. 2013. Sources of bias in measures of allele-specific expression derived from RNA-seq data aligned to a single reference genome. *BMC Genomics* **14**: 536.

Wittkopp PJ. 2011. Using pyrosequencing to measure allele-specific mRNA abundance and infer the effects of *cis*- and *trans*-regulatory differences. *Methods Mol Biol* **772**: 297–317.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821–829.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1: Methodological overview. This figure summarizes the biological samples analyzed (blue circles), the production of the raw sequencing data (green boxes), the bioinformatic methods (red boxes) used to convert the raw data into datasets 1 and 2 (blue ovals), and analyses performed on each of these datasets to examine patterns of regulatory evolution (blue boxes).

Figure S2: Independent confirmation of relative allelic expression levels inferred from RNA-seq data. Measures of total expression (**A, B**) and relative allelic expression (**C**) derived from RNA-seq read counts in Dataset 1 (**A**) and Dataset 2 (**B**, **C**) were compared to measures of total expression (**A**, **B**) and relative allelic expression (**C**) determined using pyrosequencing. For the *mel-mel* comparisons (red), cDNA samples analyzed by pyrosequencing were derived from the same RNAs used for Illumina sequencing (i.e., technical replicates). For the *sim-sech* comparison (blue), RNA samples extracted from new pools of flies (i.e., biological replicates) were analyzed by pyrosequencing. Coefficients of determination (R^2) from type 2 regressions were used to compare expression measurements based on RNA-seq and pyrosequencing.

Figure S3: Total expression levels were similar between reciprocal hybrids. Total expression levels, plotted as $\log_2$(total read count) for each gene, were compared between F_1 hybrids from reciprocal crosses for the *mel-mel* (**A**), *sim-sech* (**B**), and *mel-sim* (**C**) comparisons. Hybrid genotypes are written as maternal genotype x paternal genotype. Spearman's ρ, which makes no assumptions about normality, was used to compare the strength of the correlation in each case.

Figure S4: Most significant expression differences between reciprocal hybrids are small in magnitude. Volcano plots are shown for the comparison of total expression between reciprocal hybrids in the *mel-mel* (**A**), *sim-sech* (**B**), and *mel-sim* (**C**) comparisons. Statistical significance, represented by $\log_{10}$(q-value), is plotted on the Y-axis and the expression difference, plotted as $\log_2$(reads from hybrid 1/reads from in hybrid 2), is plotted on the X-axis. Hybrid genotypes are written as maternal genotype x paternal genotype. The vertical red lines correspond a 1.25-fold expression difference, whereas the horizontal red lines correspond to q-values with a false discovery rate of 0.05. Insets show genes with the smallest q-values in more detail.

Figure S5 : Overall expression differences increase with divergence time. For each gene, expression levels of individuals genes are compared between the *zhr* and *z30* strains of *D. melanogaster* (**A**), between *D. simulans* and *D. sechellia* (**B**), and between *D. melanogaster* (*zhr*) and *D. simulans* (**C**). Expression levels are plotted as $\log_2$(read count). Spearman's ρ, which makes no assumptions about normality, was used to measure the overall expression similarity between genotypes in each case.

Figure S6: Many small expression differences are statistically significant between genotypes. Volcano plots are shown for the comparison of total expression between the *zhr* and *z30* strains of *D. melanogaster* (**A**), between *D. simulans* and *D. sechellia* (**B**), and between *D. melanogaster* (*zhr*) and *D. simulans* (**C**). Statistical significance, represented by $\log_{10}$(q-value), is

plotted on the Y-axis, and the expression difference, plotted as $\log_2$(reads from genotype 1/reads from in genotype 2), is plotted on the X-axis. The vertical red lines correspond to a 1.25-fold expression difference, whereas the horizontal red lines correspond to q-values with a false discovery rate of 0.05. Insets show genes with the smallest q-values in more detail.

Figure S7: Expression differences between F_1 hybrids and parental species increase with divergence time. For each gene, expression levels, plotted as $\log_2$(read counts), are compared between F_1 hybrids (Y-axis) and each of the parental species (X-axis) for the *mel-mel* (A-D), *sim-sech* (E-H) and *mel-sim* (I-L) comparisons. Hybrid genotypes are written as maternal genotype x paternal genotype. Spearman's ρ, which makes no assumptions about normality, was used to compare overall expression differences in each case.

Figure S8: The proportion of genes showing misexpression in F_1 hybrids increased with divergence time of the parental genotypes. Differences in total expression between the F_1 hybrid and each parental species are compared in each panel. The difference in expression level is plotted for each gene as $\log_2$(reads from hybrid) - $\log_2$(reads from parental genotype), with the difference from parent 1 shown on the X-axis and the difference from parent 2 shown on the Y-axis. Hybrid genotypes are written as maternal genotype x paternal genotype. The Y=X line shown indicates an equal difference between expression in the F_1 hybrid and both parents. Each gene was categorized as showing either conserved (light blue), additive (orange), underdominant (red), overdominant (blue), or dominant with expression similar to parent 1 (purple) or parent 2 (green), as described in the Methods. The pie-chart insets show the proportion of genes in each class. Interestingly, in the *mel-mel* comparison (**A**, **D**), dominant expression patterns resembled the North American *zhr* strain (parent 1) more than twice as often as they resembled the African *z30* strain (parent 2). In the *sim-sec* and *mel-sim* F_1 hybrids, dominant regulatory alleles were distributed more evenly between the two parental genotypes (**B**,**E** and **C**,**F,** respectively).

Figure S9: The frequency of large expression differences between F_1 hybrids and parental species increases with divergence time. Volcano plots are shown for the comparison of expression levels between F_1 hybrids and each parent in the *mel-mel* (**A-D**), *sim-sech* (**E-H**), and *mel-sim* (**I-L**) comparisons. Statistical significance, represented by $\log_{10}$(q-value), is plotted on the Y-axis and the expression difference, plotted as $\log_2$(reads from parental genotype/reads from in hybrid genotype), is plotted on the X-axis. Hybrid genotypes are written as maternal genotype x paternal genotype. The vertical red lines correspond to a 1.25-fold expression difference, whereas the horizontal red lines correspond to q-values with a false discovery rate of 0.05. Insets show genes with the smallest q-values in more detail.

Figure S10: Allele-specific sequence reads are accurately assigned to genotypes. Relative expression levels inferred from *in silico* "mixed parental" samples after computational assignment of reads to specific alleles (Y-axis) were compared to relative expression levels determined using the separately sequenced samples (X-axis) for the *mel-mel* (**A, D, G**), *sim-sech* (**B, E, H**), and *mel-sim* (**C, F, I**) comparisons. In panels A-C, both allele-specific and shared reads from the mixed parental samples were included. In panels D-F, only allele-specific reads

from the mixed parental samples were included. In panels G-I, the allele-specific read counts in Dataset 2 (i.e., after using hypergeometric sampling to equalize power among comparisons) were used. Relative expression is plotted as $\log_2$(reads from genotype 1/reads from genotype 2) in all cases. Coefficients of determination (R^2) from linear models were used to compare relative expression between true values and those determined for *in silico* mixed samples.

Figure S11: Relative allelic expression was similar between reciprocal hybrids. Measures of relative allelic expression were compared for each gene between reciprocal hybrids for the *mel-mel* (**A**), *sim-sech* (**B**), and *mel-sim* (**C**) comparisons. Coefficients of determination (R^2) from linear models were used to compare relative allelic expression between reciprocal hybrids. The nine genes identified as having a statistically significant difference in relative allelic expression between reciprocal hybrids are shown in red. Log_2(reads from allele 1/reads from allele 2) is plotted for each hybrid genotype. **D-I**, Allelic expression levels, plotted as $\log_2$(allele-specific read counts) for each gene, are compared for each allele in reciprocal hybrids for all three comparisons. Spearman's ρ, which makes no assumptions about normality, was used to compare overall expression differences in each case. Genotypes of reciprocal hybrids are written as maternal genotype x paternal genotype.

Figure S12: Evolution of *cis*- and *trans*-regulation. A, Overall differences ($1 - \rho$) in total expression between genotypes (blue) and allele-specific expression in F_1 hybrids (red) are shown for each comparison. Relative allelic expression provides a readout of relative *cis*-regulatory activity. Data used to calculate Spearman's ρ are summarized in Figure S13. **B**, For each comparison, the proportion of genes with evidence of significant differences in total expression (blue), *cis*-regulation (red), and *trans*-regulation (green) are shown. Data used to determine these proportions is shown in Figure S16. **C**, The proportion of genes with evidence of significant *cis*- and *trans*-regulatory changes (red) is compared to the proportion of genes with evidence of *cis*- or *trans*-regulatory changes (blue). **D**, For genes with evidence of both *cis*- and *trans*-regulatory changes, the frequency of genes with *cis*- and *trans*-regulatory changes affecting gene expression in the same ("*cis*+*trans*", red) and opposite ("*cisXtrans*", blue) directions are compared. **E**, The relative frequencies of genes with *cis*- and *trans*-regulatory changes in opposite directions that do (blue) and do not (red) show evidence of misexpression in F_1 hybrids are compared. Error bars in panel **A** show the 95% quantiles from 10,000 bootstrap replicates. Comparable analyses for reciprocal hybrids are shown in Figure 3.

Figure S13: Differences in *cis*-regulatory activity increase with divergence time more rapidly than differences in total expression. For each of the 4851 genes in Dataset 2, total expression levels, plotted as $\log_2$(read count), are compared between the *zhr* and *z30* strains of *D. melanogaster* (**A**), between *D. simulans* and *D. sechellia* (**B**), and between *D. melanogaster* (*zhr*) and *D. simulans* (**C**). Levels of allele-specific expression, plotted as $\log_2$(allele-specific read count), are compared between allele 1 (X-axis) and allele 2 (Y-axis) for each F_1 hybrid from the *mel-mel* (**D**,**G**), *sim-sech* (**E**,**H**) and *mel-sim* (**F**,**I**) comparisons. Hybrid genotypes are written as maternal genotype x paternal genotype. Spearman's ρ, which makes no assumptions about normality, was used to measure the overall expression similarity between genotypes in each case.

Figure S14: Differences between gene sets used to analyze total and allele-specific expression data. Box-plots summarize the distributions of total expression differences (**A**) and sequence divergence (**B**) for the 7587 genes in Dataset 1 and the 4851 genes in Dataset 2 for the *mel-mel* (MM), *sim-sech* (SS), and *mel-sim* (MS) comparisons. The notched box-plots show the full range of values as well as the 25, 50, and 75th percentiles.

Figure S15: Evolutionary trajectories for expression divergence of individual genes. The line-plots show differences in total expression (**A**) and *cis*-regulatory activity (**B**) derived from Dataset 2 for individual genes in the *mel-mel*, *sim-sech*, and *mel-sim* comparisons plotted according to divergence time. As described in the main text, genes were classified into nine groups depending on whether they showed increased, decreased, or similar allele-specific expression differences between *mel-mel* and *sim-sech* and between *sim-sech* and *mel-sim*. The red line in each plot shows the median difference in *cis*-regulatory activity for genes in that class for each comparison. The pie-chart shows the relative frequency of genes in each class.

Figure S16: The relative contributions of *cis*- and *trans*-regulatory changes to expression divergence change with divergence time. For each gene, relative expression between parental genotypes, plotted as $\log_2$(reads from parent 1/ reads from parent 2) on the X-axis, is compared to relative allele-specific expression in F_1 hybrids, plotted as $\log_2$(reads from allele 1/ reads from allele 2) on the Y-axis. Hybrid genotypes are written as maternal genotype x paternal genotype. Each gene was categorized as showing either conserved *cis*- and *trans*-regulation (yellow, "conserved"), only *cis*-regulatory differences (black, "all *cis*"), only *trans*-regulatory differences (red, "all *trans*"), *cis*- and *trans*-regulatory differences with no expression difference between parental genotypes (orange, "compensatory"), or *cis*- and *trans*-regulatory differences with effects on expression in the same (blue, "*cis* + *trans*") or opposite (green, "*cis* X *trans*") directions, as described in the Methods and Table S5. Genes with results from significance tests inconsistent with any of these categories (see Methods) were classified as "ambiguous" (grey). The pie-chart insets show the proportion of genes in each class, for each comparison. The slopes reported in each panel are from a linear regression model fitted to the corresponding dataset. A larger slope indicates a larger contribution of *cis*-regulatory divergence to total expression divergence.

Figure S17: Effects of *cis*-regulatory divergence. A, The percentage of total regulatory divergence attributable to *cis*-regulatory divergence (%*cis*) is shown for the *mel-mel*, *sim-sech*, and *mel-sim* comparisons. **B-D**, %*cis* is compared among sets of genes with different levels of total expression differences, reported as the absolute value of the $\log_2$(reads from genotype 1/ reads from genotype 2) ratio, for the *mel-mel* (**B**), *sim-sech* (**C**), and *mel-sim* (**D**), comparisons. **E**, %*cis* is compared for sets of genes showing additive ("A") and non-additive ("NA", dominant or misexpression) inheritance for each comparison. In all panels, notched box-plots show the full range of values as well as the 25, 50, and 75th percentiles, with the width of the box-plots proportion to the number of genes in each class. Analyses comparable to panels **A** and **H** using reciprocal hybrids are shown in Figure 4.

SUPPLEMENTAL TABLES

Table S1: Summary of sequencing depth for RNA-seq and gDNA.

Table S2: Number of genes suitable for quantifying total expression in each genotype.

Table S3: Number of genes suitable for quantifying allele-specific expression in each genotype.

Table S4: Accuracy of mapping maternally inherited mitochondrial alleles in interspecific F_1 hybrids.

Table S5: Criteria for assigning genes to regulatory evolution classes.

Table S6: Pyrosequencing assays for quantification of allelic expression ratios.

Figure S1

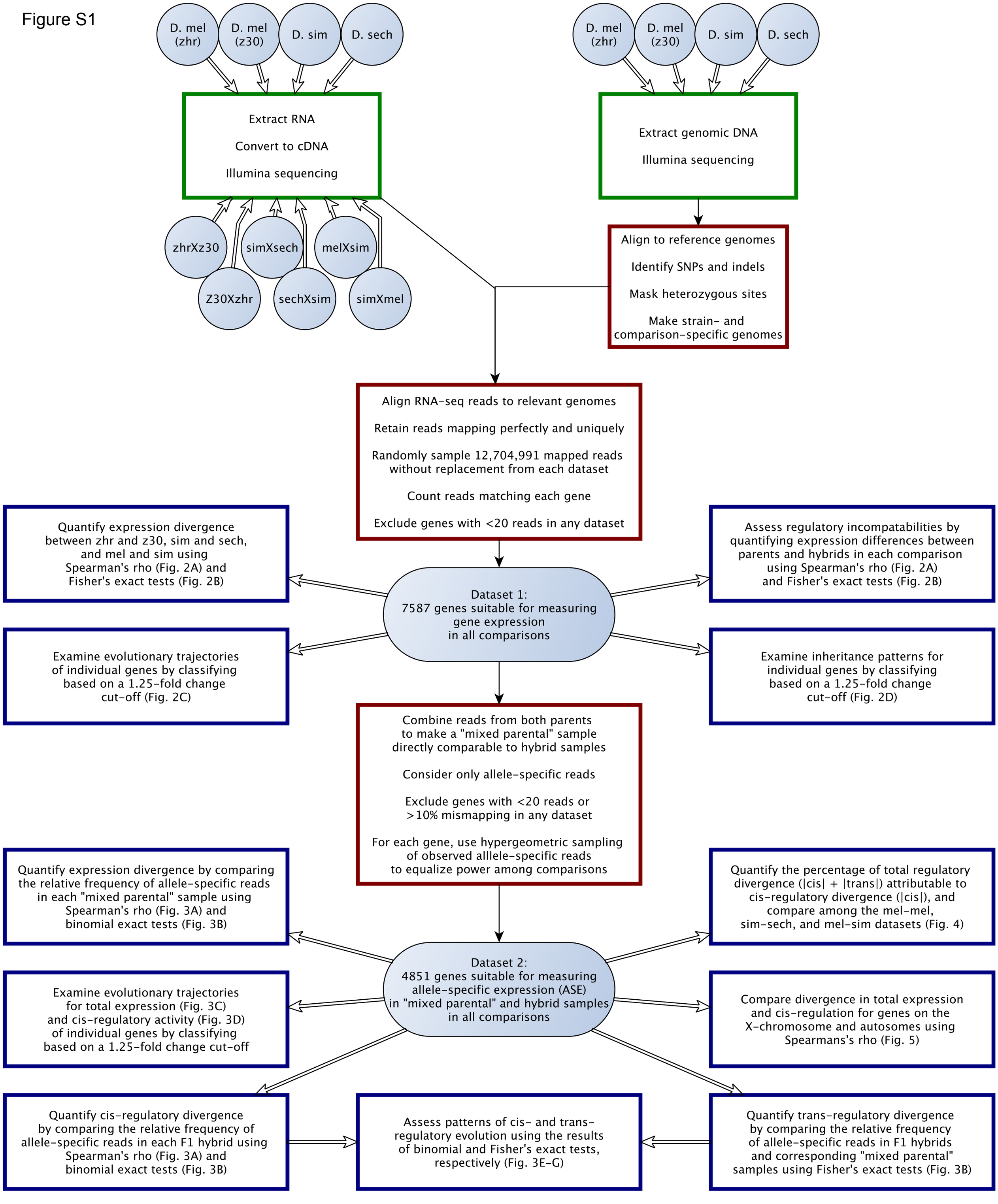

D. mel (zhr)
D. mel (z30)
D. sim
D. sech
Extract RNA
Convert to cDNA
Illumina sequencing
zhrXz30
simXsech
melXsim
Z30Xzhr
sechXsim
simXmel
D. mel (zhr)
D. mel (z30)
D. sim
D. sech
Extract genomic DNA
Illumina sequencing
Align to reference genomes
Identify SNPs and indels
Mask heterozygous sites
Make strain- and comparison-specific genomes
Align RNA-seq reads to relevant genomes
Retain reads mapping perfectly and uniquely
Randomly sample 12,704,991 mapped reads without replacement from each dataset
Count reads matching each gene
Exclude genes with <20 reads in any dataset
Quantify expression divergence between zhr and z30, sim and sech, and mel and sim using Spearman's rho (Fig. 2A) and Fisher's exact tests (Fig. 2B)
Assess regulatory incompatabilities by quantifying expression differences between parents and hybrids in each comparison using Spearman's rho (Fig. 2A) and Fisher's exact tests (Fig. 2B)
Dataset 1: 7587 genes suitable for measuring gene expression in all comparisons
Examine evolutionary trajectories of individual genes by classifying based on a 1.25-fold change cut-off (Fig. 2C)
Examine inheritance patterns for individual genes by classifying based on a 1.25-fold change cut-off (Fig. 2D)
Combine reads from both parents to make a "mixed parental" sample directly comparable to hybrid samples
Consider only allele-specific reads
Exclude genes with <20 reads or >10% mismapping in any dataset
For each gene, use hypergeometric sampling of observed alllele-specific reads to equalize power among comparisons
Quantify expression divergence by comparing the relative frequency of allele-specific reads in each "mixed parental" sample using Spearman's rho (Fig. 3A) and binomial exact tests (Fig. 3B)
Quantify the percentage of total regulatory divergence (|cis| + |trans|) attributable to cis-regulatory divergence (|cis|), and compare among the mel-mel, sim-sech, and mel-sim datasets (Fig. 4)
Dataset 2: 4851 genes suitable for measuring allele-specific expression (ASE) in "mixed parental" and hybrid samples in all comparisons
Examine evolutionary trajectories for total expression (Fig. 3C) and cis-regulatory activity (Fig. 3D) of individual genes by classifying based on a 1.25-fold change cut-off
Compare divergence in total expression and cis-regulation for genes on the X-chromosome and autosomes using Spearmans's rho (Fig. 5)
Quantify cis-regulatory divergence by comparing the relative frequency of allele-specific reads in each F1 hybrid using Spearman's rho (Fig. 3A) and binomial exact tests (Fig. 3B)
Assess patterns of cis- and trans-regulatory evolution using the results of binomial and Fisher's exact tests, respectively (Fig. 3E-G)
Quantify trans-regulatory divergence by comparing the relative frequency of allele-specific reads in F1 hybrids and corresponding "mixed parental" samples using Fisher's exact tests (Fig. 3B)

Figure S2

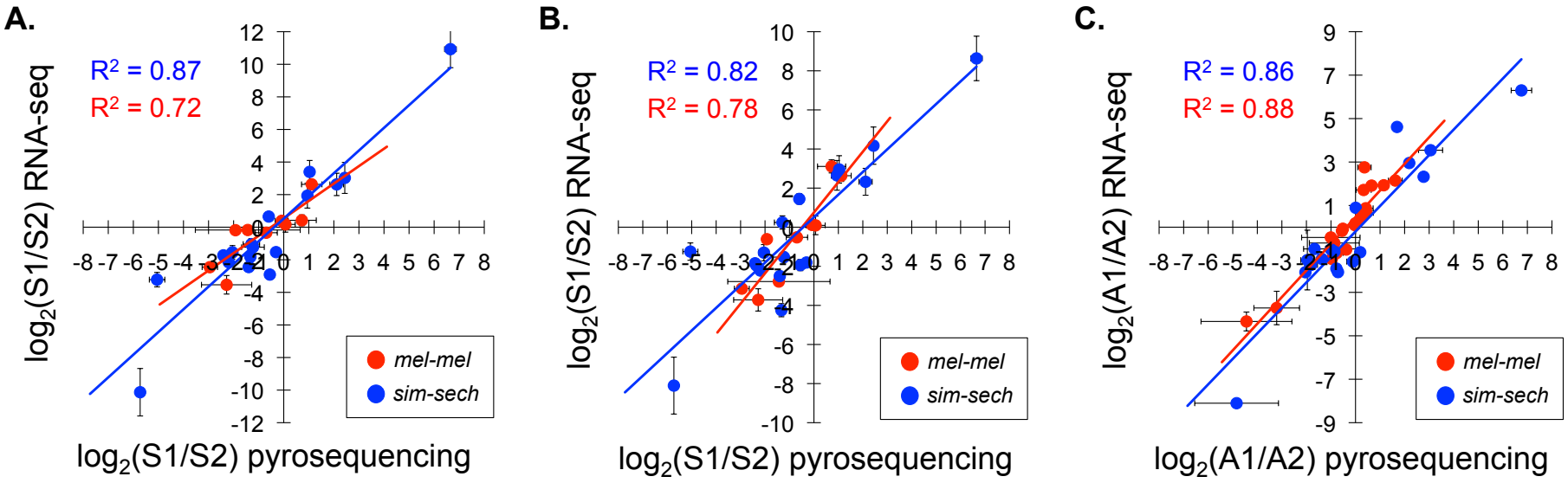

A.
R² = 0.87
R² = 0.72
log2(S1/S2) RNA-seq
log2(S1/S2) pyrosequencing
mel-mel
sim-sech
B.
R² = 0.82
R² = 0.78
log2(S1/S2) RNA-seq
log2(S1/S2) pyrosequencing
mel-mel
sim-sech
C.
R² = 0.86
R² = 0.88
log2(A1/A2) RNA-seq
log2(A1/A2) pyrosequencing
mel-mel
sim-sech

Figure S3

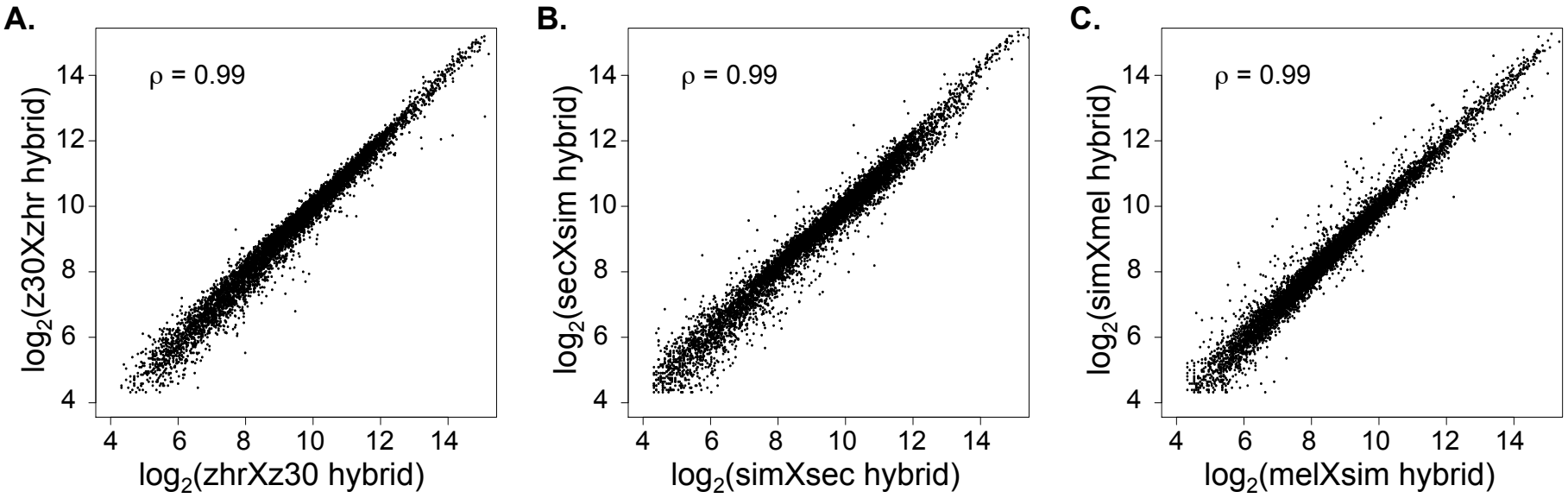

A.
ρ = 0.99
log2(z30Xzhr hybrid)
log2(zhrXz30 hybrid)
B.
ρ = 0.99
log2(secXsim hybrid)
log2(simXsec hybrid)
C.
ρ = 0.99
log2(simXmel hybrid)
log2(melXsim hybrid)
4
6
8
10
12
14

Figure S4

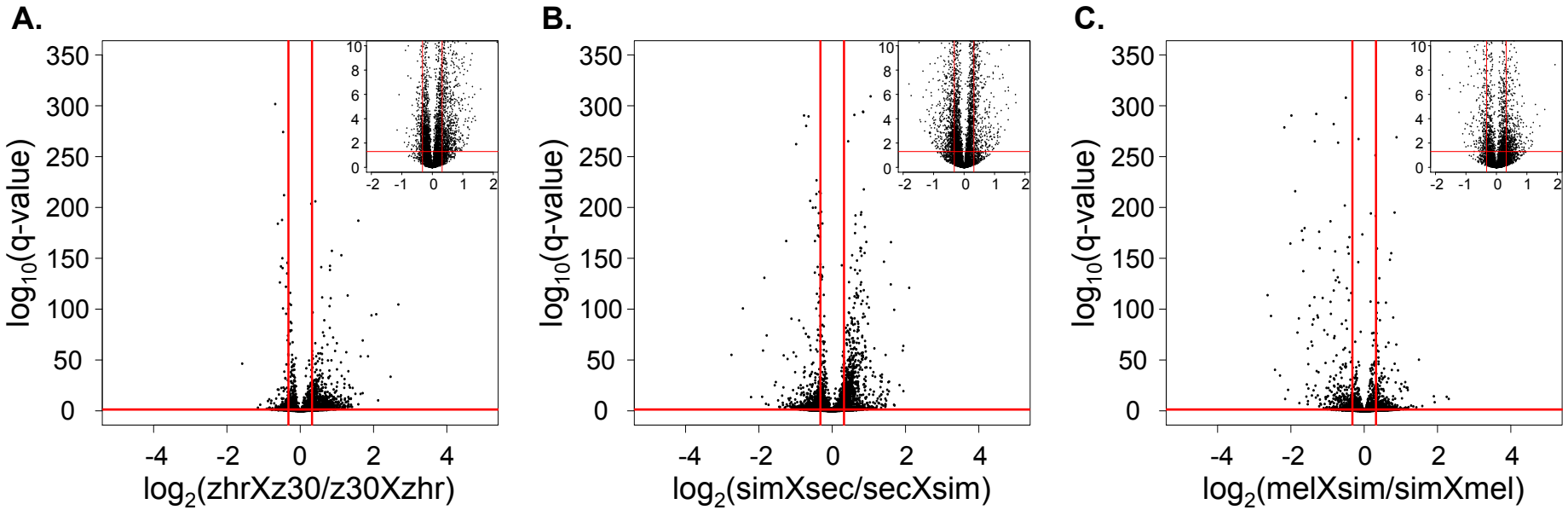

A.
B.
C.
log10(q-value)
log2(zhrXz30/z30Xzhr)
log2(simXsec/secXsim)
log2(melXsim/simXmel)

Figure S5

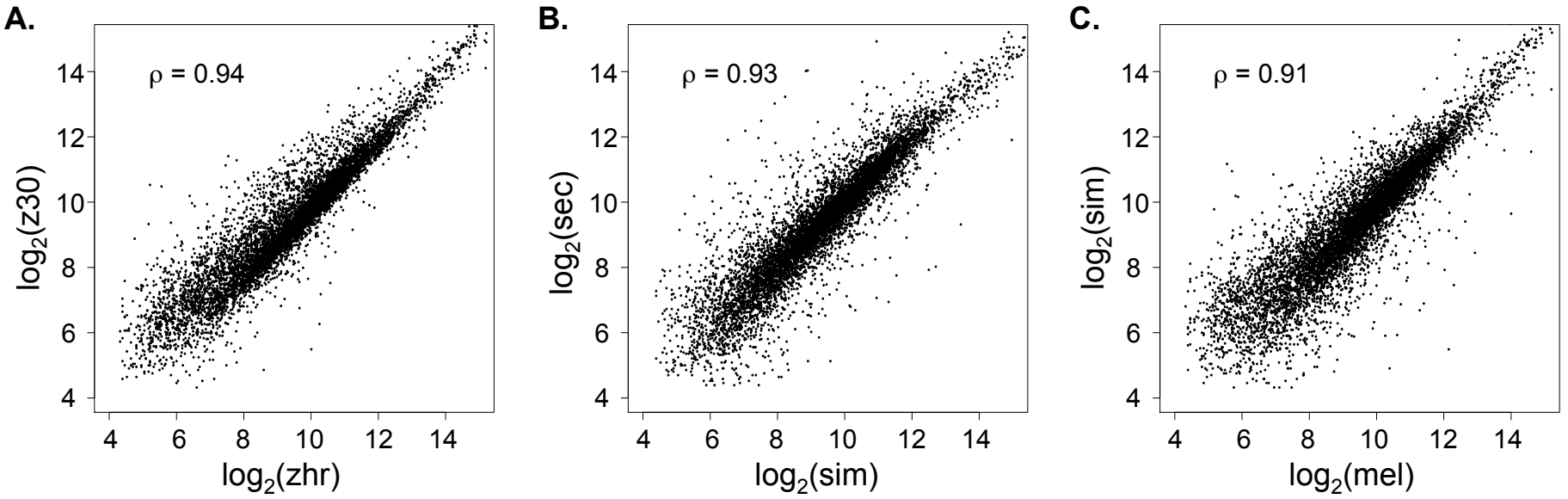

A.
ρ = 0.94
log2(z30)
log2(zhr)
B.
ρ = 0.93
log2(sec)
log2(sim)
C.
ρ = 0.91
log2(sim)
log2(mel)
4
6
8
10
12
14

Figure S6

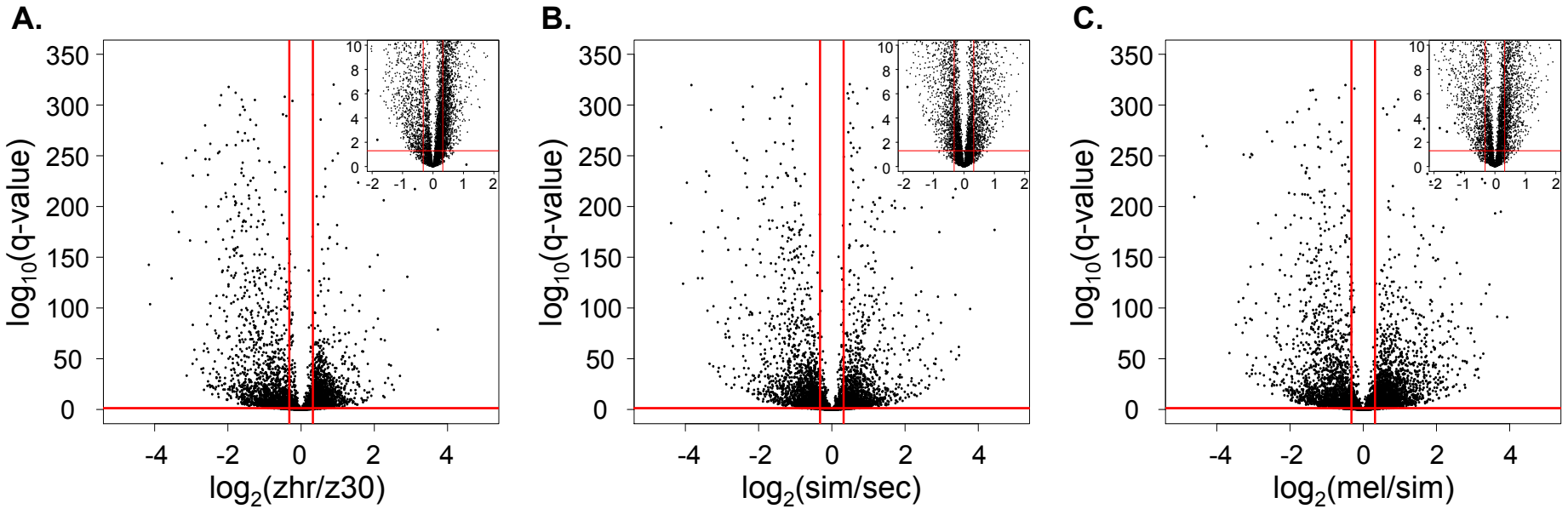

A.
B.
C.
log10(q-value)
log2(zhr/z30)
log2(sim/sec)
log2(mel/sim)
350
300
250
200
150
100
50
0
-4
-2
0
2
4
10
8
6
4
2
0
-2
-1
0
1
2

Figure S7

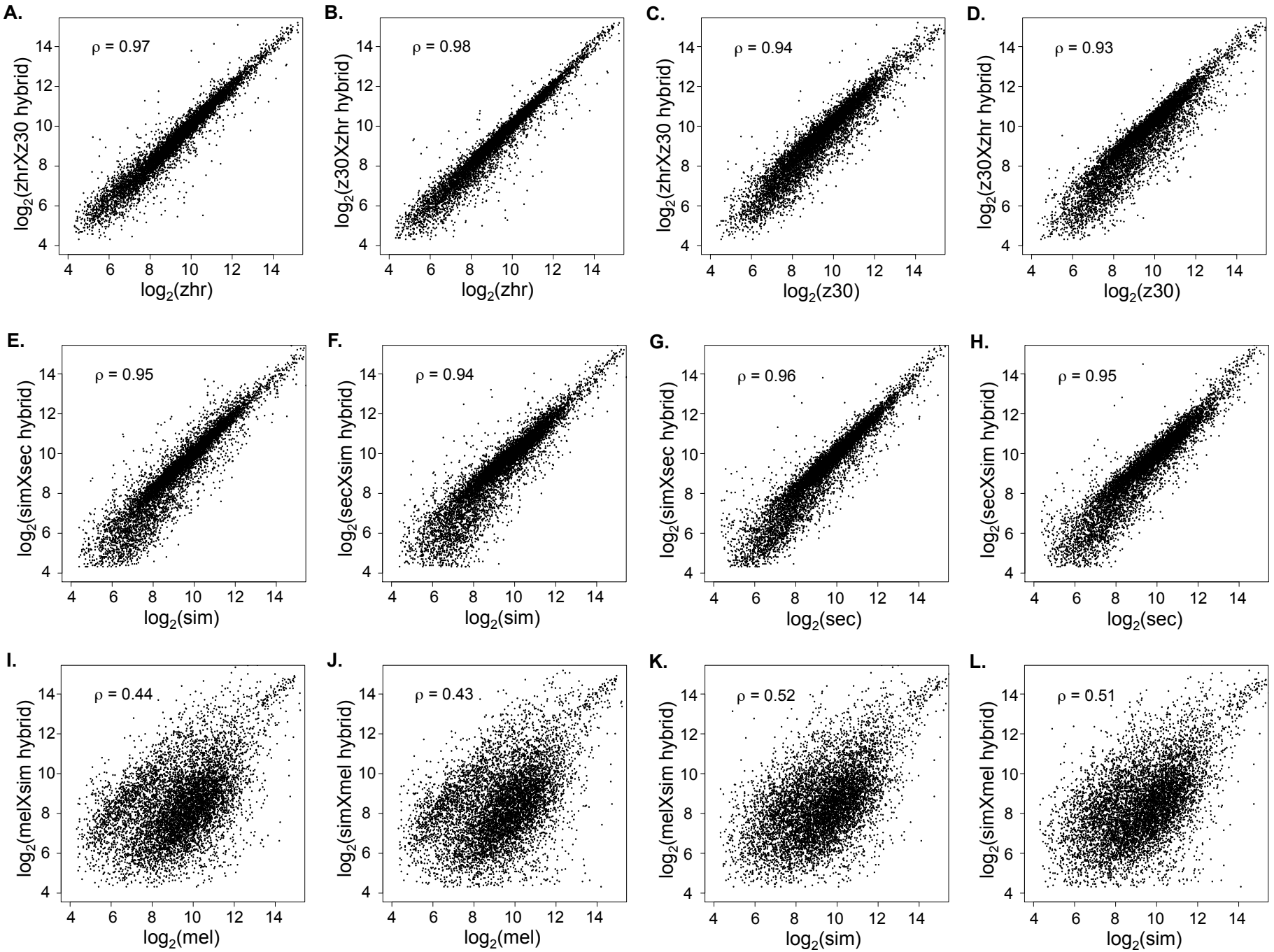

A.
ρ = 0.97
log2(zhrXz30 hybrid)
log2(zhr)
B.
ρ = 0.98
log2(z30Xzhr hybrid)
log2(zhr)
C.
ρ = 0.94
log2(zhrXz30 hybrid)
log2(z30)
D.
ρ = 0.93
log2(z30Xzhr hybrid)
log2(z30)
E.
ρ = 0.95
log2(simXsec hybrid)
log2(sim)
F.
ρ = 0.94
log2(secXsim hybrid)
log2(sim)
G.
ρ = 0.96
log2(simXsec hybrid)
log2(sec)
H.
ρ = 0.95
log2(secXsim hybrid)
log2(sec)
I.
ρ = 0.44
log2(melXsim hybrid)
log2(mel)
J.
ρ = 0.43
log2(simXmel hybrid)
log2(mel)
K.
ρ = 0.52
log2(melXsim hybrid)
log2(sim)
L.
ρ = 0.51
log2(simXmel hybrid)
log2(sim)

Figure S8

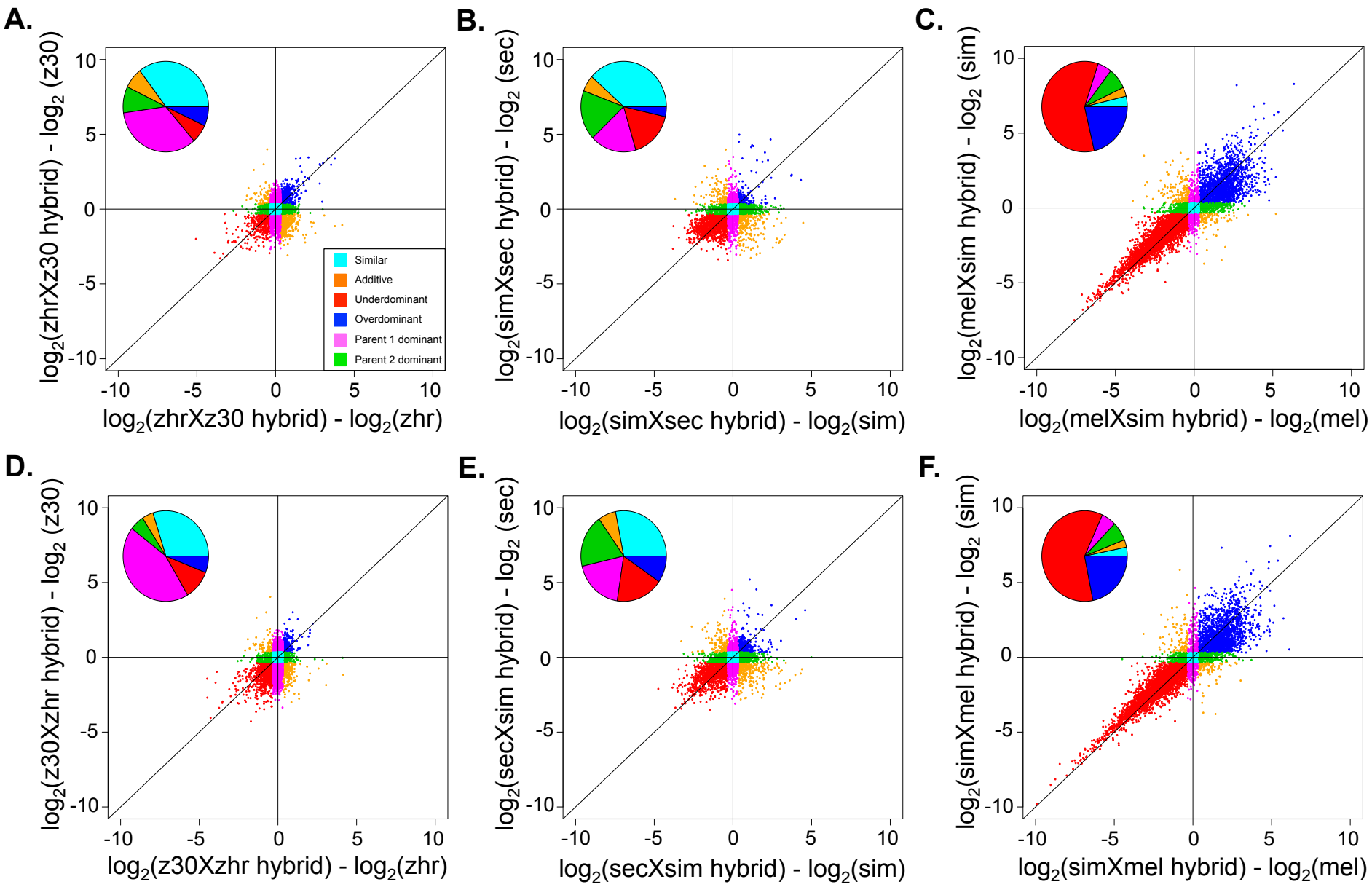

A.
B.
C.
D.
E.
F.
log2(zhrXz30 hybrid) - log2 (z30)
log2(zhrXz30 hybrid) - log2(zhr)
log2(simXsec hybrid) - log2 (sec)
log2(simXsec hybrid) - log2(sim)
log2(melXsim hybrid) - log2 (sim)
log2(melXsim hybrid) - log2(mel)
log2(z30Xzhr hybrid) - log2 (z30)
log2(z30Xzhr hybrid) - log2(zhr)
log2(secXsim hybrid) - log2 (sec)
log2(secXsim hybrid) - log2(sim)
log2(simXmel hybrid) - log2 (sim)
log2(simXmel hybrid) - log2(mel)
Similar
Additive
Underdominant
Overdominant
Parent 1 dominant
Parent 2 dominant

Figure S9

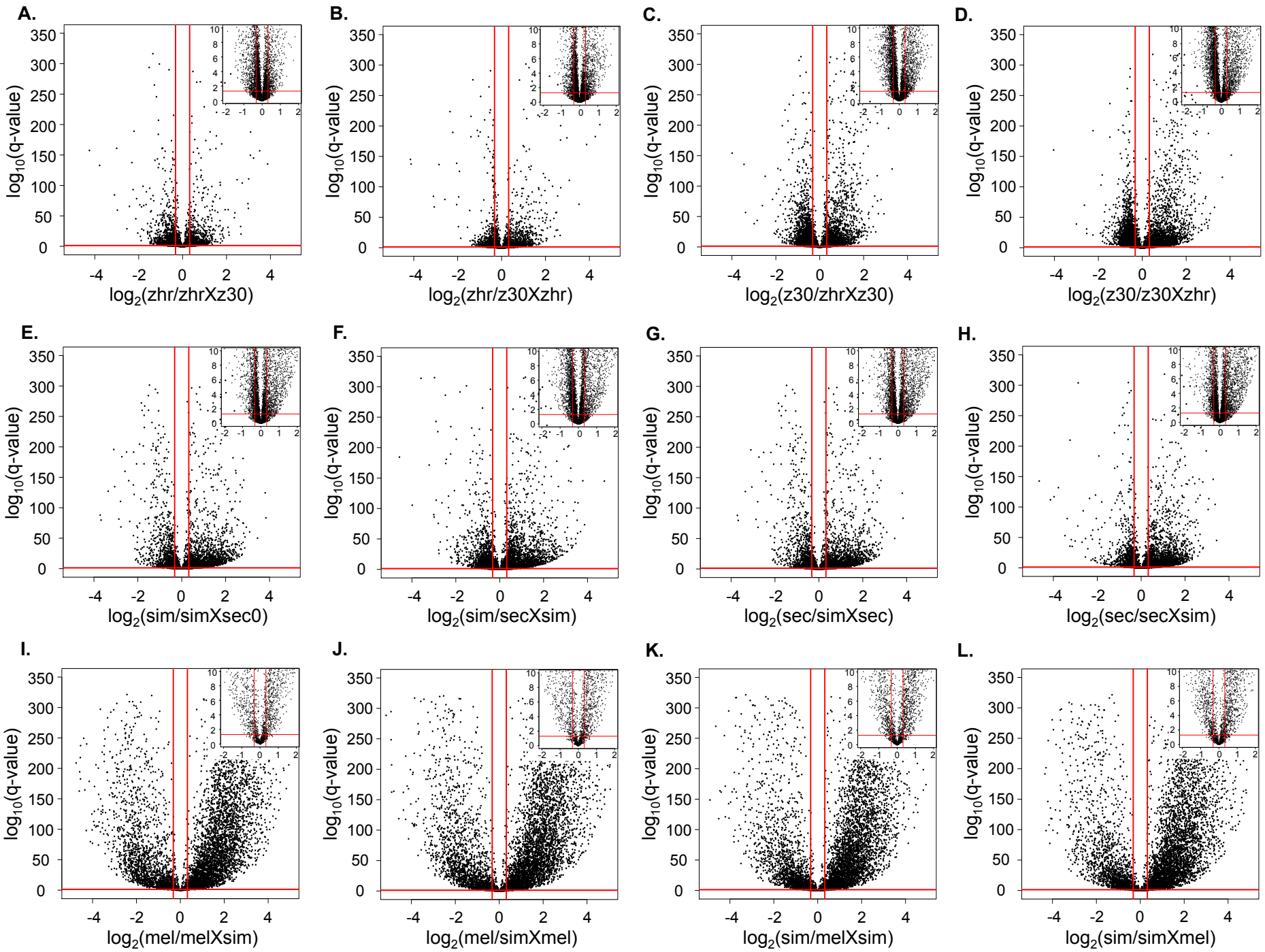

A.
B.
C.
D.
E.
F.
G.
H.
I.
J.
K.
L.
log10(q-value)
log2(zhr/zhrXz30)
log2(zhr/z30Xzhr)
log2(z30/zhrXz30)
log2(z30/z30Xzhr)
log2(sim/simXsec0)
log2(sim/secXsim)
log2(sec/simXsec)
log2(sec/secXsim)
log2(mel/melXsim)
log2(mel/simXmel)
log2(sim/melXsim)
log2(sim/simXmel)

Figure S10

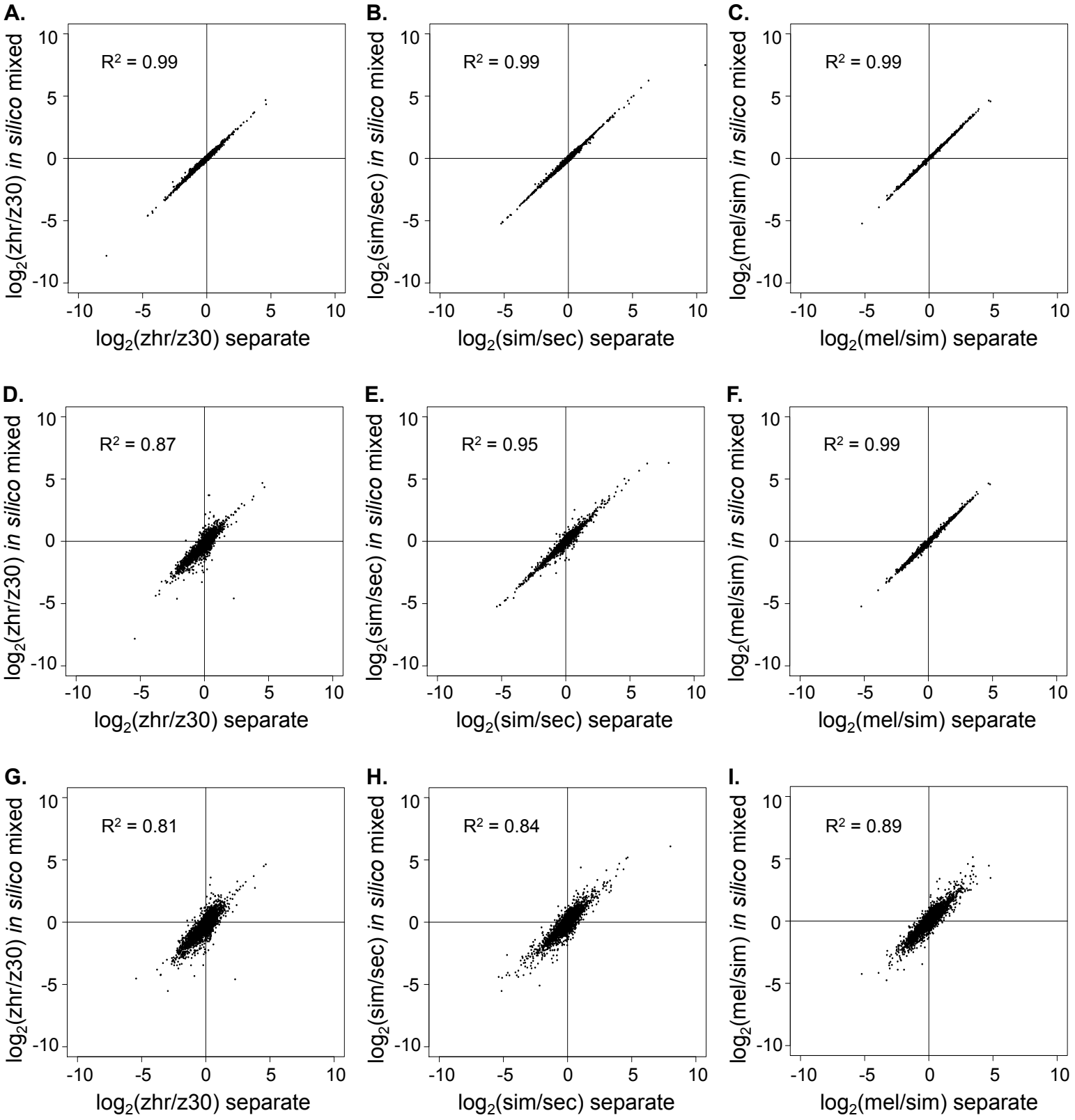

A.
B.
C.
D.
E.
F.
G.
H.
I.
R² = 0.99
R² = 0.99
R² = 0.99
R² = 0.87
R² = 0.95
R² = 0.99
R² = 0.81
R² = 0.84
R² = 0.89
log2(zhr/z30) in silico mixed
log2(sim/sec) in silico mixed
log2(mel/sim) in silico mixed
log2(zhr/z30) separate
log2(sim/sec) separate
log2(mel/sim) separate
10
5
0
-5
-10

Figure S11

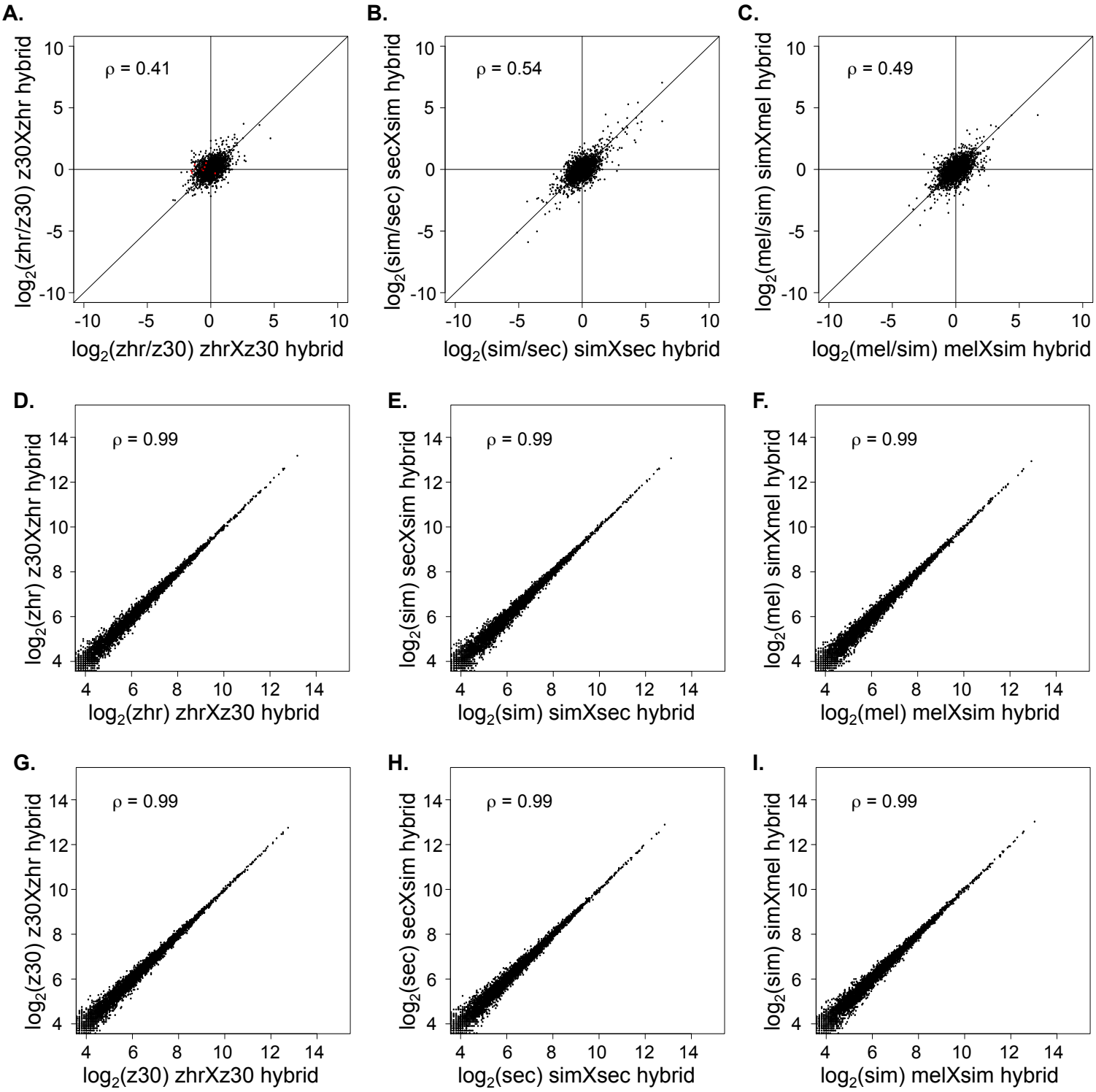

A.
ρ = 0.41
log2(zhr/z30) z30Xzhr hybrid
log2(zhr/z30) zhrXz30 hybrid
B.
ρ = 0.54
log2(sim/sec) secXsim hybrid
log2(sim/sec) simXsec hybrid
C.
ρ = 0.49
log2(mel/sim) simXmel hybrid
log2(mel/sim) melXsim hybrid
D.
ρ = 0.99
log2(zhr) z30Xzhr hybrid
log2(zhr) zhrXz30 hybrid
E.
ρ = 0.99
log2(sim) secXsim hybrid
log2(sim) simXsec hybrid
F.
ρ = 0.99
log2(mel) simXmel hybrid
log2(mel) melXsim hybrid
G.
ρ = 0.99
log2(z30) z30Xzhr hybrid
log2(z30) zhrXz30 hybrid
H.
ρ = 0.99
log2(sec) secXsim hybrid
log2(sec) simXsec hybrid
I.
ρ = 0.99
log2(sim) simXmel hybrid
log2(sim) melXsim hybrid

Figure S12

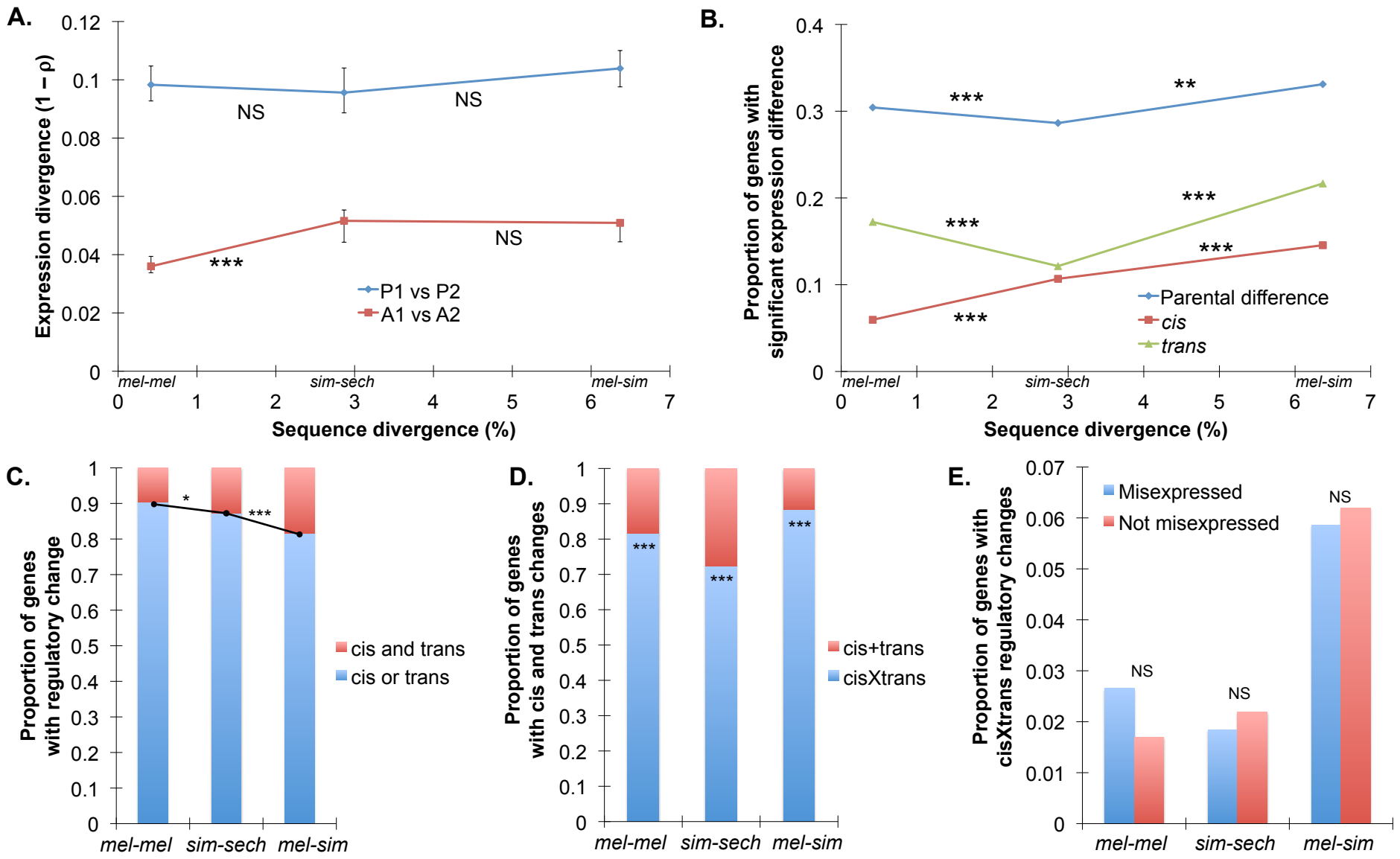

A.
Expression divergence (1 – ρ)
0.12
0.1
0.08
0.06
0.04
0.02
0
NS
NS
NS
***
P1 vs P2
A1 vs A2
mel-mel
sim-sech
mel-sim
0 1 2 3 4 5 6 7
Sequence divergence (%)
B.
Proportion of genes with significant expression difference
0.4
0.3
0.2
0.1
0
***
**
***
***
***
***
Parental difference
cis
trans
mel-mel
sim-sech
mel-sim
0 1 2 3 4 5 6 7
Sequence divergence (%)
C.
Proportion of genes with regulatory change
1 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0
*
***
cis and trans
cis or trans
mel-mel
sim-sech
mel-sim
D.
Proportion of genes with cis and trans changes
1 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0
***
***
***
cis+trans
cisXtrans
mel-mel
sim-sech
mel-sim
E.
Proportion of genes with cisXtrans regulatory changes
0.07 0.06 0.05 0.04 0.03 0.02 0.01 0
Misexpressed
Not misexpressed
NS
NS
NS
mel-mel
sim-sech
mel-sim

Figure S13

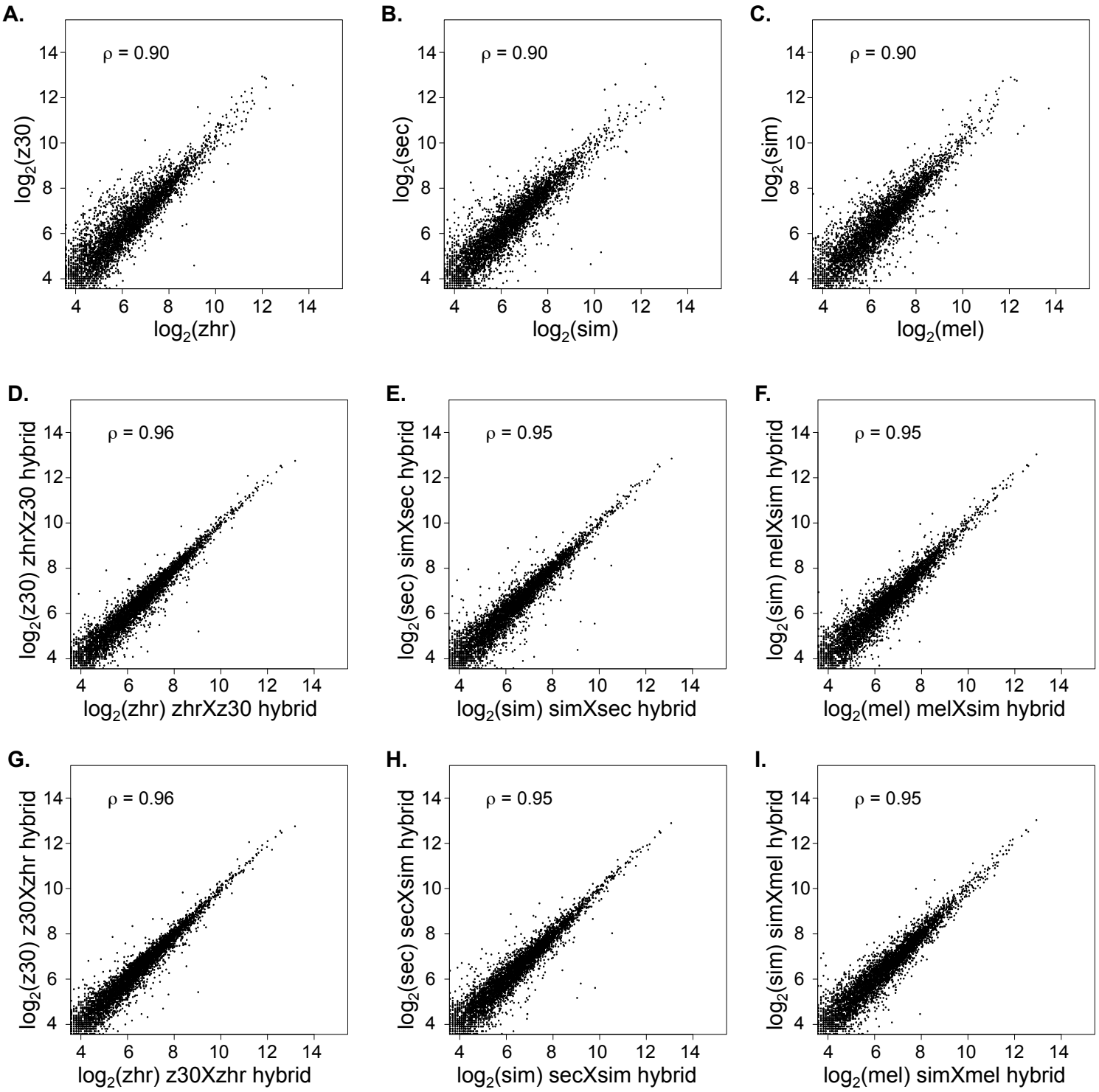

A.
ρ = 0.90
log2(z30)
log2(zhr)
B.
ρ = 0.90
log2(sec)
log2(sim)
C.
ρ = 0.90
log2(sim)
log2(mel)
D.
ρ = 0.96
log2(z30) zhrXz30 hybrid
log2(zhr) zhrXz30 hybrid
E.
ρ = 0.95
log2(sec) simXsec hybrid
log2(sim) simXsec hybrid
F.
ρ = 0.95
log2(sim) melXsim hybrid
log2(mel) melXsim hybrid
G.
ρ = 0.96
log2(z30) z30Xzhr hybrid
log2(zhr) z30Xzhr hybrid
H.
ρ = 0.95
log2(sec) secXsim hybrid
log2(sim) secXsim hybrid
I.
ρ = 0.95
log2(sim) simXmel hybrid
log2(mel) simXmel hybrid

Figure S14

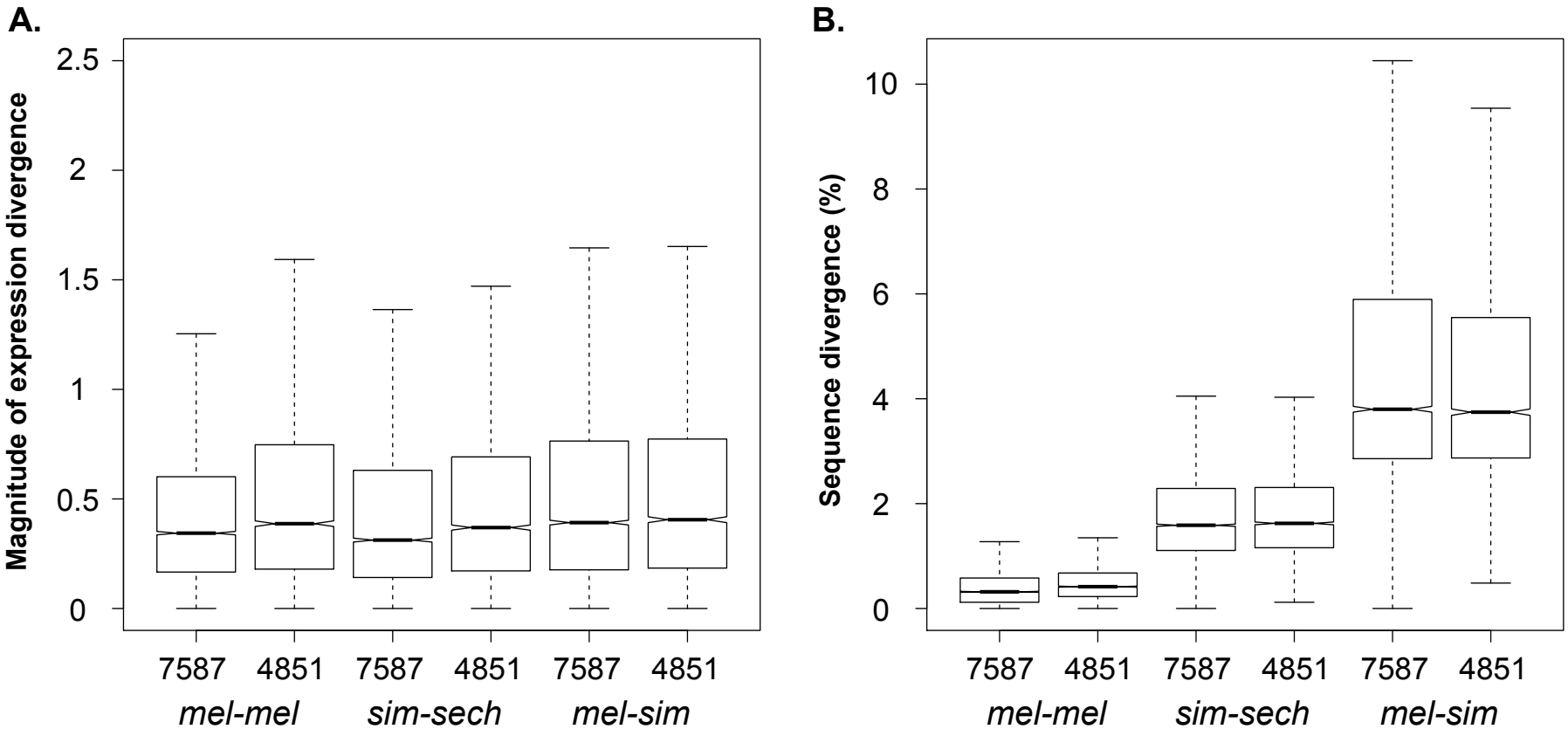

A.
Magnitude of expression divergence
2.5
2
1.5
1
0.5
0
7587
4851
7587
4851
7587
4851
mel-mel
sim-sech
mel-sim
B.
Sequence divergence (%)
10
8
6
4
2
0
7587
4851
7587
4851
7587
4851
mel-mel
sim-sech
mel-sim

Figure S15

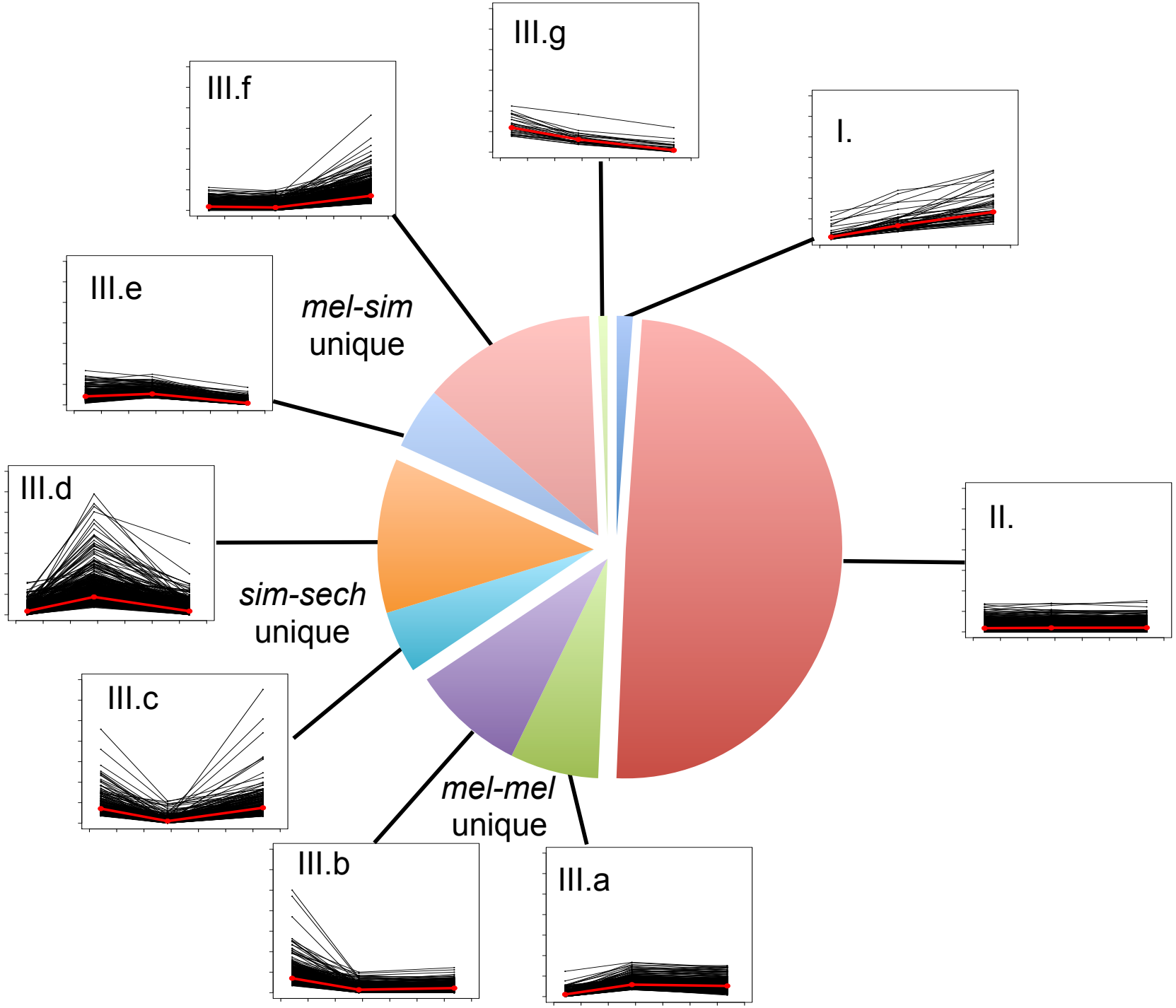

III.g
III.f
I.
III.e
mel-sim
unique
III.d
II.
sim-sech
unique
III.c
mel-mel
unique
III.b
III.a

Figure S16

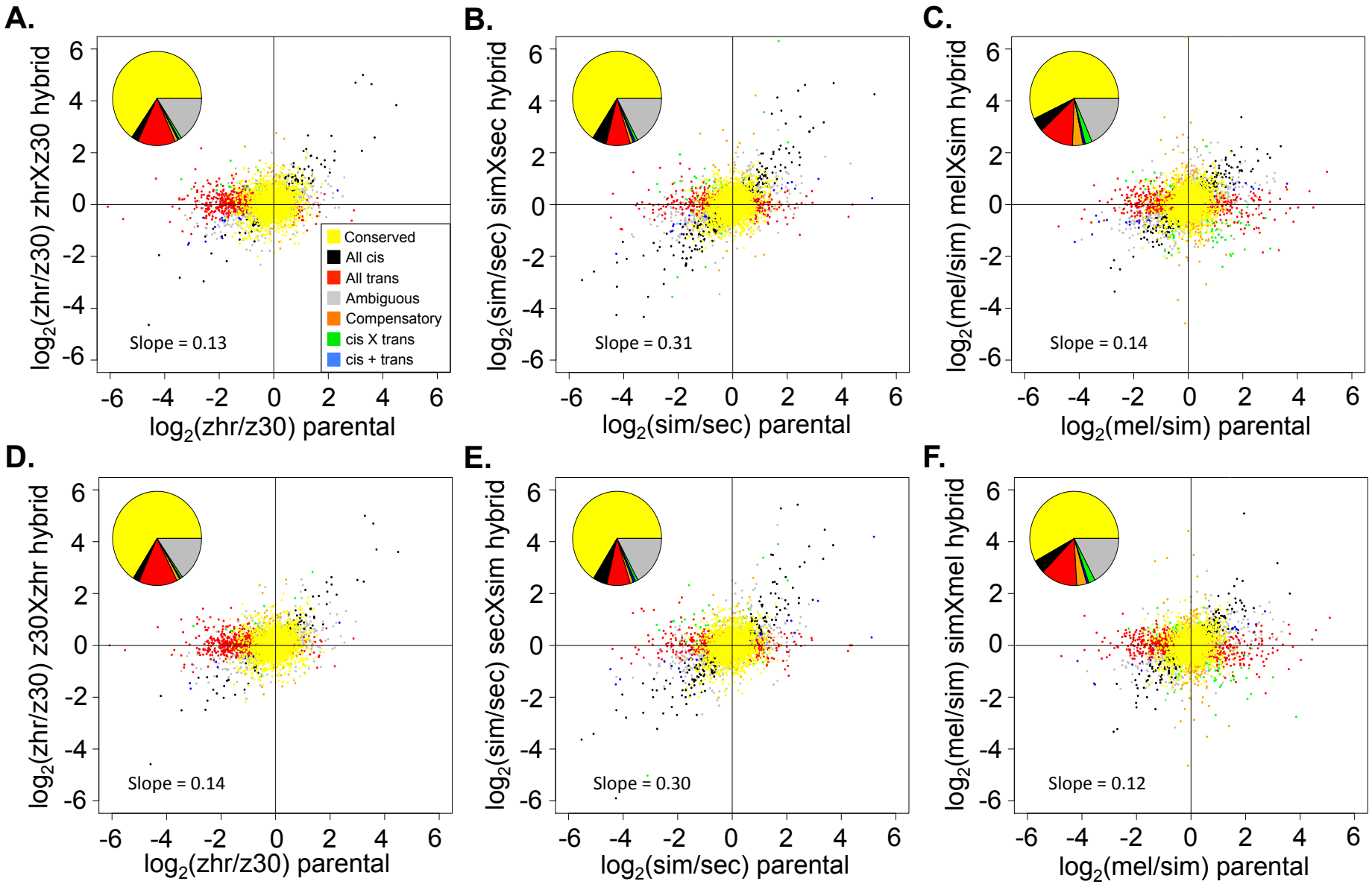

A.
B.
C.
D.
E.
F.
log2(zhr/z30) zhrXz30 hybrid
log2(zhr/z30) parental
Slope = 0.13
Conserved
All cis
All trans
Ambiguous
Compensatory
cis X trans
cis + trans
log2(sim/sec) simXsec hybrid
log2(sim/sec) parental
Slope = 0.31
log2(mel/sim) melXsim hybrid
log2(mel/sim) parental
Slope = 0.14
log2(zhr/z30) z30Xzhr hybrid
log2(zhr/z30) parental
Slope = 0.14
log2(sim/sec) secXsim hybrid
log2(sim/sec) parental
Slope = 0.30
log2(mel/sim) simXmel hybrid
log2(mel/sim) parental
Slope = 0.12

Figure S17

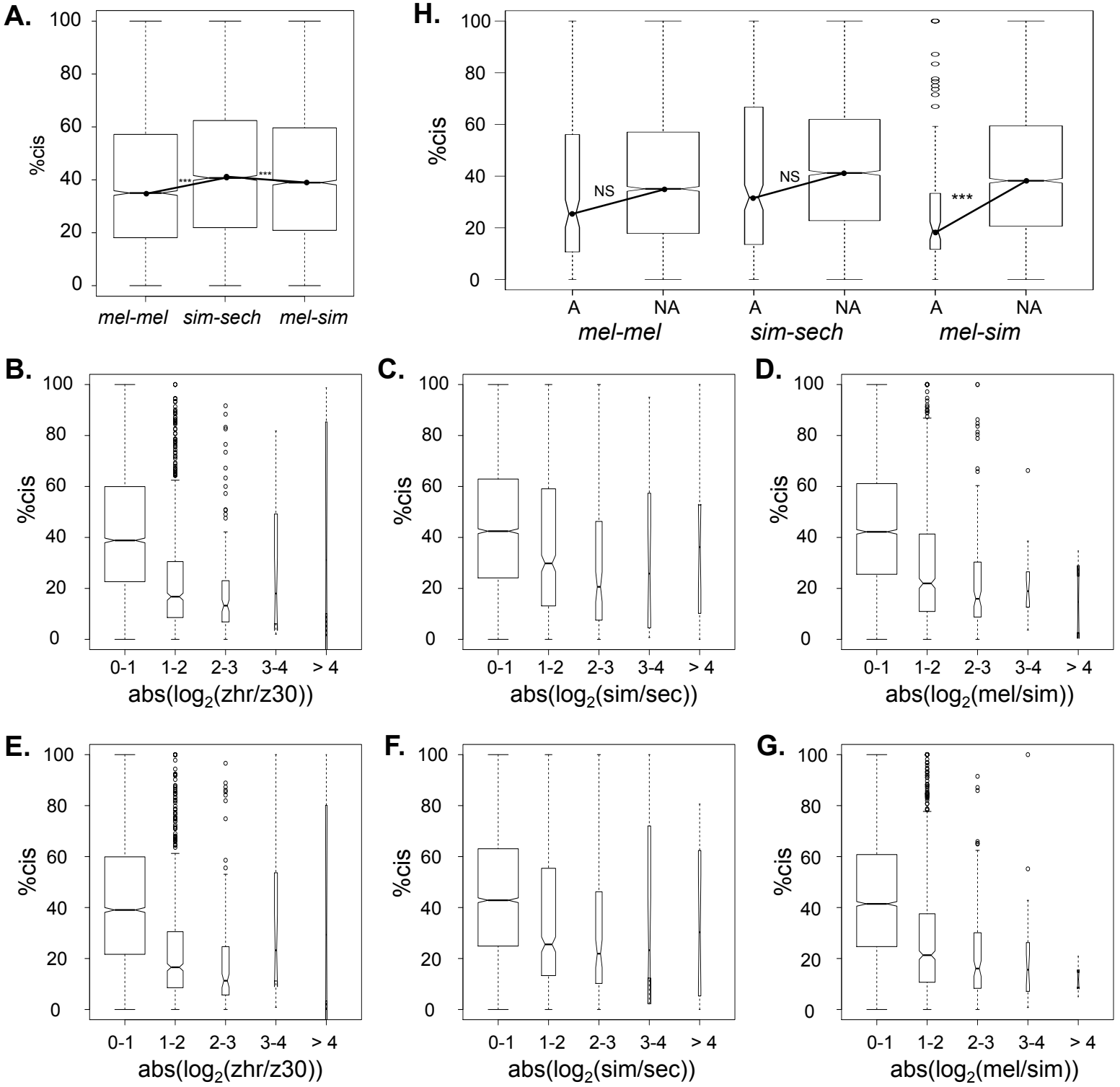

A.
%cis
mel-mel
sim-sech
mel-sim
***
***
H.
%cis
NS
NS
***
A
NA
A
NA
A
NA
mel-mel
sim-sech
mel-sim
B.
%cis
0-1
1-2
2-3
3-4
> 4
abs(log2(zhr/z30))
C.
%cis
abs(log2(sim/sec))
D.
%cis
abs(log2(mel/sim))
E.
%cis
abs(log2(zhr/z30))
F.
%cis
abs(log2(sim/sec))
G.
%cis
abs(log2(mel/sim))

Table S1. Summary of sequencing depth for RNA-seq and gDNA.						
	cDNA	**gDNA**				
Single genotypes						
D. melanogaster (zhr)	16464075	27051150*				
D. melanogaster (z30)	21806797	25863911				
D. simulans (sim)	18006673	31859837				
D. sechellia (sech)	15817452	27706737				
F1 hybrids						
zhrXz30	31432754					
z30Xzhr	31439998					
simXsech	19787136					
sechXsim	20059660					
zhrXsim	23929242					
simXzhr	25875801					
*This dataset was supplemented with an additional 15692412 single end reads from *D. melanogaster zhr* gDNA.						

Table S2. Number of genes suitable for quantifying total expression in each genotype.				
Dataset				
Parental genotypes	**Comparison**	**# mapped reads**	**# after downsampling**	**# genes > 20 reads**
D. melanogaster zhr	*mel-mel*	12915170	12704991	8949
D. melanogaster z30	*mel-mel*	18231082	12704991	9191
D. simulans Tsimbazaza	*sim-sech*	14811651	12704991	9257
D. sechellia droSec1	*sim-sech*	12704991	12704991	8981
D. melanogaster zhr	*mel-sim*	12905538	12704991	8947
*D. simulans Tsimbazaz*a	*mel-sim*	14818358	12704991	9259
F_1 hybrids				
D. melanogaster zhr X *D. melanogaster z30*	*mel-mel*	25957434	12704991	9204
D. melanogaster z30 X *D. melanogaster zhr*	*mel-mel*	26515174	12704991	8978
D. simulans Tsimbazaza X *D. sechellia droSec1*	*sim-sech*	17169600	12704991	8477
D. sechellia droSec1 X *D. simulans Tsimbazaza*	*sim-sech*	17517914	12704991	8484
D. melanogaster zhr X *D. simulans Tsimbazaza*	*mel-sim*	19484335	12704991	9307
D. simulans Tsimbazaza X *D. melanogaster zhr*	*mel-sim*	20936819	12704991	9215
All samples				7587

Table S3. Number of genes suitable for quantifying allele-specific expression in each genotype.						
Comparison	# with <10% mapping to wrong allele	# (A1+A2) ≥ 20	# in all three per comparison	# in all comparisons	# "imprinted"	Final number
mel-mel						
parents vs hybrid 1		6126				
parents vs hybrid 2		6003				
hybrid 1 vs hybrid 2		5954			9	
parents	12433		5921	4860		4851
sim-sech						
parents vs hybrid 1		7274				
parents vs hybrid 2		7279				
hybrid 1 vs hybrid 2		7159			0	
parents	12325		7149	4860		4851
mel-sim						
parents vs hybrid 1		8602				
parents vs hybrid 2		8526				
hybrid 1 vs hybrid 2		8592			0	
parents	13281		8439	4860		4851

Table S4. Accuracy of mapping maternally inherited mitochondrial alleles in interspecific F1 hybrids.

Hybrid cross	# correct	Total #	% correct
D. simulans Tsimbazaza X D. sechellia droSec1	4879	4905	99.47
D. sechellia droSec1 X D. simulans Tsimbazaza	9304	9357	99.43
D. melanogaster zhr X D. simulans Tsimbazaza	34791	34833	99.88
D. simulans Tsimbazaza X D. melanogaster zhr	18257	18290	99.82

Table S5. Criteria for assigning genes to regulatory evolution classes.					
Classification	**Fisher exact test[1]**	**Binomial exact test[2]**	**Fisher exact test**	**Additional criteria**	
conserved	P1 = P2	A1 = A2	P1/P2 = A1/A2	N/A	
all cis	P1 ≠ P2	A1 ≠ A2	P1/P2 = A1/A2	N/A	
all trans	P1 ≠ P2	A1 = A2	P1/P2 ≠ A1/A2	N/A	
cis+trans	P1 ≠ P2	A1 ≠ A2	P1/P2 ≠ A1/A2	log2(P1/P2)/log2(A1/A2) > 1	
cisXtrans	P1 ≠ P2	A1 ≠ A2	P1/P2 ≠ A1/A2	log2(P1/P2)/log2(A1/A2) < 1	
compensatory	P1 = P2	A1 ≠ A2	P1/P2 ≠ A1/A2	N/A	
ambiguous	P1 ≠ P2	A1 = A2	P1/P2 = A1/A2	N/A	
ambiguous	P1 = P2	A1 ≠ A2	P1/P2 = A1/A2	N/A	
ambiguous	P1 = P2	A1 = A2	P1/P2 ≠ A1/A2	N/A	
[1]Compares expression between parental genotype 1 (P1) and parental genotype 2 (P2).					
[2]Compares expression between allele 1 (A1) and allele 2 (A2) in F_1 hybrids.					

Table S6: Pyrosequencing assays for quantification of allelic expression ratios

mel-mel							
Gene	**Biotin[1]**	**Forward primer**	**Reverse primer**	**Pyro primer**	**Sequence to analyze[2]**	**Dispensation[3]**	**Fomula *(zhr/z30)*[4]**
CG12819	F	GAGCAATGACGACTCAAGTGAGA	GGCCGCTTCGCTCAATTTA	GCATTGGATGAGCAG	KTACTTTG GTTTTTGGGC TTCTCTGT	c**TGT**cACTG	(G1+T2)/T1
CG17530	R	AACTTTTCTCAAGGGACAGGATTA	ATCCGCAATGGTCAGTTGA	CAGGATTACATTGCTGG	SAATCAAC TGACCATTGC GGATTTCA	t**CG**tATCACT	C1/G1
CG18444	F	ACTGCGCTGCAGGGAGAT	TTGGATGGCCGCATTGTC	ATGGCCGCATTGTCG	GYGGCTCT GCTACCACCA TCAGCAGCT	a**GCT**aGCTCTG	C1/T1
CG6253	F	ACCGCTTCAAGCTGCGTTATG	TGGTGCGCTTCTTCAGTGTG	CAGCTTGTTGCGCTG	RCGCTTGG CATAACGCAG CTTGAAGC	c**AG**tCGCTGC	(A1*0.86)/G1
CG6906	R	TCCCAGATTGACCGCAAAG	TCGCCCAGTGGAAGTGGAT	AAAGGCAAGAGCGTTA	AYATGACC AATCCACTTC CACTGGGCG	g**AT**CgATGAC	T1/C1
CG7269	F	AGGGGTGCCCACCACAAT	AAAGGACGAGGAGACCCTCAAG	GAGCGGCACCCCGCA	YATTGTGG TGGGCACCCC TGGCCGAA	g**CT**gATGTGT	T1/C1
CG7874	F	GTGGCCAGAGGTTCGAGC	CTTCAGCTCCCGCTTGAT	CTCCTGTCCGGGCAC	RATCTGCG TCCAGTCATA GCGCCAGC	c**AGA**cTCTGC	(A1*0.86)/(G1+(A2*0.86))
CG9089	R	GCGTGGCCATTGGACTTAG	CCAACGATACCTCCAAAGGATCT	CCATTGGACTTAGCCTT	YGAGGGAG ATCCTTTGGA GGTATCGT	a**CT**aGAGAGAT	C1/T1
CG9497	R	CCTTTTCCTGCGGTGCCT	CGAGAAGATCGGCAAGTACATG	GTTCCAAGTTGGCCC	KCTGATTT TGCATGTACT TGCCGATC	a**TG**aCTGA	T1/G1
CG9916	F	GAAACCCAACATGTCTATTGAAAA	TCGAATCTTATGGATCGCAGTC	GGTTCTCTTTAAGAAGGTG	YGATGCAA AAGCAACAGC AACATTGC	g**CT**aGATGC	T1/C1
sim-sec							
Gene	**Biotin**	**Forward primer**	**Reverse primer**	**Pyro primer**	**Sequence to analyze**	**Dispensation**	**Fomula (*sim/sec*)**
CG1644	F	TTCTGCGAACCCATTGTGTC	ATCCGGCTCTGATCAAGCG	AGACGACTGATCCCAC	TTGYGACACA ATGGGTTCGC AGAACCTT	aTTTG**CT**aGAC	C1/T4
CG2233	R	AAGGGACCGGGAAAATTGT	AACACCCTCTTCGCCAGACT	AAAGGCGGGAAACTC	GWGGAGTCTG GCGAAGAGGG TGTTCCT	cG**AT**cGAGTC	T1/A1
CG3209	F	TGTGCGATGTGTGGTACCTG	CACTCCTTCTTTGGCTTCATGC	CGATGACGCTCTTCAC	KCGGTTCGCG AAGTCGATCG CCGACT	a**GT**aCGTCGC	G1/T1
CG4703	F	CTCGTTCGCCGGTGGTAA	GAAATCTCAAGGGGATGTGC	CCCTGTATCTGGGACC	RCTTCTCTCC TTTGGAAACG ATGCCC	t**AG**tCTCTCT	G1/A1
CG4783	F	GGCACGAAAAACCAGAAATC	ATCCGAAAGTTTGGAGCTCA	GTCGTACAATCCCACATA	MGATTTCTGG TTTTTCGTGC CAACTG	g**AC**tGATCTG	C1/A1
CG4797	R	CACTCATTTTGGCCGGCTATA	CGCGACTACTTGGTTACATAAAGG	TTTGGCCGGCTATAATG	TASTTCAAGT TGTTTAGTTC AAGTTGTT	cTA**CG**aTCAG	G1/C2
CG5269	F	CAATCAAAAGTTCCAGCCTACA	TAATCGTTCGACAATAGGGATAGG	CCAATGAGCCCTTGG	GYGCCAATAG TCTTATGTAG GCTGGAA	t**GCT**aGCATA	T2/C1
CG6201	R	GTATTCCGCTTGGGTCGTACTG	CACACAATGCCGCACTTCT	GCTGGATCGCCCAGG	GYGGCGCTAA TCCCCTCGTG ATCGATC	c**GCT**aGCGCT	T1/C2
CG7702	R	ACTGCTGATCTTGCCGATGT	TTTCCGTTTTCGGCTTCA	AAGAAACACGCGCGA	RTGCTGAAGC CGAAAACGGA AACGTT	t**AG**cTGCTGA	A1/G1
CG7923	R	TGTCGACGGAGCTGAGCAAT	AGCCACACTCCATATAGCGAAGAT	ACCGATGGCAGGAAA	MTGRAGCTGG TATGGTTCAA TATAGT	t**AC**gTGAGCT	A1/C1
CG9009	F	GTCGGGCTTTTGGAGATTGAA	TTCGCCCAAATGCGAGAT	TGCCAGCGCCGCTTT	YGCCGTTCGT CTGCAGACCA AGTTCA	a**CT**aGCGTCG	T1/C1
CG9673	R	CGTATCCGGCTTCCCATT	TGCATCCTACAAGCAGCAGAAG	GCGATCGGCTCAGCGT	RTTGTAGTTG GCCTTCTGCT GCTTGT	t**AG**cTGTAGT	G1/A1
CG10877	F	AAAACTTTCCACACACATCATACA	TGGTTCTTTTTTGTTGCTACTTGT	TGTGGGTCAACTAATCTG	TMGCCAGTAC ACA	cT**AC**tGCAGT	A1/C2
CG10916	R	AGCTCAGTATCCATAGGGACGTAG	GTCATTTGTGGGTGCAGTTTCT	TGGAAGGTACCGGCG	SAGAAGTCRC AGAAACTGCA CCCACA	a**CG**tAGAGTC	C1/G1
CG11391	R	GTAGAGACTGGCGATTGTTTTGTG	TTGGCCTACGGTTGCTTG	GTTGACGGCGTGAAA	YGGAGTTCCG TTGAACAAGC AACCGT	a**CT**aGAGTCG	T1/C1
CG11407	R	TCCGATGGACAGATGTGG	TGATCGCAAGATCTGGAGC	CAGATGTGGGTCAAAGTAG	YATTCCAATT GGTCTCCGCT CCAGAT	g**CT**gATCATG	T1/C1
CG15883	R	AAATTACGAGGCGATAGTTCCG	GACACGTGCGACTATCCATACCT	GCCCTTAAAATAACACTG	KAGAATCAGG TATGGATAGT CGCACG	a**GT**cAGATCA	T1/G1
CG17186	F	ATGCGGAGGAAACACGATAC	ATTCATTCGGCATCAACTGG	TTTGCACAACCTCAAT	TTGYGCCTGG CTTTTTGCTT GCATCTTGA	cTTTG**CT**aGCT	T4/C2

[1]Indicates whether the Forward (F) or Reverse (R) primer was biotinylated.

[2]Sequence immediately following the 3' end of the Pyro primer. Sites used to differentiate the two alleles are indicated with IUPAC ambiguity codes.

[3]Order of nucleotides added to Pyrosequencing reaction. Lowercase bases are not expected to be incorporated and serve as controls. Peak heights resulting from incorporation of the bases in bold were used to calculate relative allelic expression.

[4]Describes the way in which peak heights from individual bases were used to calculate relative allelic expression. Each base is indicated by XN, corresponding to the Nth occurance of the X base. A correction factor of 0.86 was used for peaks produced by incorporation of A, as recommended by the manufacturer.