Nucleic Acids Research Advance Access published November 17, 2015

doi: 10.1093/nar/gkv1193

Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes

Reazur Rahman[1], Gung-wei Chirn[1], Abhay Kanodia[1], Yuliya A. Sytnikova[1], Björn Brembs[2], Casey M. Bergman[3] and Nelson C. Lau[1,*]

[1]Department of Biology and Rosenstiel Basic Medical Science Research Center, Brandeis University, Waltham, MA 02454, USA, [2]Institute of Zoology, Universität Regensburg, Regensburg, Germany and [3]Faculty of Life Sciences, University of Manchester, Manchester M21 0RG, UK

Received September 23, 2015; Revised October 22, 2015; Accepted October 24, 2015

ABSTRACT

To understand how transposon landscapes (TLs) vary across animal genomes, we describe a new method called the Transposon Insertion and Depletion AnaLyzer (TIDAL) and a database of >300 TLs in *Drosophila melanogaster* (TIDAL-Fly). Our analysis reveals pervasive TL diversity across cell lines and fly strains, even for identically named sub-strains from different laboratories such as the ISO1 strain used for the reference genome sequence. On average, >500 novel insertions exist in every lab strain, inbred strains of the *Drosophila* Genetic Reference Panel (DGRP), and fly isolates in the *Drosophila* Genome Nexus (DGN). A minority (<25%) of transposon families comprise the majority (>70%) of TL diversity across fly strains. A sharp contrast between insertion and depletion patterns indicates that many transposons are unique to the ISO1 reference genome sequence. Although TL diversity from fly strains reaches asymptotic limits with increasing sequencing depth, rampant TL diversity causes unsaturated detection of TLs in pools of flies. Finally, we show novel transposon insertions negatively correlate with Piwi-interacting RNA (piRNA) levels for most transposon families, except for the highly-abundant *roo* retrotransposon. Our study provides a useful resource for *Drosophila* geneticists to understand how transposons create extensive genomic diversity in fly cell lines and strains.

INTRODUCTION

Transposons comprise major portions of nearly all sequenced animal genomes because they continue to successfully proliferate in spite of host mechanisms that suppress their activity. One conserved transposon-suppressing mechanism is the Piwi/piRNA pathway, in which germ cells produce piRNAs, small RNAs antisense to transposon sequences that target the Piwi proteins to transposon transcripts to engage silencing processes. For transposons to persist and spread, they must evade suppression mechanisms and mobilize to new genomic loci that either benefit or do not harm the fitness of the host (1). Changing copy numbers and locations of transposons within genomes can be perceived as a dynamic 'landscape' of transposons that can profoundly affect the architecture of the host animal genome.

To better understand how truly diverse transposon landscapes (TLs) are across broad numbers of animal genomes, we need computational tools that efficiently and accurately quantify new transposon Insertion and Deletion (InDel) events in genomic data. With short-read deep sequencing becoming commonplace, a plethora of model organism genomes is now available that enable new insights into the dynamics of TLs across individuals and populations. A prominent example of this genomics revolution is the immense trove of *Drosophila melanogaster* genomes (currently >600) that have been sequenced to high coverage, including genomes from worldwide populations, cell lines, and laboratory strains (2–14).

Drosophila melanogaster's compact genome and global cosmopolitan distribution makes it a prime model system for population genomics studies. As such, the majority of resequenced genomes in this species have been sampled from natural populations, including: (i) the *Drosophila* Genetic Reference Panel (DGRP), a bank of ∼192 highly-inbred strains from Raleigh, North Carolina, maintained in the Bloomington *Drosophila* Stock Center (BDSC) for Genome-Wide Association Studies of biomedical-relevant traits (8,11,14); (ii) the *Drosophila* Genome Nexus (DGN), a broad collection of genome sequences from several in-

*To whom correspondence should be addressed. Tel: +1 781 736 2445; Email: nlau@brandeis.edu
Present addresses:
Gung-wei Chirn, iBioinfo Group LLC, Lexington, MA, USA.
Yuliya A. Sytnikova, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

dependent population studies of *D. melanogaster* strains isolated from Europe, the Middle East and Sub-Saharan Africa (10,12) and (iii) various pools of flies sampled from several locales in the United States, Austria, Italy, Portugal and Australia (6,7,9,13).

Since Illumina short-read sequences are currently the dominant format for these population genomics studies, most TL analyses in *D. melanogaster* entail comparing reads to the reference genome sequence of the ISO1 strain (genotype of y^1; cn^1 bw^1 sp^1) (15–18). In this genome, the large bulk of transposons is densely packed near the telomeres and in pericentromeric heterochromatin, as well as specific transposon-dense chromosomes like the Y and fourth (19–22), which are still notoriously challenging to assemble and annotate with even current genomics technologies. Although heterochromatin is generally thought to be inert (23), it serves important structural roles for chromosome maintenance (24–27), and reproduction (28). However, another ∼10% of the euchromatic genome is also filled with transposons that can reside within gene introns or near gene promoters, and thus able to influence gene expression (21,22,29–35). The most recent version of the reference *D. melanogaster* genome sequence, called Release6 (or dm6), has merged all major euchromatic and heterochromatic scaffolds into one assembly (20,36). RepeatMasker (37) annotates ∼32 750 transposon loci in this release, comprised of 135 well-characterized transposon families. These transposons can also be mainly broken down to Class 1 retrotransposons (∼70%), Class 2 DNA transposons (∼10%) and rolling-circle transposons (∼18%, *DINE-1* elements, (38)).

Several programs have been previously described for detecting *de novo* transposon InDels relative to the Release5/dm3 *D. melanogaster* genome (9,39–41), whereas other programs have been specifically developed for determining the presence and absence of reference-genome annotated transposons (42,43). These tools have been used to define preferred Target Site Duplication (TSD) sequences for transposon insertions (39,41,42), and reveal the frequent occurrence of transposons at low allele frequency at many genomic loci (40). These different programs have typically not been applied to identical datasets, however previous work found that only a small minority of transposon insertions were called in common by three programs on the same genomic sequence input (44). Thus, the best approach for determining TLs from Illumina sequences remains an unresolved problem for geneticists.

Here, we introduce a new bioinformatics pipeline that generates annotation-rich outputs of TLs for individual genome inputs called the Transposon Insertions and Depletion AnaLyzer (TIDAL). In this study, we have chosen to refer to transposon absences as 'depletions' rather than 'deletions' because in pooled or heterogeneous samples the absence of a reference transposon may not be complete. Furthermore, the absence of a reference transposon in a sample may not actually be a deletion but rather an insertion in the reference genome sequence. Our tool rapidly generates TL datasets in a format that is amenable to aggregate analyses, and tracks the ratio between the reads supporting the transposon InDel relative to the reads corresponding to the unaltered genome sequence.

Using this new tool, we have generated transposon annotations for >300 *D. melanogaster* genomes from wild populations, lab strains and cell lines, and created a website database called TIDAL-Fly (http://www.bio.brandeis.edu/laulab/Tidal_Fly/Tidal_Fly_Home.html) that displays these data in user-friendly format for simple sorting and text searching in spreadsheet programs as well as aggregate analysis with Structured Query Language (SQL). Importantly, the TLs in TIDAL-Fly were run on the latest genome assembly improvements of Release 6/dm6 (20). Furthermore, we benchmarked TIDAL against three other programs on transposon insertions using identical DGRP datasets and we conducted genomic PCR to experimentally assess TIDAL's outputs. Finally, we use TIDAL-Fly outputs to provide new insights to key genomics questions such as: how frequently and how many transposon loci differ between individual fly strains, and which transposon families are most prevalent in these differences? How do TLs differ between *D. melanogaster* cell lines to individual fly strains to pools of flies? Do fly strains from different labs with identical strain name possess similar or different TLs? How does read length and sequencing depth influence the number of discovered transposon InDels?

MATERIALS AND METHODS

Library construction and deep sequencing of genomic DNA from *D. melanogaster* cells and lab strains

Genomic DNA (gDNA) for the S2c1 cell line was extracted from 5×10^6 cells lysed overnight in SDS lysis buffer containing 20 uM Proteinase K, and phenol/chloroform extraction. 8 ug of gDNA was then fragmented in a Bioruptor sonicator (Diagenode) with 8 cycles (20 s pulse and 90 s pause) at the 'High' power level. Fragmented DNA was used for library construction performed essentially as described (45). After end repair and 3′-A tailing, we performed adaptor oligo duplex ligation (duplex of PE_Tdot_common_C* and barcoded linker, Supplementary Table S2). Ligation products of size between 400 and 450 bp were agarose-gel purified and used for PCR amplification with primer oligos PE-POSTPCR_1 and PE-POSTPCR_2. Amplified libraries were gel purified and quantified on an Agilent Bioanalyzer. A single-end 150 nt sequencing run was performed on the MiSeq with the version 3 kit.

DNA for CanS sub-strains from (46) was extracted from 20 females using the Qiagen DNeasy Blood and Tissue Kit (Cat. No. 69504), with the final elution step replaced by an ethanol precipation. Input DNA was tagmented using the Nextera DNA sample preparation kit (Cat. No. FC-121-1030). Following a cleanup using the Zymo-Spin kit (Cat. No. D4023) the purified, tagmented DNA was then amplified via limited-cycle PCR which also added the indices (i7 and i5) and sequencing primers. AMPure XP beads (Cat. No. A63881) were then used to purify and size select the library DNAs. The libraries were then normalized to 2nM and pooled prior to cluster generation using a cBot instrument. The loaded flow-cell was then paired-end sequenced (2×101 nt) on an Illumina HiSeq2500 instrument. Demultiplexing of the output data (allowing one mismatch) was performed with bcl2fastq 1.8.4.

Genomic sequence data that we generated for this study has been deposited in the Sequencing Reads Archive (SRA) under the accession numbers SRR1983913 and ERP009394. All the other genomic DNA and small RNA datasets used in this study were downloaded from the SRA, with all accession numbers recorded in Supplementary Table S1.

Construction of the TIDAL pipeline

The initial framework for the TIDAL pipeline was built for detecting TE insertions from single-end Illumina reads using a split-read approach (33). Here, we improved and expanded the TIDAL pipeline to also detect TE depletions from the same input of single-end reads, and automated the pipeline using a combination of shell, C and PERL scripts that follows the detailed flowchart path in Supplementary Figure S1. Source code for the TIDAL program can be found on GitHub at: https://github.com/laulabbrandeis/TIDAL.

TIDAL starts with an input file listing the library name, the SRA Run accession numbers (SRR#), and a user-defined read length reflecting the size of reads in the library. The specified SRA file is downloaded and converted to a FASTQ file with the fastq-dump command from the SRA Tool kit using the '–split-spot [readlength]' parameter to convert all paired-end reads into single-end reads. In case of the DGRP libraries, one SRA accession might have several libraries of varying length reads, so only reads within 10 nt of the specified length is extracted with a custom perl script. Trimmomatic (47) was used to trim low quality bases from 5′ and 3′ end of the read and if the average quality of the read is below 20 (using parameters 'LEADING:20 TRAILING:20 AVGQUAL:20').The trimmed reads are then mapped to the Release6/dm6 reference genome sequence with Bowtie2 (48) using parameters '–sensitive –end-to-end'. The alignments are filtered with samtools (49) to retain those with MAPQ scores $\geq$ 10. These alignments are used for coverage analysis later in the pipeline and by Control-FREEC application (50) to generate a Copy Number Variation genome chart in PDF format (using parameters step = 5000, dm6 chr file and pre-computed 100mers GEM mappability tracks). All the unmapped reads from the reference genome mapping are aligned to *Drosophila* viruses, structural RNAs, and transposon consensus sequences curated from Repbase (51) and Flybase (20) using Bowtie1 (52) (using parameters '-v 3 –k 2 –m 100000'), the unmapped reads used for split read analysis. Less than 5% of the reads from the input library is left at this stage and fraction of reads culled at each step of the pipeline is recorded in a file. For TE insertions, the 22 nt long 5′ and 3′ termini of these unmappable reads were mapped with Bowtie1 to TE consensus sequences with parameters '-v 2 -k 5 –m 5', and to the masked reference genome with parameters '-v 1 -k 5 –m 5′. Reads with one termini mapping to $\leq$5 TEs and one termini mapping uniquely to the genome are potential candidate reads that shows evidence of de novo TE insertion. These candidate reads are aligned to the reference genome with BLAT (53), where the BLAT alignment score reflects the fraction of bases from the read that can mapped to the reference genome. If almost the entire read maps to the reference genome then it will have a very high BLAT score and is likely to be a false positive read. These reads are then grouped into clusters where the genomic coordinates of the split reads mapping are from same strand and fall within 300 nt of each other and the other end correspond to the same TE family. We only retain clusters with at least 4 reads and cluster size that is greater than half of the read length minus 22 nt, whereas clusters in specific repeat-rich regions (not marked by RepeatMasker) are discarded. The last key validation step is to discard false-positive clusters which display an average score >83% for all of the BLAT score ratios of the full length reads within the cluster. A coverage ratio (CR) for each TE insertion is then determined for the interval of the cluster window, expanded by 22 nt at the 5′ coordinate, by dividing the number of TE insertions reads over the number of reference genome mapped reads, which are estimated by using coveragebed (54) from earlier Bowtie2 alignment to reference genome, plus a pseudocount of 1. For example, a TE insertion that has 30 TE-insertion reads and 2 reference genome mapping reads results in a CR of 10 = [30/(2 + 1)]. By using Refseq annotation tracks from UCSC genome browser track, we then annotated the identity of nearby genes for each TE insertion; and if no genes are nearby then it is annotated as intergenic.

For TE depletions, both split read termini are mapped to the masked reference genome with Bowtie1 using parameters '-v 3 -k 5 –m 5'. Reads are marked as candidates where both termini maps uniquely with the same orientation to the same chromosome and distance between the 22mers is greater than read length of the library. These split reads are then grouped into clusters if their genomic coordinates and orientation fall within a maximum of 300 nt of each other. Clusters with at least four reads are retained and clusters in specific repeat-rich regions (not marked by RepeatMasker) are discarded. This step generates extensive lists of depletions, for which a majority are relatively small genomic deletions of <500nt that may not correspond to any TE sequence. Since TIDAL does not determine precise breakpoints of TE depletions, we instead rely on the coordinates of the read clusters for estimate of coverage. The CRs for each of these depletion sites are determined with this formula, CR=(depletion reads)/(1 + average(RefGen_5p, RefGen_3p)), where depletion reads is the number of reads in the cluster, RefGen_5p represents the number reference genome reads at a fixed interval near 5′ breakpoint and RefGen_3p represents the number of reference genome reads at a fixed interval near 3′ break point. The interval near 5′ breakpoint is defined as the 5′ read cluster − end coordinate − 2/5*(read length) to 5′ end coordinate + 22 + 1/5*(read length). The interval near 3′ breakpoint is defined as the 3′ read cluster start coordinate − 1/5*(read length) to the 3′ start coordinate +1/5*(read length). The 'coveragebed' tool is used to count the number of mapped reads in these interval from the earlier reference genome alignment by Bowtie2

The genomic coordinates of each TE InDel is then searched against the RefSeq and RepeatMasker annotation tracks downloaded from the UCSC Genome Browser (55) for the Release6/dm6 genome, to assign the closest gene names and the closest TE name for depletions. If no genes are nearby, the annotation defaults to 'intergenic, not near

genes' whereas a blank is listed for depletions that do not encompass a reference TE sequence. The depletions table is filtered to keep the entries annotated with a TE name, and the final data both from the insertion and depletion components of the pipeline is saved in BED format. The reference genome coordinates are then binned into 5kb intervals, and the counts of the TE Insertions and Depletions per 5kb bin are tabulated into the Fixed-Bin table. Finally, this fixed bin table is used by a custom R script to generate the final transposon landscape genome charts as PDFs. All text file and PDF outputs were then connected to hyperlinks to form the TIDAL-Fly database website that is hosted at: http://www.bio.brandeis.edu/laulab/Tidal_Fly/ Tidal_Fly_Home.html.

Analyses of transposon landscapes (TLs) and small RNAs

TIDAL tables were imported into Microsoft Access and aggregate analyses were conducted with queries written in Structured Query Language (SQL). Profiles of TE families were then compiled and plotted in Microsoft Excel, tracking individual families with ≥20 InDel events while grouping the remainder in an aggregated category. The TE families were then ranked according to their frequency of InDels, and then plotted as proportions in the stacked column charts. To compare transposon InDels between different TE-prediction programs, the genomic coordinates for each InDel was first converted to Release6/dm6 with the UCSC Genome Browser LiftOver tool (55). Coordinates were then rounded to the nearest kilobase (kb) to normalize the small numerical differences in the InDel coordinates calls between the different programs. Empirically, <3% of InDels had coordinates within 1 kb of another InDel, therefore the numerical rounding strategy was sufficient to maintain unique configurations for each program's TL. Venn diagrams of overlapping TE InDel coordinates between up to five libraries were conducted in R-studio with the Vennerable package (https://r-forge.r-project.org/projects/vennerable).

The libraries for DGRP fly strains ovarian small RNAs were downloaded from the SRA with the project accession number SRP019948 (44). The only small RNA library from that project not analysed for this study was RAL-427 (SRR1572816) because no Illumina genomic sequence was available for this DGRP line (only low coverage 454-genomic sequence). Small RNA reads were quality checked by FastQC (http://http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), sorted according to the barcode sequence in their 5′ adaptor, and then adaptor sequences were trimmed by FASTX-Toolkit http://hannonlab.cshl.edu/fastx_toolkit/. Structural RNAs were determined by cross-mapping to a custom database, and removed from subsequent analyses. Mapping to TEs was performed with Bowtie1 against a list of *Drosophila* consensus TE sequences obtained from the Repbase database (51) and from FlyBase (22), while virus sequences were obtained from Genbank. Up to three mismatches were allowed in the small RNA mapping to the TE consensus sequences. The basic processing pipeline is written in shell script (process-quick.sh) described in (33).

Genomic PCR assays of transposon insertions and depletions

We arbitrarily selected 49 TE InDel candidates with CRs≥3.0 from the S2c1 cell line (15 insertions, 10 depletions) and the ISO1-BL fly strain (12 insertions, 12 depletions). We also selected 48 total TE insertions that were either predicted only by TIDAL, by the LnB program (which were also predicted by TIDAL), only by TEMP, and only by the CnT program (12 sites each). Primers flanking the candidate TE InDel were designed to initially amplify a short amplicon of between 150 and 500 bp, or a long amplicon containing the TE which can range from 800bp (a solo LTR) to >9 kb (full-length intact TE). Noting that very large amplicons can be challenging to amplify, we employed multiple control genomic DNA samples to compare with the target genomic DNA sample, reasoning that the same PCR capable of amplifying the small non-TE amplicon in control samples might simply have greatly reduced or absent amplicons in the target genomic DNA sample because of the presence of the TE. We also tested one set of primer combinations that used a reverse primer base-pairing to the TE sequence, but found such TE-pairing primers to fail more frequently or generate amplicon artifacts. The list of primer sequences are listed in Supplementary Table S2.

Since we extracted one bulk sample of genomic DNA for each *D. melanogaster* cell line (S2c1) or fly strain (ISO1-BL, RAL-362, RAL-517, and RAL-765), these single DNA samples were used throughout the genomic PCR assays to demonstrate consistency in the PCRs and primers. 50ul PCR mixes contained 10 ng of gDNA template, 0.5 uM primers, 0.3 mM dNTPs, 1× GC buffer, 1 M Betaine, and 1 unit Phusion polymerase that was added only after the reaction reached 95°C for a hot start protocol. Annealing temperatures ranged from 52 and 68°C as optimized for particular primer pairs, and 10 minute extension times were used in 35 total PCR cycles. Amplicons were electrophoresed in 1% agarose gels.

RESULTS

TIDAL discovers transposon insertions and depletions with a more accurate split-read approach

While attempting to apply previously developed transposon InDel programs on OSS and OSC cell genomes (33), we discovered certain pitfalls in short read mapping to the Release5/dm3 *D. melanogaster* genome and transposon consensus sequences that affected the calls and total counts of transposon insertions (9,39–41). First, many reads in nearly all Illumina-sequenced *D. melanogaster* genomic libraries contain multiple single nucleotide polymorphisms (SNPs) and short (i.e. 5–25nt) InDels when compared to the reference genome sequence. Thus, a very-short read mapping algorithm like Bowtie1 (52,56) can only map ∼70% of genomic library reads, while other short read algorithms like BWA (57) and Bowtie2 (48) that accept longer reads can map >90% of genomic reads because they can accommodate these SNPs and InDels. Second, we found better reliability and simpler interpretations of novel transposon InDel patterns by only tracking single-end reads rather than considering paired-end reads. We found paired-end reads frequently generated InDel prediction artifacts from one of

the paired-ends mapping to a genomic region missed by RepeatMasker (a commonly applied algorithm for the prediction and masking of transposons in reference genomes (58)), and therefore being mis-interpreted as unique euchromatic sequence (39,40). We frequently spotted these potential artifacts in un-masked heterochromatic regions that have extreme transposon density, such as many portions within the Y and 4th chromosomes, telomeric and pericentromeric regions, and contigs in 'Chr#Het' or 'ChrU' ('Het' for heterochromatin, 'U' for unknown). When these unannotated regions were re-queried with BLAT, it was apparent that these unmasked regions were in fact repetitive sequences.

To overcome these obstacles, we developed a sensitive single-end, split-read transposon detection approach called TIDAL that leverages the strengths of different short read mapping tools. TIDAL first counts paired-end reads as two independent single-end reads, and culls the majority of reads which fully map to the reference genome and transposon consensus sequences. It then applies split-read mapping and validation procedures to the 'unmappable' reads (Figure 1A, detailed flow chart in Supplementary Figure S1). We reasoned that both transposon insertions and depletions should be accurately represented by split reads both upstream and downstream of the InDel breakpoints (Figure 1B and C). Bona fide split-reads representing the transposon InDel are then clustered together such that the genomic interval containing a cluster of split reads does not exceed a sequence window of 300 nucleotide (nt) (twice the length of the longest150 nt reads. This cluster also defines a clear window for counting reads representing the reference genome allele unaltered by a transposon InDel. By dividing the count of transposon InDel split-reads by the count of unaltered reference genome reads plus a pseudo-count of 1 in the interval defined by a cluster, we derived a relative metric called the CR, which approximates the frequency of the transposon InDel allele versus the reference genome allele of the same locus in the library. The importance of the CR is discussed further below in genomic PCR experiments and when comparing transposon landscapes between individual fly strains to pools of flies.

A second key improvement in TIDAL is the use of the BLAT algorithm (53) to validate that candidate reads are truly a split between a transposon-mapping end and a euchromatic genome end. BLAT is a sequence alignment program that yields a score which scales with different read lengths, and this score allowed us to address those genomic regions that were clearly repetitive yet were missed by RepeatMasker. Since a bona fide candidate split read cannot map to the genome in its entirety, we only retained reads whose ratios of the BLAT score to read length were $\leq$83%. This empirical threshold was found to be effective at clearing 92% of false positive reads such as degenerated transposon reads which can still map to multiple loci with BLAT. BLAT is slower compared to highly efficient short read aligners, so we perform BLAT alignments only for the relatively short list of candidate split reads first identified by Bowtie1 (Figure 1A). The importance of the BLAT algorithm is discussed further below when TIDAL is benchmarked against other transposon insertion algorithms applied to DGRP fly strains.

A final major improvement in TIDAL is generation of over 300 TLs based upon the latest *D. melanogaster* genome build, Release6/dm6 (see Supplementary Table S1 for accession numbers). Supplementary Figure S2 shows substantial changes in the genome build structure between the latest Release6/dm6 build and the previous Release5/dm3 build for which the bulk of modEncode (2,3) and previous transposon insertion programs were based upon (9,39–41). TIDAL outputs are accessible from a website database located at: http://www.bio.brandeis.edu/laulab/Tidal_Fly/Tidal_Fly_Home.html and will be maintained in future phases to incorporate additional *D. melanogaster* genomes (4,5). The outputs in TIDAL-Fly are grouped according to cell lines, common lab wild-type reference strains, DGRP flies, DGN flies and pools of flies from population genetics studies (Figure 1C).

Retrotransposons make up the bulk of TL diversity amongst *D. melanogaster* cell lines

We applied TIDAL to 21 *D. melanogaster* cell line genomes, one of which is new to this study (S2c1), some of which we and others had sequenced (33,34), and other lines from modEncode (2). Between ~800 to ~3000 novel transposable element (TE) insertions could be detected across cell lines, with LTR-retrotransposons making up the bulk of these new insertions, as they do in the reference genome (22) (Figure 2A). The composition of these LTR-retrotransposon insertions varied widely between cell lines, ranging from abundant *mdg1* insertions in lines originating from Oregon-R embryos (S3, S4, W2, Clone.8 and L1), *297* in several other lines, *gypsy* and *springer* in OSC lines, and an explosion of *ZAM* in one OSS line that was previously observed (33). In contrast, TE depletions were much more similar between cell line genomes, averaging at ~480 transposons that included Class I (retrotransposons) and Class II (DNA-cut-and-paste) transposons, with very consistent patterns of the same types of transposons being depleted (Figure 2B).

When comparing the genomic coordinates of TE insertions between cell lines, gratifyingly two S2 cell variants, S2c1 (this study) and S2R+ (from modENCODE (2)) shared the greatest proportions of the same TE insertions, whereas the other cell lines tended to have cell-line-specific TE insertions (Figure 2C). These cell-line-specific insertions may represent transposition events that occurred either in cell culture or in the original fly strains used to generate these cell lines. In contrast, the sites of TE depletions at the locus level also tended to show greater similarity between cell lines. The number of TE depletions discovered by TIDAL scaled with library read depth, but no correlation existed between the number of TE insertions and read depth (Figure 2D), perhaps reflecting the extensive copy number variation amongst various loci in cell lines (2) that could contribute to major fluctuations in TE insertion number. TE insertions in exons, which would likely disrupt protein function, are rare as they are in the reference genome (35) but have a higher than expected proportion in introns (Supplementary Figure S2B).

To evaluate the efficacy of TIDAL predictions with experimental validations, we focused on TE InDels for the

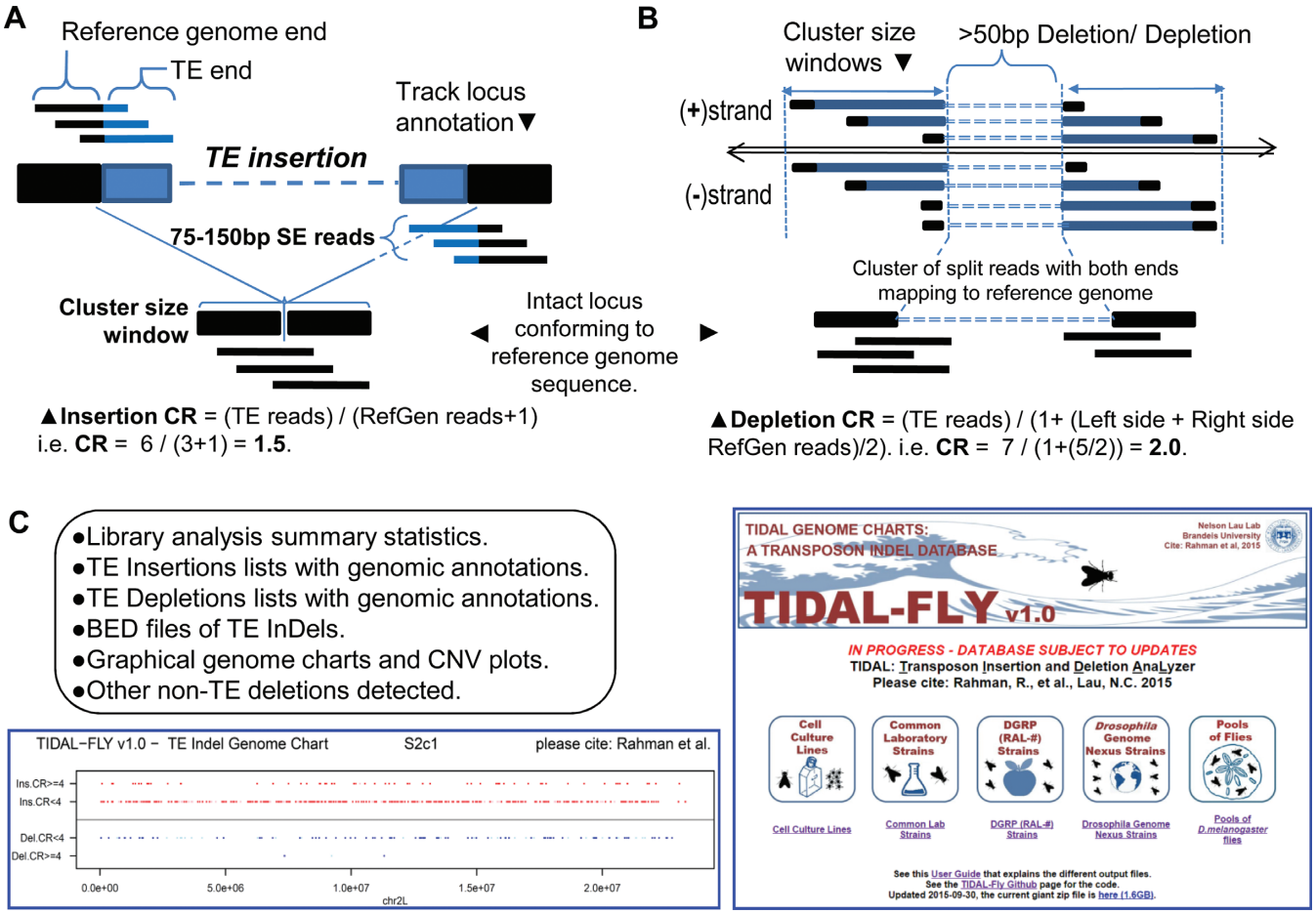


Figure 1. The design of the transposon insertion and depletion analyzer (TIDAL). (**A**) Diagram of the split-read approach for detecting transposon insertions, and (**B**) transposon depletions that include the calculation of a coverage ratio (CR) for each insertion and depletion. Detailed flowchart of the bioinformatics pipeline is shown in Supplementary Figure S1. (**C**) List of the output files accessible from the database, screenshot of the genome charts of transposon landscapes determined by TIDAL, and screenshot of the TIDAL-Fly database website homepage.

S2c1 cell line because we had enough genomic DNA remaining from initial library construction to also conduct genomic PCR on this sample. Our previous empirical studies showed that even a single primer designed for one end of a TE could lead to multiple amplicons that obfuscated the genomic PCR analyses (33). Therefore, we optimized a long-amplicon genomic PCR protocol using only primers mapped to the euchromatic genome and directly flanking the TE InDel. In addition to S2c1 cell genomic DNA, we included PCR tests of genomic DNA from ISO1 obtained from the BDSC (ISO1-BL) and RAL-362 fly strains, which were predicted to lack the TE InDels; these provided a set of controls for interpreting some results that were limited by inherent challenges with amplifying long multi-kilobase genomic amplicons (Supplementary Figure S3). In genomic DNA mixtures, shorter amplicons lacking the TE insertion will preferentially amplify over the rarer, longer amplicons in the PCR. Therefore, we also considered additional proxies for an insertion such as some large amplicons unable to electrophorese into the gel, and significant decreases of the short amplicons due to the insertion being too large to be amplified but still reflecting a reduced amount of short amplicon template. We arbitrarily selected 15 insertion and 10 depletion sites predicted by TIDAL with CR>3 in the S2c1 genomes, and ~66–70% of these predictions could be validated by the genomic PCR analyses. These data are in line with our previous estimate of an empirical false discovery rate of <12% from earlier comparisons of TIDAL outputs to PCR assays with genomic DNA from OSS and OSC cells (33). We attribute the differences between the current and previous analyses to the small sample sizes of PCR studies and other undetermined genomic DNA variations between the different cell lines.

Extensive TL diversity in common laboratory *D. melanogaster* strains

Next, we applied TIDAL to the genomes of sets of common laboratory 'wild-type' strains like Oregon-R (OreR), sequenced as part of modEncode and other projects (3,59), and Canton-S (CanS), used in a recent behavioral study (46) and from isogenized lines (60). We compared the TLs between five different OreR and six different CanS sub-strains isolated from different labs (Figure 3A, B), and discovered very different TLs among sub-strains of lab lines labeled with the same strain name. Whereas the OreR from TO2,

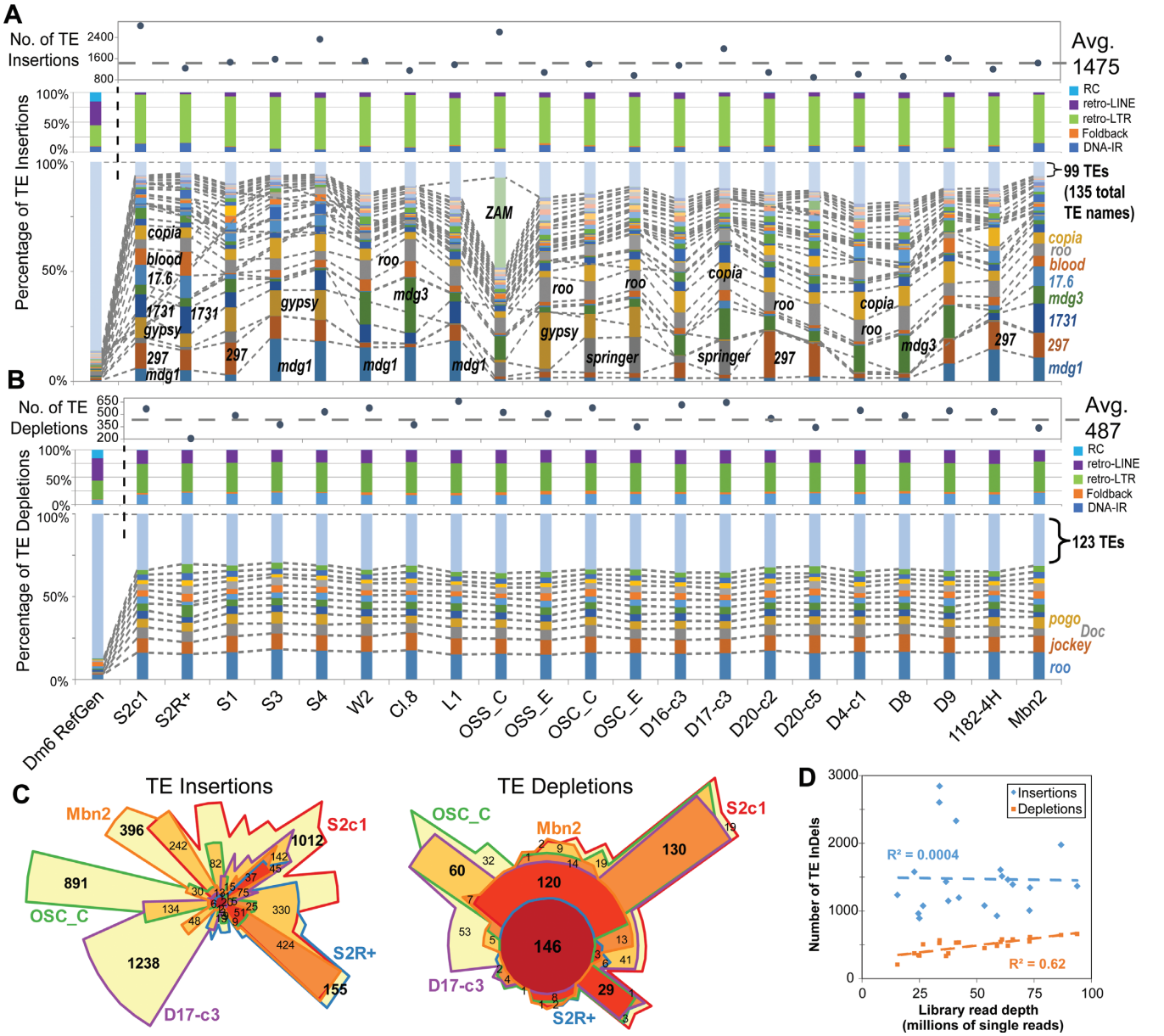


Figure 2. Transposon landscapes of *D. melanogaster* cell lines. (**A**) Profiles of transposon insertions, and (**B**) transposon depletions for 21 cell lines, grouped by their annotated strain origin (Supplementary Table S1A). The top panels show the total counts of transposon InDels, the middle panels show the proportions of each transposon class, and the bottom panels show the proportions of the TE families with ≥20 InDels (the rest are grouped together at the top). The dashed lines and labels in the bottom panels mark notable TE families. (**C**) Euler plots (area-proportional Venn diagrams) comparing between five representative cell lines the overlap of shared TE InDels based on their genomic coordinates. Bold numbers highlight the notably large numbers of cell-line specific TE insertions and broadly shared TE depletions. (**D**) Scatterplot comparing cell lines library read depth to TE InDels numbers.

SE, and BG labs shared many TE insertions, the OreR from PB1 and Dw1 had markedly distinct TE insertions. Likewise, all six CanS sub-strains also exhibited their own distinct TE insertion patterns. Differences among CanS sub-strains is not simply a reflection of TE InDels with low CR values, because a comparison of InDels with CR ≥3 still indicated highly distinct TLs (Supplementary Figure S4A-D). Most of the TE insertion differences between CanS sub-strains were represented by the *roo* retrotransposon (Figure 3C). However, in some of the OreR sub-strains, the *P-element* was a major factor in the TL differences. This observation suggests introgression of *P-element*-containing lab strains into certain stocks of OreR, which originally should be free of *P-elements* (61). Similar to cell lines, novel TE insertions in lab fly strains rarely occur in exons, but are over-represented introns and intergenic regions (Supplementary Figure S4E), which are genomic regions that can still affect gene expression in *D. melanogaster* (33,34).

In contrast to the diversity in TE insertion landscapes, the TE depletion landscapes amongst the lab strains was highly similar both in the genomic coordinates and TE family (Figure 3A, B and D). Similar to the greater homogene-

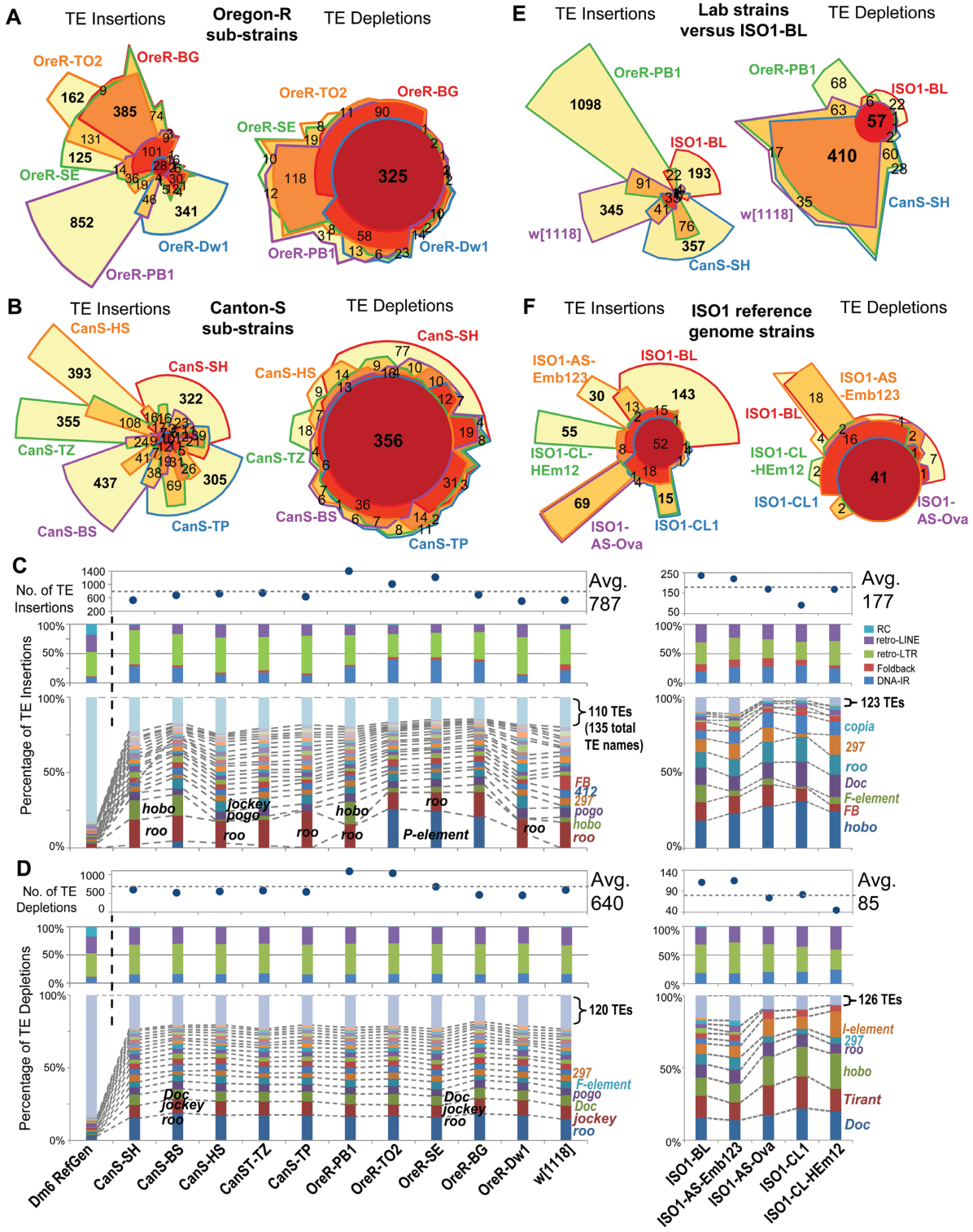


Figure 3. Transposon landscapes of common *D. melanogaster* lab strains. Euler plots comparing overlapping transposon InDels detected in Oregon-R (OreR) sub-strains (**A**) and Canton-S (CanS) sub-strains (**B**) from five different laboratories. Bold numbers highlight the notably large numbers of sub-strain-specific TE insertions and broadly shared TE depletions. Additional analyses of the CanS sub-strains are in Supplementary Figure S4. Profiles of transposon (**C**) insertions and (**D**) depletions for 11 lab strains (left set of panels) and five genomes from three sub-strains of the ISO1 strain (right panels, -BL: Bloomington, -AS: Spradling lab, -CL: Langley lab). In (**C**) and (**D**), the top panels show the total counts of transposon InDels, the middle panels show the proportions of each transposon class, and the bottom panels show the proportions of the TE families with ≥20 InDels (the rest are grouped together at the top). The dashed lines and labels in the bottom panels mark notable TE families. Euler plots comparing overlapping transposon InDels between the ISO1-BL strain and other lab strains (**E**) and other sub-strains of ISO1 (**F**). Bold numbers highlight the notably large numbers of strain-specific and sub-strain-specific TE insertions compared to broadly shared TE depletions. Additional analyses of the ISO1 sub-strains are in Supplementary Figure S5.

ity in TE depletion patterns amongst cell lines, we interpret these depletions as representing TE insertions in the ISO1 strain that was used for *D. melanogaster* reference genome sequence. Surprisingly, TIDAL analysis of genomic data from ISO1-BL, a recently sequenced isolate of the ISO1 strain from the BDSC (18), revealed an additional 236 and 111 novel TE insertions and depletions, respectively, with regards to the reference genome (Figure 3C and D). The number of TE InDels in ISO1-BL relative to the ISO1 reference genome is lower than other lab strains (Figure 3E), supporting its close relationship to the original isolate of ISO1 used for genome sequencing (15,16). We were able to validate several of these TE InDels with genomic PCR analyses using DNA from ISO1-BL compared to the ISO1 substrain from the Celniker lab (ISO1-UC) that was used for the reference genome (Supplementary Figure S3C and D).

Other ISO1 sub-strains from the Spradling (62) and Langley (63) labs also displayed ∼100–300 novel TE insertions (Figure 3F, Supplementary Figure S5), with each ISO1 substrain appearing to have its own unique TE landscape (Supplementary Figure S5A and B). In the Spradling lab samples, there were potential signatures of tissue-specific TLs (Supplementary Figure S5C). The most commonly mobilizing TE family in ISO1 sub-strains was *hobo* (Supplementary Figure S5D and E), supporting previous results that *hobo* is unstable in ISO1 sub-strains (64,65). These data indicate that TLs are much more diverse amongst common lab fly strains than previously appreciated, even for substrains of the ISO1 strain used for the *D. melanogaster* reference genome sequence. These results add additional considerations to previous studies using piRNA-seq and ChIP-seq to study transposon control (66–68), since it cannot be assumed that the TL of a lab strain is the same as the reference genome.

Several of the CanS sub-strains and the ISO1-BL strain contained many individual Illumina-sequencing runs that we initially combined into single libraries for TIDAL analysis. We analyzed different subsets of these runs as 'technical replicates' in order to conduct an internally-controlled computational experiment that evaluates TIDAL's performance as a function of sequencing depth. In both CanS and ISO1-BL libraries, there is a linear increase in computation processing time with increasing numbers of reads (Supplementary Figure S4F, S5F, S5G), with TIDAL run times ranging from ∼100 min/∼25M reads to ∼330 min/∼100M reads. The whole-genome alignments by Bowtie2 are the most computationally intensive steps in the TIDAL pipeline. Final TIDAL outputs as displayed on the TIDAL-Fly database are only <50MB, however the intermediate BAM and FASTQ files can put disk usage at ∼5–50GB per library. TIDAL consistently detected more TE insertions than TE depletions amongst both CanS and ISO1-BL individual Illumina-sequencing runs, reaching an asymptote of fewer new TE InDels with ever increasing library depth (Supplementary Figure S4F, S5F, S5G). These analyses suggest that for lab-maintained fly stocks, there are diminishing returns in TE InDel discovery using TIDAL for genomic sequencing beyond 100M reads, as most TL patterns are already well defined at this depth.

Benchmarking TIDAL on DGRP strain genomes

Most *D. melanogaster* genomic sequence libraries consist of a single read length. However, many DGRP strains have sequencing runs with different paired-end read lengths (i.e. 75nt, 95nt, 100nt, and 125nt long reads) that have been consolidated under single SRA entries (8,11). Since TIDAL's split-read approach requires a minimum of 50nt reads for TE InDel discovery, we asked whether read lengths impacted TE InDel determination. These DGRP libraries with multiple sequencing runs of different read lengths provided an excellent opportunity to assess this question in an internally-controlled experiment. Gratifyingly, the majority of TIDAL TE insertions calls in the shorter read library were identically represented in the longer read library for the same DGRP strain, with many new TE insertions only called in the longer-read library, which also tended to be sequenced at greater depth (Figure 4A). This result confirmed the reproducibility of the TIDAL outputs and the independent deep-sequencing runs. By plotting the total number of TE InDels as a function of sequencing coverage (read length*depth/genome size) across these DGRP libraries, we also observed a logarithmic trend in TE InDel discovery with greater genome sequencing coverage that was consistent with the lab fly strains analyses (Figure 4B). However, unlike cell lines or lab strains, the capacity to detect TE depletions in DGRP fly libraries was similar to the detection of TE insertions. By classifying the data points according to read length, we detect a trend suggesting library depth may have a greater influence on TE InDel detection sensitivity than read length (Figure 4B).

Song *et al.* (44) previously showed for a single DGRP strain (RAL-391) that there was relatively little agreement between three TE insertion prediction programs called: TEMP (39), a program by Cridland *et al.* (40), and a program by Linheiro and Bergman (41) (we refer to the two latter programs as CnT and LnB, respectively). We extended this observation by analyzing several DGRP libraries to benchmark TIDAL's TE insertion calls against the calls made by TEMP, CnT, and LnB programs. Because these programs had outputs of Release5/dm3 genomic coordinates for TE insertions, we first converted their predictions to 1-kb binned Release6/dm6 coordinates for comparison. We conducted overlap analysis on the TE insertion genomic coordinates, and observed a very low number of TE insertions called by all four programs (i.e. <∼30 common calls out of >∼1000 TE insertions) (Figure 4C and Supplementary Figure S6).

Indeed, a substantial number of common TE insertions calls were shared by only two programs (Figure 4C and Supplementary Figure S6A), and significant numbers of TE insertions were uniquely called by TIDAL, TEMP and CnT. The LnB program always had far fewer TE insertion calls compared to TIDAL, TEMP and CnT, because it was designed to optimize specificity over sensitivity (41). TIDAL consistently displayed the best overlap in common TE insertion calls with the LnB split-read method, more so than CnT and TEMP, which both use BWA as its main read-mapping algorithm and paired-end information. We found more commonality in TE depletion calls between TIDAL to TEMP than for TE insertion calls (Supplementary Fig-

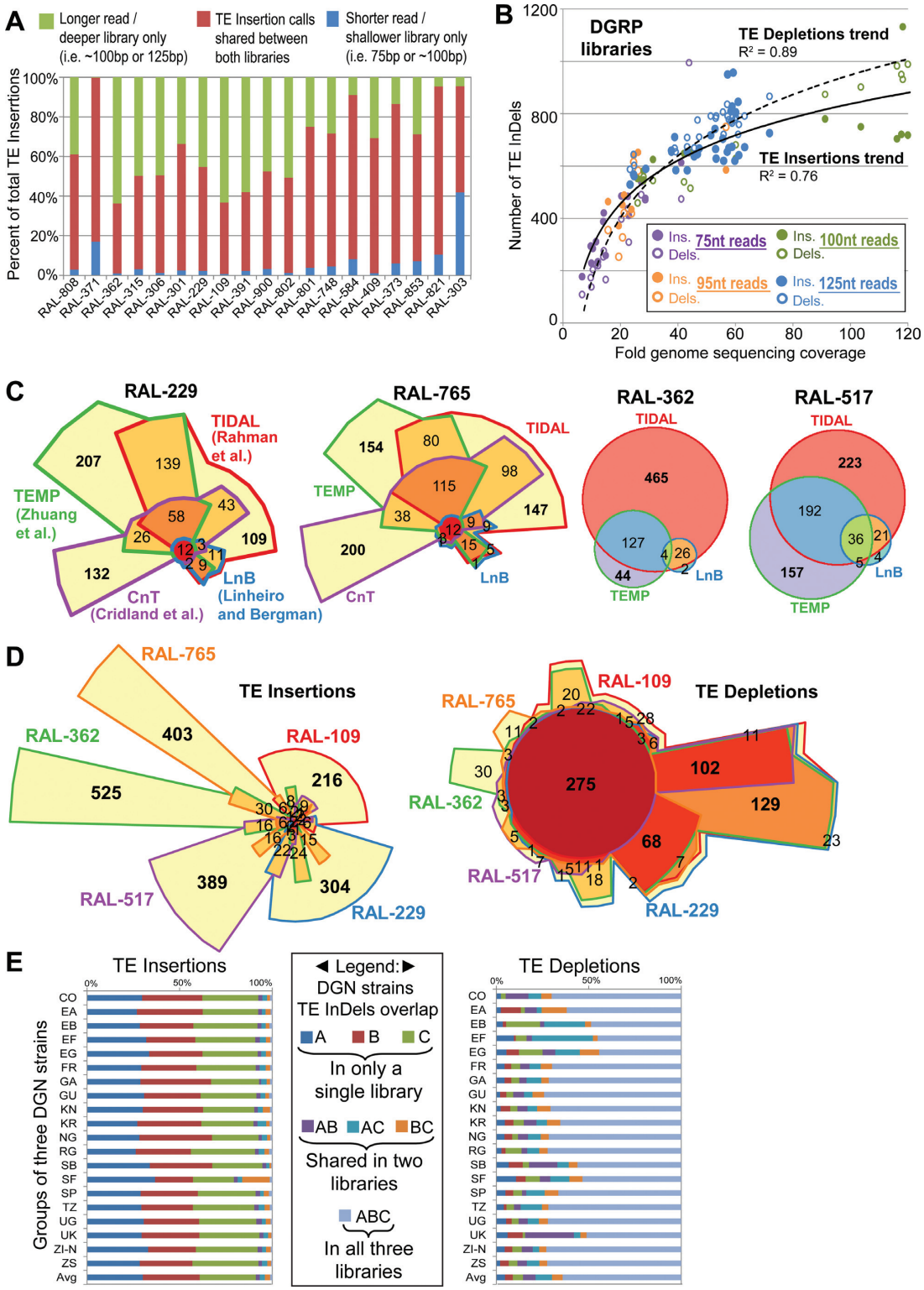


Figure 4. TIDAL analyses of DGRP and DGN fly strains. (**A**) Significant overlap in TE Insertions called by TIDAL for 19 DGRP strains that were sequenced twice by different Illumina genomic libraries of different lengths and depths. (**B**) Scatterplot comparing the fold genome sequencing coverage (read length*depth/genome size) to the number of TE InDels for 68 libraries from 23 DGRP strains. Logarithmic trend lines were fitted to the TE InDel points. (**C**) Euler plots comparing the genomic locus overlap in transposon insertion calls made by TIDAL and three other transposon insertion programs for four DGRP strains. The CnT program lacked any insertion predictions for RAL-362 and RAL-517. Bold numbers highlight the notably large numbers of program-specific TE insertion calls. (**D**) Euler plots comparing the TE InDels called by TIDAL for five DGRP strains. Bold numbers highlight the notably large numbers of strain-specific TE insertions and broadly shared TE depletions. (**E**) Comparisons of TE InDel calls within cohorts of three strains per geographic isolates of flies from the DGN (see Supplementary Table S1D). Across multiple geographic isolates, the vast majority of TE insertions are unique to a single strain within each 3-strain cohort, whereas TE depletions are broadly shared across DGN fly strains.

ure S6A). Neither the LnB nor CnT program call TE depletions.

To better understand why so many TE insertion calls in DGRP lines differed between the programs, we noticed that many TEMP and CnT output coordinates were in heterochromatic regions that are already dense with TEs, such as Chr4, ChrY, Chr2RHet and ChrU (Supplementary Figure S2, S6B), and these can be removed without affecting the overlapping calls between these programs and TIDAL (Supplementary Figure S6C). Closer inspection in the genome browser indicate that these heterochromatic regions are often not masked by RepeatMasker, yet TIDAL and LnB can avoid these regions because the pure split-read and BLAT approach can remove these reads from too high a BLAT mapping score. For 48 arbitrarily-selected sites with less-obvious reasons for discrepancy, we used PCR to evaluate TE insertion calls for the RAL-765 strains that were predicted only by TIDAL, by LnB (which overlapped with TIDAL calls), only by TEMP, and only by CnT (Supplementary Figure S7). Using the same criteria in the genomic PCR assay for the S2c1 cell line and ISO1 sub-strain data above, the number of PCR events supporting a TE insertion was approximately the same for TIDAL-only predictions (7/12) compared to TEMP-only (5/12) and CnT-only (6/12) predictions, while the agreement between TIDAL and LnB predictions was strongly supported very frequent detection of insertions by PCR (11/12). These data show that while no single method yet can identify all TE insertions in *D. melanogaster* genomes, the majority of TIDAL predictions are likely to be real insertions.

Widespread TE insertion diversity amongst wild fly strains across the globe

To address the question of whether the TL diversity observed in cell lines and lab strains was particularly large or small compared to naturally wild fly strains, we applied TIDAL to sets of *D. melanogaster* genomes in the DGRP, DGN, and pools of flies. For the first phase of this project, we analyzed the TLs in 57 and 70 strains in the DGRP and DGN resources, respectively, and 52 different pools of flies. We chose these particular strain libraries for general qualities such as high library read depth and widespread global distributions (Supplementary Table S1), and assumed these initial subsets would represent the greater trends of TLs that would apply to the rest of the strains.

We conducted overlap analyses of the TE InDel genome coordinates between DGRP fly strains to assess the similarity of TLs. A representative Euler diagram comparing five DGRP strains shows that the overwhelming majority of novel euchromatic TE insertions are unique to each strain (Figure 4D). In contrast, the TE depletions are frequently shared between DGRP strains. To examine these TL patterns globally, we conducted overlap analyses of DGN strains by comparing cohorts of three strains each from various geographically disparate regions. Similar to the DGRP flies, all the DGN 3-strain cohorts from the same region exhibited TE insertion patterns that were frequently unique to each strain, whereas TE depletion patterns were more frequently shared between all three strains (Figure 4E). These TL depletion patterns in DGRP and DGN fly strains are remarkably consistent with lab fly strains and even cell lines, reinforcing the notion that these actually represent TE insertions in the ISO1 reference genome sequence. Although there are a few individual fly strains with unique depletion patterns, commonality of the TE depletion patterns is the prevailing picture at all of the individual genomic sites (Figure 4E) or TE family level (Supplementary Figure S8). Thus, until we have additional *D. melanogaster* reference genome sequences as a basis for comparison, we believe the current TE depletion analyses cannot be accurately used to consider TE absence differences between genomes.

A minority of families make up the majority of TE insertion diversity

As a result of limited insight that can be gained by examining TE depletion patterns, we decided to focus our remaining attention to the TE insertion diversity amongst DGRP and DGN fly strains and pools of flies. Abridged plots (Figure 5A–C) that are representative of broader comparisons between strains and pools of flies (Supplementary Figure S9) reveal three new insights into *D. melanogaster* TL patterns: (i) The average number of novel euchromatic TE insertions in wild fly strains (∼550–670) from the DGRP and DGN is similar but somewhat lower than the average insertions inbred lab strains like OreR and CanS (∼750) and much lower than cell lines (>1400); (ii) compared to their ∼10% proportion of the total TEs in the reference genome (22), insertions of Class 2 DNA transposons are proportionally more abundant in both lab and wild fly strains (≥25%) relative to the Class 1 retrotransposons, which dominate in cell lines; and 3) in cell lines, lab and wild strains, a minor proportion (∼16–27%) of all characterized *D. melanogaster* TE families make up the bulk of the TE insertion diversity (≥75%).

Interestingly, the average number of TE insertions in pools of flies was frequently much greater (>2000, Figure 5C) than samples from DGRP, DGN and lab fly strains. We investigated this further by plotting the log10-transfrormed numbers of TE InDels relative to the library sequencing depth (fold genomic coverage, Figure 5D). Linear trends were fitted to each group of libraries, and the slopes for the DGRP, DGN and lab fly strains trend lines for TE InDels were all <1, indicative of the diminishing returns of TE InDel discovery with increasing read depths. However, the trend line for TE insertions in pools of flies is much higher (1.34) than the individual fly strains (0.12–0.6), suggesting that in pools of flies there is no obvious limit to discovering more TE insertions with greater depth in sequencing. This result is consistent with TE insertion landscapes varying widely among individual flies, which when pooled, lends to greater TE insertion landscape diversity in genomic libraries relative to individual strains.

If TE insertion landscape diversity is widespread between individual flies, we would predict that TE insertion reads in pools of flies would be diluted because a sample of individuals with many rare TE insertions would contribute more reads representing the unaltered reference genome sequence at any given loci. The CR for each TE InDel provides this relative measurement, and indeed the vast majority of TE insertions in pools of flies have CR <1.0, whereas the ma-

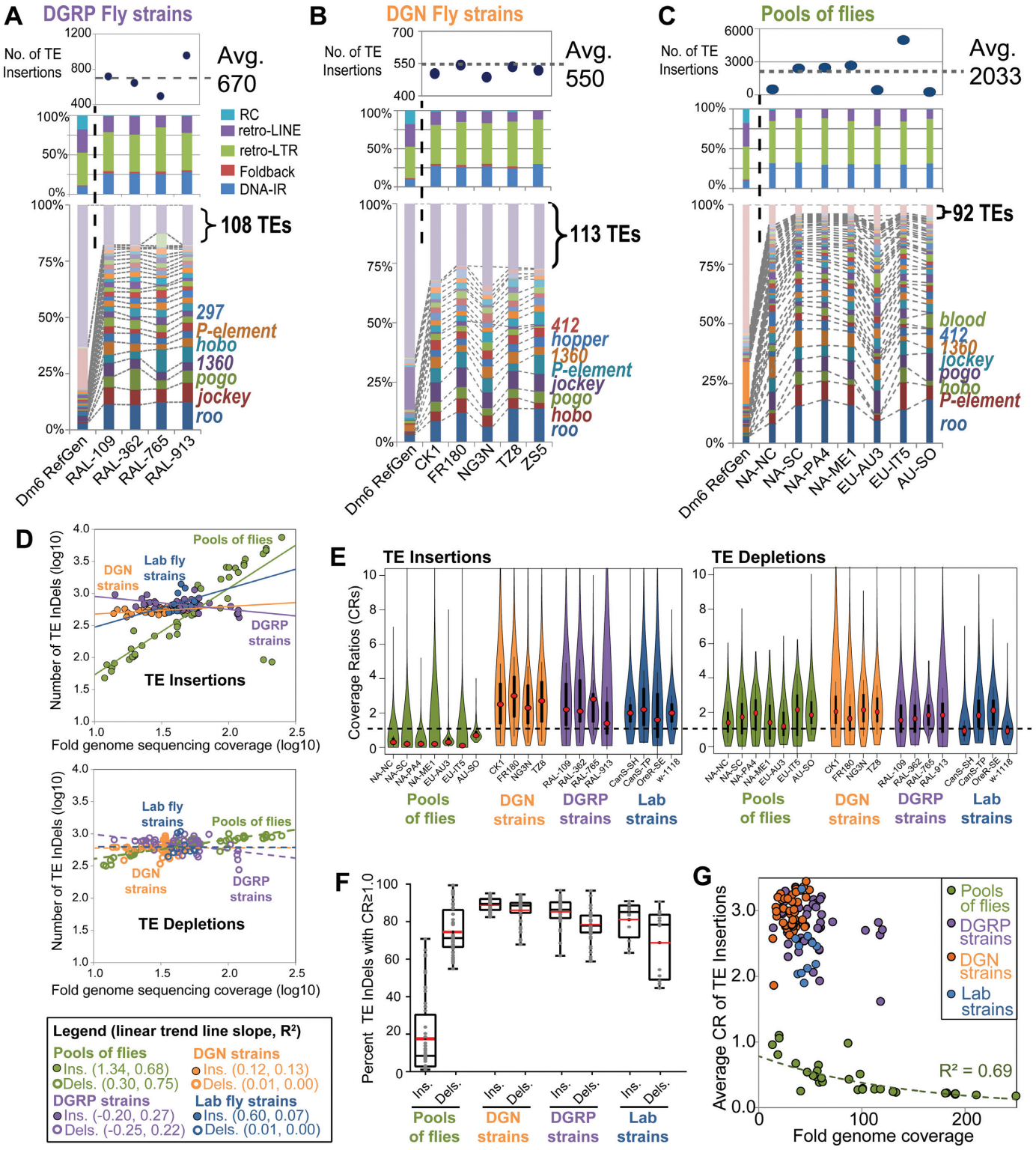


Figure 5. DGRP and DGN fly strains have distinct transposon landscape characteristics compared to pools of flies. Abridged profiles of transposon insertions for (**A**) DGRP fly strains, (**B**) DGN fly strains, and (**C**) Pools of flies. Detailed transposon insertion profiles for all analyzed libraries are shown in Supplementary Figure S9. The top panels show the total counts of transposon insertions, the middle panels show the proportions of each transposon class, and the bottom panels show the proportions of the TE families with ≥20 insertions (the rest are grouped together at the top). The dashed lines and labels in the bottom panels mark notable TE families. (**D**) Scatterplots comparing the fold genome sequencing coverage (reads*read length/genome size) to the number of TE InDels for the *D. melanogaster* libraries analyzed by TIDAL for this study. Linear trend lines were fitted to each group of insertions (top graph, solid lines) and depletions (bottom graph, dashed lines), with the slopes and R^2 values for each trend line listed in parenthesis in the legend. (**E**) Violin plots of the distribution of CRs for TE insertions and depletions from select sets of fly libraries. (**F**) Box plot of the distributions of percentages of TE InDels with CR≥1.0 for each group of fly strains and pools of flies. (**G**) Scatterplots of the same data points in the box plot, but plotted against the fold genome coverage to show a unique exponential decay trend for the CR for TE insertions from pools of flies.

jority of TE InDels in individual fly strains have CR ≥1.0 (Figure 5E and F). In fact, the mean CR for TE insertions across the different pools of flies is 0.45 ($N = 32$), whereas the mean CR for DGRP and DGN lines is 2.98 ($N = 46$) and 2.73 ($N = 44$), respectively. The difference in CR values between pools of flies and fly strains was only apparent in TE insertions but not TE depletions (Figure 5E), reinforcing the interpretation that TE depletions are unlikely to be absences from a single strain but rather insertions that are unique to the ISO1 reference. Additionally, TE insertion CRs in pools of flies follow an exponential decay trend with greater sequencing depth (Figure 5G), which we interpret as the effect of progressively diluting the representation of each TE insertion during deeper sampling of genomic diversity in pools of flies.

Abundance of TE families besides *roo* negatively correlates with piRNA expression

Our study confirms earlier findings that TE insertion patterns vary widely among *D. melanogaster* strains (30,65,69), but our TIDAL-Fly data also indicates that only a small subset of TE families make up the majority of this *D. melanogaster* TE insertion landscape diversity. How does this TE insertion diversity relate to host TE suppression mechanisms such as the Piwi/piRNA pathway? A previous study sequenced piRNA libraries from 16 DGRP strains, but their analysis, which relied on earlier predictions of TE insertions, was unable to detect a significant correlation across strains between piRNA levels and the number of novel TE insertions for any TE family (44). We re-analyzed these and other previously published piRNA datasets (33,44) with a different approach that compared normalized proportions of piRNA counts for all TE families to the proportions of novel TE insertions predicted by TIDAL within a given cell line or strain. By measuring all piRNAs mapped directly to TE consensus sequences without normalization to the reference genome sequence, we were then able to correlate these proportions of TE-directed piRNAs with the proportions of TE insertions across all TE families within each *D. melanogaster* OSS/OSC cell lines (Figure 6A) and DGRP strain (Figure 6B). This approach quite effectively normalizes the variation in absolute counts of piRNAs and TE insertions, which can differ greatly between different small RNA and genomic DNA libraries.

In both *D. melanogaster* cell lines that express piRNAs and in 15 DGRP strains (1 strain lacked Illumina genomic sequences), the TE families with the greatest proportions of TE-directed piRNAs were in the lowest proportions of novel TE insertions (Figure 6). This relationship is consistent with the mechanism that abundant Piwi/piRNA complexes engage TE transcripts to direct silencing (33,34,70–72). Accordingly, many TE families with large proportions of insertions tended to have lower proportions of piRNAs directed at these families. However, in all the DGRP strains, the *roo* retrotransposon stands out as the sole TE family with large numbers of novel insertions and high proportions of piRNAs directed against it. Although we do not have a mechanistic explanation for *roo's* exceptionally high number of new insertions and high levels of *roo* piRNAs in DGRP flies, this result may be related to *roo's* high transposition rate (73,74) and possible horizontal transfer of *roo* between different Drosophilid species (75).

DISCUSSION

We report a new tool optimized for accurate determination of *D. melanogaster* TLs and demonstrate the TIDAL-Fly database's value as resource for the genome-wide analysis of TLs in *D. melanogaster*. In addition to discovering how different genomic library characteristics can impact TLs, our analysis has also uncovered notable biological and genomic insights regarding TLs in the *D. melanogaster* genome. Cell lines differ widely in their TLs, with transposon insertion numbers showing weak correlation with sequencing depth (Figure 2D), perhaps reflecting the frequent aneuploidy and/or polyploidy in cell cultures that may allow transposon insertion numbers to proliferate in less predictable manners (2,33). In contrast, lab and wild fly strains have more defined limits of new transposon insertions, with each strain carrying several hundred new TE insertions. Although additional insertions can be detected with greater sequencing depth, these new insertions tend to have low (<1.0) CRs, indicating that these TE copies are rarer amongst all the genomes sampled in each library. Common lab strains, and even sub-strains with the same lab strain name, also differ in their TLs, which in some cases is almost certainly due to contamination or introgression (i.e. the introduction of the *P-element* in some OreR strains, Figure 3C). Finally, pooled fly samples typically reveal many more TE insertions relative to individual lab or wild strains, with additional sequencing coverage being able to identify more insertions in pooled samples.

The CRs that TIDAL determines can inform which TE insertions can be detected by genomic PCR, because it reflects the 'penetrance' of each transposon InDel within each genomic library (Figure 1B and C). However, the striking difference of CR distributions between fly strains and pools of flies also suggest a need to reexamine how TLs should be interpreted if they are derived from pools of flies versus collecting various individual strains. In fly strains, TE InDels with very high CRs (>4) may likely represent germline transposition events in past generations that are present in most cells in the organism. However, the meaning of the bulk of insertions with CR values <4.0 (Figure 5E) is less clear, and some proportion of these events may reflect somatic insertions. Future sequencing experiments with individual flies and comparisons of the soma to germline is needed to test if somatic transposition events are perhaps more prevalent than appreciated. Despite the economy of sequencing pools of flies compared to the more demanding collection and sequencing of many individual fly strains (9), we suggest caution in evaluating TL differences between pools of flies without considering CRs because differences in allele frequencies may confound interpretation of TE location or abundance. For example, each of the many TE insertions in pooled fly samples that have low CRs might be germline events within a few individual flies that are then diluted by the pervasive diversity of unique TLs from other flies within the pool. Alternatively, if somatic transposition is common, the interpretation of low CRs insertions from

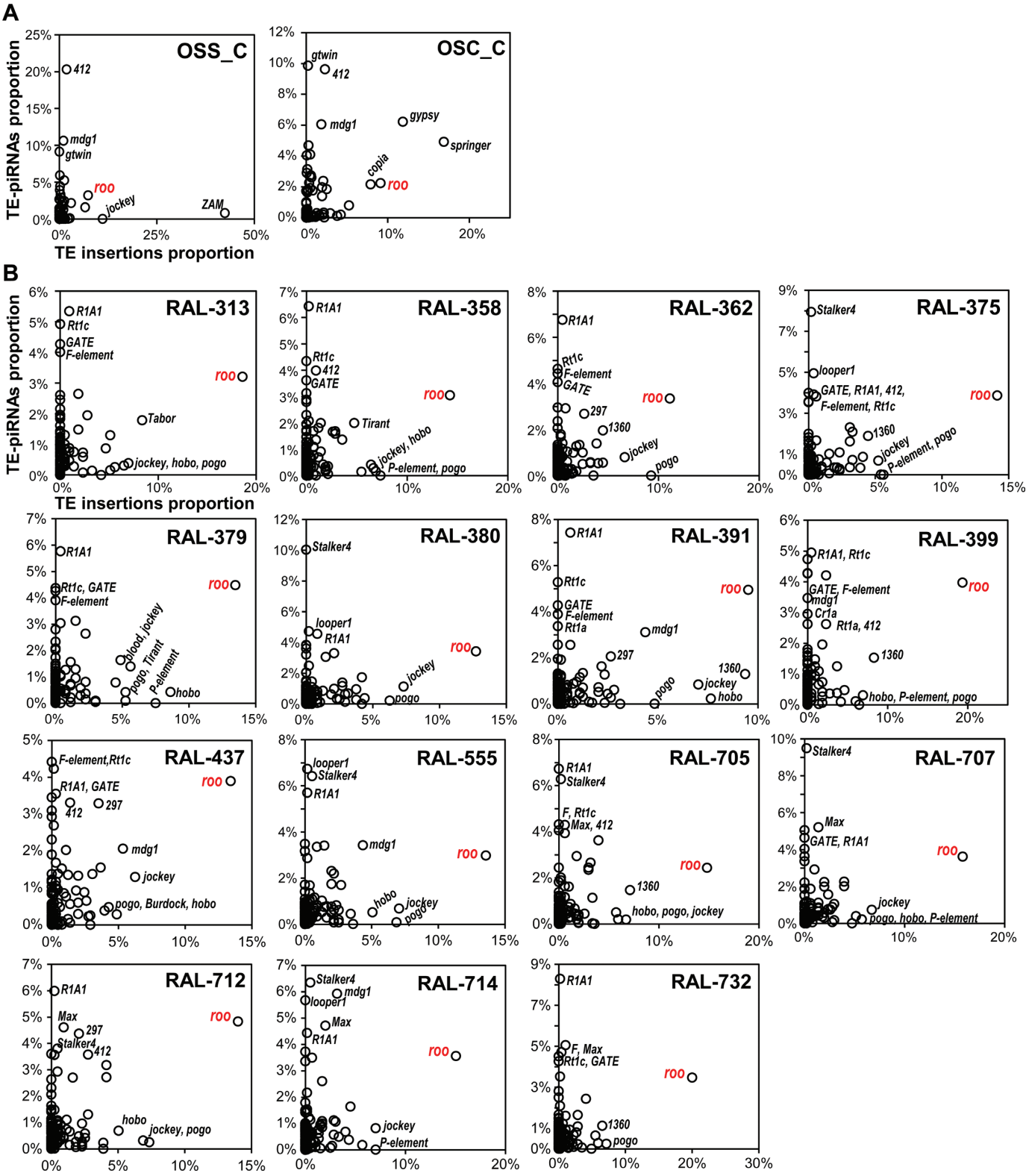


Figure 6. Comparisons of TE-directed piRNAs versus TE Insertions for OSS/OSC cells and DGRP fly strains. Scatterplots compare the relative proportions of TE-directed piRNAs and TE insertions for each sample in (**A**) OSS and OSC cells and (**B**) DGRP fly strains with sequenced piRNAs and genomes sequenced by Illumina. The *roo* transposon is highlighted for its exceptionally high proportion of strain-specific insertions despite the high proportions of *roo*-directed piRNAs.

pooled samples would still be challenging since somatic insertions could be conflated with low frequency alleles.

Recently, genomic mosaicism was observed between *D. melanogaster* follicle cells and salivary glands (62), and transposons have been described as sources of genome variation amongst individual neurons (76,77). With TIDAL, we are better poised to further explore these and other phenomena in *D. melanogaster*. For example, an emerging hypothesis for animal aging proposes the progressive incapacity of older cells to reign in transposons (78), and several lines of evidence in *D. melanogaster* suggest aging animals display signatures of heterochromatin mark alterations that would be molecularly indicative of TL expansion (77,79). Furthermore, some transposon insertions have been reported to show signatures of adaptation (31,32,80), and TIDAL could enable deeper scrutiny of individual strains possessing the most penetrant TE insertions near genes in these fly strains.

TIDAL has the limitation of not being able to analyze new transposon InDels in TE-dense heterochromatic regions, including large intergenic piRNA clusters that serve as genetic traps for new TE insertions that then prevent future transposon mobilization events (81–83). This limitation is tied to the short length of Illumina reads that cannot be mapped uniquely in repetitive regions nor span the lengths of full transposons. Extremely-long read (>20 kb) sequencing technologies like the PacBio system may pave the way for future routine sequencing efforts of *D. melanogaster* genomes when its costs can approach the Illumina platform. For example, the ISO1 strain's genome was recently re-sequenced and completely re-assembled *de novo* from PacBio runs, with impressive closure of previous genome gaps that included transposons (84,85). Our data suggest that frequent transposon depletions overlapping between disparate *D. melanogaster* genomes may actually reflect specific TE insertions to the ISO1 sub-strain used for the reference genome. We therefore propose that the field could benefit from PacBio sequencing of new *D. melanogaster* reference genomes such as population-specific type strains or commonly-used cell lines to provide additional reference states that would improve our understanding of the transposon architecture in these genomes.

Our bioinformatics results also allow us to compare genome-wide data against the previous classic studies of transposon accumulation and transposition rates in *D. melanogaster.* Related lab stocks have been previously shown to display distinct restriction length polymorphisms from transposon probes (86), while other lab lines undergoing mutation accumulation have been examined with low-resolution approaches like in situ hybridization of individual TE families in polytene chromosomes to estimate transposition rates (64,73,74,87–97). These studies described transposition rates ranging from $\sim 10^{-4}$ to 10^{-5} per TE copy per generation for several TE families. Making an assumption of ~15 years separating the ISO1-BL strain from the original ISO1 strain whose DNA comprises the reference genome sequence (18), and using a 12-day average *Drosophila* generation time, our TIDAL outputs derived transposition rates for active families that are on the same order of magnitude as previous studies (Supplementary Figure S5H). Finally, Nuzhdin et al. predicted that inbreeding flies would foster increases in transposition (87,98). TIDAL outputs also support this prediction by revealing higher average numbers of TE insertions in lab strains inbred for decades (~790), versus DGRP fly strains which have only recently been inbred (~670), and these are also higher than outbred DGN fly strains (~550).

Understanding rates of transposition in other genetic and environmental contexts remains an open question, and this can now be studied accurately using high-resolution genome sequencing and TIDAL. For example, it is now appreciated that the Piwi/piRNA pathway acting in the animal germline may strongly modulate the evolution of TLs (88,99–101), to a point where TL diversity might be an expected outcome of the Piwi/piRNA pathway by preventing detrimental explosions of transposition, but allowing some low level of transposition to occur. If wild *D. melanogaster* are constantly allowing low levels of transposition activity (102), this activity can be intensified by various natural mechanisms such as horizontal transfer (103–106), stressful environmental conditions (107,108), permissive alleles in the host (109), or crosses between distant strains (110). In addition, our comparisons of TE insertions versus piRNA content in DGRP strains suggest that *roo* may be an example of a transposon that has been able to evade Piwi/piRNA silencing by unknown mechanisms (Figure 6B). Our future interests will be to re-examine the rates of transpositions by genome sequencing and TIDAL analysis of direct lineages of flies subjected to stress and aging, and in sub-fertile mutants of the Piwi pathway. With this high-resolution genomics approach, we will better understand the détente between genomes and the genetic parasites that reside within them.

NOTE ADDED IN PROOF

In a very recent update of the TEMP program applied to two DGRP libraries (RAL-109 and RAL-229) and the Release 6 genome, there was greater overlap in TE insertion calls between TEMP and LnB and TIDAL. However, many program-specific calls remain, requiring future continuing development of all TE discovery programs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dianne Schwarz for manuscript comments and Michael Marr for suggesting the TIDAL acronym. We also thank Alex Ferrazoli and Susan Lovett for providing access to the Illumina MiSeq. Finally, we thank the University of Manchester Genomic Technologies Core Facility for assistance with genome sequencing. NCL was a Searle Scholar.

FUNDING

Human Frontier Science Program [RGY0093/2012 to C.M.B.]; National Institutes of Health [R00HD057298 to N.C.L.]; Searle Scholars Foundation to N.C.L. Funding for open access charge: Brandeis University Open Access Fund.

Conflict of interest statement. None declared.

REFERENCES

1. Levin,H.L. and Moran,J.V. (2011) Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.*, **12**, 615–627.
2. Lee,H., McManus,C.J., Cho,D.Y., Eaton,M., Renda,F., Somma,M.P., Cherbas,L., May,G., Powell,S., Zhang,D. *et al.* (2014) DNA copy number evolution in Drosophila cell lines. *Genome Biol.*, **15**, R70.
3. mod,E.C., Roy,S., Ernst,J., Kharchenko,P.V., Kheradpour,P., Negre,N., Eaton,M.L., Landolin,J.M., Bristow,C.A., Ma,L. *et al.* (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science*, **330**, 1787–1797.
4. Bergman,C.M. and Haddrill,P.R. (2015) Strain-specific and pooled genome sequences for populations of Drosophila melanogaster from three continents. *F1000Res*, **4**, 31.
5. Grenier,J.K., Arguello,J.R., Moreira,M.C., Gottipati,S., Mohammed,J., Hackett,S.R., Boughton,R., Greenberg,A.J. and Clark,A.G. (2015) Global diversity lines – a five-continent reference panel of sequenced Drosophila melanogaster strains. *G3 (Bethesda)*, **5**, 593–603.
6. Bastide,H., Betancourt,A., Nolte,V., Tobler,R., Stobe,P., Futschik,A. and Schlotterer,C. (2013) A genome-wide, fine-scale map of natural pigmentation variation in Drosophila melanogaster. *PLoS Genet.*, **9**, e1003534.
7. Bergland,A.O., Behrman,E.L., O'Brien,K.R., Schmidt,P.S. and Petrov,D.A. (2014) Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLoS Genet.*, **10**, e1004775.
8. Huang,W., Massouras,A., Inoue,Y., Peiffer,J., Ramia,M., Tarone,A.M., Turlapati,L., Zichner,T., Zhu,D., Lyman,R.F. *et al.* (2014) Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Res.*, **24**, 1193–1208.
9. Kofler,R., Betancourt,A.J. and Schlotterer,C. (2012) Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. *PLoS Genet.*, **8**, e1002487.
10. Lack,J.B., Cardeno,C.M., Crepeau,M.W., Taylor,W., Corbett-Detig,R.B., Stevens,K.A., Langley,C.H. and Pool,J.E. (2015) The Drosophila genome nexus: a population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. *Genetics*, **199**, 1229–1241.
11. Mackay,T.F., Richards,S., Stone,E.A., Barbadilla,A., Ayroles,J.F., Zhu,D., Casillas,S., Han,Y., Magwire,M.M., Cridland,J.M. *et al.* (2012) The Drosophila melanogaster Genetic Reference Panel. *Nature*, **482**, 173–178.
12. Pool,J.E., Corbett-Detig,R.B., Sugino,R.P., Stevens,K.A., Cardeno,C.M., Crepeau,M.W., Duchen,P., Emerson,J.J., Saelao,P., Begun,D.J. *et al.* (2012) Population Genomics of sub-saharan Drosophila melanogaster: African diversity and non-African admixture. *PLoS Genet.*, **8**, e1003080.
13. Reinhardt,J.A., Kolaczkowski,B., Jones,C.D., Begun,D.J. and Kern,A.D. (2014) Parallel geographic variation in Drosophila melanogaster. *Genetics*, **197**, 361–373.
14. Zichner,T., Garfield,D.A., Rausch,T., Stutz,A.M., Cannavo,E., Braun,M., Furlong,E.E. and Korbel,J.O. (2013) Impact of genomic structural variation in Drosophila melanogaster based on population-scale sequencing. *Genome Res.*, **23**, 568–579.
15. Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F. *et al.* (2000) The genome sequence of Drosophila melanogaster. *Science*, **287**, 2185–2195.
16. Myers,E.W., Sutton,G.G., Delcher,A.L., Dew,I.M., Fasulo,D.P., Flanigan,M.J., Kravitz,S.A., Mobarry,C.M., Reinert,K.H., Remington,K.A. *et al.* (2000) A whole-genome assembly of Drosophila. *Science*, **287**, 2196–2204.
17. Brizuela,B.J., Elfring,L., Ballard,J., Tamkun,J.W. and Kennison,J.A. (1994) Genetic analysis of the brahma gene of Drosophila melanogaster and polytene chromosome subdivisions 72AB. *Genetics*, **137**, 803–813.
18. Gutzwiller,F., Carmo,C.R., Miller,D.E., Rice,D.W., Newton,I.L., Hawley,R.S., Teixeira,L. and Bergman,C.M. (2015) Dynamics of Wolbachia pipientis gene expression across the Drosophila melanogaster life cycle. *G3 (Bethesda)*, doi:10.1534/g3.115.021931.
19. Smith,C.D., Shu,S., Mungall,C.J. and Karpen,G.H. (2007) The Release 5.1 annotation of Drosophila melanogaster heterochromatin. *Science*, **316**, 1586–1591.
20. dos Santos,G., Schroeder,A.J., Goodman,J.L., Strelets,V.B., Crosby,M.A., Thurmond,J., Emmert,D.B., Gelbart,W.M. and FlyBase,C. (2015) FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.*, **43**, D690–D697.
21. Bergman,C.M., Quesneville,H., Anxolabehere,D. and Ashburner,M. (2006) Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. *Genome Biol.*, **7**, R112.
22. Kaminker,J.S., Bergman,C.M., Kronmiller,B., Carlson,J., Svirskas,R., Patel,S., Frise,E., Wheeler,D.A., Lewis,S.E., Rubin,G.M. *et al.* (2002) The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biol.*, **3**, RESEARCH0084.
23. Elgin,S.C. and Reuter,G. (2013) Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harbor Perspect. Biol.*, **5**, a017780.
24. Chow,J.C. and Heard,E. (2010) Nuclear organization and dosage compensation. *Cold Spring Harbor Perspect. Biol.*, **2**, a000604.
25. Schoeftner,S. and Blasco,M.A. (2009) A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. *EMBO J.*, **28**, 2323–2336.
26. Khurana,J.S., Xu,J., Weng,Z. and Theurkauf,W.E. (2011) Distinct functions for the Drosophila piRNA pathway in genome maintenance and telomere protection. *PLoS Genet.*, **6**, e1001246.
27. George,J.A., DeBaryshe,P.G., Traverse,K.L., Celniker,S.E. and Pardue,M.L. (2006) Genomic organization of the Drosophila telomere retrotransposable elements. *Genome Res.*, **16**, 1231–1240.
28. Riddle,N.C., Shaffer,C.D. and Elgin,S.C. (2009) A lot about a little dot - lessons learned from Drosophila melanogaster chromosome 4. *Biochem. Cell. Biol.*, **87**, 229–241.
29. Ashburner,M. and Bergman,C.M. (2005) Drosophila melanogaster: a case study of a model genomic sequence and its consequences. *Genome Res.*, **15**, 1661–1667.
30. Cridland,J.M., Thornton,K.R. and Long,A.D. (2015) Gene expression variation in Drosophila melanogaster due to rare transposable element insertion alleles of large effect. *Genetics*, **199**, 85–93.
31. Mateo,L., Ullastres,A. and Gonzalez,J. (2014) A transposable element insertion confers xenobiotic resistance in Drosophila. *PLoS Genet.*, **10**, e1004560.
32. Guio,L., Barron,M.G. and Gonzalez,J. (2014) The transposable element Bari-Jheh mediates oxidative stress response in Drosophila. *Mol. Ecol.*, **23**, 2020–2030.
33. Sytnikova,Y.A., Rahman,R., Chirn,G.W., Clark,J.P. and Lau,N.C. (2014) Transposable element dynamics and PIWI regulation impacts lncRNA and gene expression diversity in Drosophila ovarian cell cultures. *Genome Res.*, **24**, 1977–1990.
34. Sienski,G., Donertas,D. and Brennecke,J. (2012) Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell*, **151**, 964–980.
35. Lipatov,M., Lenkov,K., Petrov,D.A. and Bergman,C.M. (2005) Paucity of chimeric gene-transposable element transcripts in the Drosophila melanogaster genome. *BMC Biol.*, **3**, 24.
36. Hoskins,R.A., Carlson,J.W., Wan,K.H., Park,S., Mendez,I., Galle,S.E., Booth,B.W., Pfeiffer,B.D., George,R.A., Svirskas,R. *et al.* (2015) The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Res.*, **25**, 445–458.
37. Smit,A.F., Toth,G., Riggs,A.D. and Jurka,J. (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.*, **246**, 401–417.
38. Thomas,J., Vadnagara,K. and Pritham,E.J. (2014) DINE-1, the highest copy number repeats in Drosophila melanogaster are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). *Mob. DNA*, **5**, 18.
39. Zhuang,J., Wang,J., Theurkauf,W. and Weng,Z. (2014) TEMP: a computational method for analyzing transposable element polymorphism in populations. *Nucleic Acids Res.*, **42**, 6826–6838.

40. Cridland,J.M., Macdonald,S.J., Long,A.D. and Thornton,K.R. (2013) Abundance and distribution of transposable elements in two Drosophila QTL mapping resources. *Mol. Biol. Evol.*, **30**, 2311–2327.
41. Linheiro,R.S. and Bergman,C.M. (2012) Whole genome resequencing reveals natural target site preferences of transposable elements in Drosophila melanogaster. *PLoS One*, **7**, e30008.
42. Fiston-Lavier,A.S., Barron,M.G., Petrov,D.A. and Gonzalez,J. (2015) T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data. *Nucleic Acids Res.*, **43**, e22.
43. Fiston-Lavier,A.S., Carrigan,M., Petrov,D.A. and Gonzalez,J. (2011) T-lex: a program for fast and accurate assessment of transposable element presence using next-generation sequencing data. *Nucleic Acids Res.*, **39**, e36.
44. Song,J., Liu,J., Schnakenberg,S.L., Ha,H., Xing,J. and Chen,K.C. (2014) Variation in piRNA and transposable element content in strains of Drosophila melanogaster. *Genome Biol. Evol.*, **6**, 2786–2798.
45. Ensminger,A.W., Yassin,Y., Miron,A. and Isberg,R.R. (2012) Experimental evolution of Legionella pneumophila in mouse macrophages leads to strains with altered determinants of environmental survival. *PLoS Pathog.*, **8**, e1002731.
46. Colomb,J. and Brembs,B. (2014) Sub-strains of Drosophila Canton-S differ markedly in their locomotor behavior. *F1000Res*, **3**, 176.
47. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
48. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
49. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
50. Boeva,V., Popova,T., Bleakley,K., Chiche,P., Cappo,J., Schleiermacher,G., Janoueix-Lerosey,I., Delattre,O. and Barillot,E. (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, **28**, 423–425.
51. Kapitonov,V.V. and Jurka,J. (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.*, **9**, 411–412.
52. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
53. Kent,W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
54. Quinlan,A.R. (2014) BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics*, **47**, 11–34.
55. Fujita,P.A., Rhead,B., Zweig,A.S., Hinrichs,A.S., Karolchik,D., Cline,M.S., Goldman,M., Barber,G.P., Clawson,H., Coelho,A. *et al.* (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.*, **39**, D876–D882.
56. Langmead,B. (2010) Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinformatics*, Chapter 11, Unit 11.17.
57. Li,H. and Durbin,R. (2010) Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*, **26**, 589–595.
58. Smit,A.F. (1993) Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.*, **21**, 1863–1872.
59. Chandler,C.H., Chari,S., Tack,D. and Dworkin,I. (2014) Causes and consequences of genetic background effects illuminated by integrative genomic analysis. *Genetics*, **196**, 1321–1336.
60. Miller,D.E., Takeo,S., Nandanan,K., Paulson,A., Gogol,M.M., Noll,A.C., Perera,A.G., Walton,K.N., Gilliland,W.D., Li,H. *et al.* (2012) A whole-chromosome analysis of Meiotic Recombination in Drosophila melanogaster. *G3 (Bethesda)*, **2**, 249–260.
61. Robertson,H.M. and Engels,W.R. (1989) Modified P elements that mimic the P cytotype in Drosophila melanogaster. *Genetics*, **123**, 815–824.
62. Yarosh,W. and Spradling,A.C. (2014) Incomplete replication generates somatic DNA alterations within Drosophila polytene salivary gland cells. *Genes Dev.*, **28**, 1840–1855.
63. Langley,C.H., Crepeau,M., Cardeno,C., Corbett-Detig,R. and Stevens,K. (2011) Circumventing heterozygosity: sequencing the amplified genome of a single haploid Drosophila melanogaster embryo. *Genetics*, **188**, 239–246.
64. Zakharenko,L.P., Kovalenko,L.V. and Mai,S. (2007) Fluorescence in situ hybridization analysis of hobo, mdg1 and Dm412 transposable elements reveals genomic instability following the Drosophila melanogaster genome sequencing. *Heredity (Edinb.)*, **99**, 525–530.
65. Moschetti,R., Dimitri,P., Caizzi,R. and Junakovic,N. (2010) Genomic instability of I elements of Drosophila melanogaster in absence of dysgenic crosses. *PLoS One*, **5**, e13142.
66. Shpiz,S., Ryazansky,S., Olovnikov,I., Abramov,Y. and Kalmykova,A. (2014) Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the drosophila germline. *PLoS Genet.*, **10**, e1004138.
67. Marinov,G.K., Wang,J., Handler,D., Wold,B.J., Weng,Z., Hannon,G.J., Aravin,A.A., Zamore,P.D., Brennecke,J. and Toth,K.F. (2015) Pitfalls of mapping high-throughput sequencing data to repetitive sequences: Piwi's genomic targets still not identified. *Dev. Cell*, **32**, 765–771.
68. Lin,H., Chen,M., Kundaje,A., Valouev,A., Yin,H., Liu,N., Neuenkirchen,N., Zhong,M. and Snyder,M. (2015) Reassessment of Piwi binding to the genome and Piwi impact on RNA polymerase II distribution. *Dev. Cell*, **32**, 772–774.
69. Charlesworth,B. and Langley,C.H. (1989) The population genetics of Drosophila transposable elements. *Annu. Rev. Genet.*, **23**, 251–287.
70. Rozhkov,N.V., Hammell,M. and Hannon,G.J. (2013) Multiple roles for Piwi in silencing Drosophila transposons. *Genes Dev.*, **27**, 400–412.
71. Saito,K., Nishida,K.M., Mori,T., Kawamura,Y., Miyoshi,K., Nagami,T., Siomi,H. and Siomi,M.C. (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev.*, **20**, 2214–2222.
72. Vagin,V.V., Sigova,A., Li,C., Seitz,H., Gvozdev,V. and Zamore,P.D. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, **313**, 320–324.
73. Diaz-Gonzalez,J., Dominguez,A. and Albornoz,J. (2010) Genomic distribution of retrotransposons 297, 1731, copia, mdg1 and roo in the Drosophila melanogaster species subgroup. *Genetica*, **138**, 579–586.
74. Diaz-Gonzalez,J., Vazquez,J.F., Albornoz,J. and Dominguez,A. (2011) Long-term evolution of the roo transposable element copy number in mutation accumulation lines of Drosophila melanogaster. *Genet Res. (Camb.)*, **93**, 181–187.
75. de la Chaux,N. and Wagner,A. (2009) Evolutionary dynamics of the LTR retrotransposons roo and rooA inferred from twelve complete Drosophila genomes. *BMC Evolut. Biol.*, **9**, 205.
76. Erwin,J.A., Marchetto,M.C. and Gage,F.H. (2014) Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat. Rev. Neurosci.*, **15**, 497–506.
77. Li,W., Prazak,L., Chatterjee,N., Gruninger,S., Krug,L., Theodorou,D. and Dubnau,J. (2013) Activation of transposable elements during aging and neuronal decline in Drosophila. *Nat. Neurosci.*, **16**, 529–531.
78. Gorbunova,V., Boeke,J.D., Helfand,S.L. and Sedivy,J.M. (2014) Human genomics. Sleeping dogs of the genome. *Science*, **346**, 1187–1188.
79. Wood,J.G. and Helfand,S.L. (2013) Chromatin structure and transposable elements in organismal aging. *Front. Genet.*, **4**, 274.
80. Gonzalez,J., Karasov,T.L., Messer,P.W. and Petrov,D.A. (2010) Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila. *PLoS Genet.*, **6**, e1000905.
81. Brennecke,J., Aravin,A.A., Stark,A., Dus,M., Kellis,M., Sachidanandam,R. and Hannon,G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*, **128**, 1089–1103.
82. Brennecke,J., Malone,C.D., Aravin,A.A., Sachidanandam,R., Stark,A. and Hannon,G.J. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science*, **322**, 1387–1392.
83. Khurana,J.S., Wang,J., Xu,J., Koppetsch,B.S., Thomson,T.C., Nowosielska,A., Li,C., Zamore,P.D., Weng,Z. and Theurkauf,W.E.

(2011) Adaptation to P element transposon invasion in Drosophila melanogaster. *Cell*, **147**, 1551–1563.
84. Kim,K.E., Peluso,P., Babayan,P., Yeadon,P.J., Yu,C., Fisher,W.W., Chin,C.S., Rapicavoli,N.A., Rank,D.R., Li,J. *et al.* (2014) Long-read, whole-genome shotgun sequence data for five model organisms. *Sci. Data*, **1**, 140045.
85. Berlin,K., Koren,S., Chin,C.S., Drake,J.P., Landolin,J.M. and Phillippy,A.M. (2015) Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat. Biotechnol.*, **33**, 623–630.
86. Junakovic,N. and Angelucci,V. (1986) Polymorphisms in the genomic distribution of copia-like elements in the related laboratory stocks of Drosophila melanogaster. *J. Mol. Evol.*, **24**, 83–88.
87. Nuzhdin,S.V. and Mackay,T.F. (1995) The genomic rate of transposable element movement in Drosophila melanogaster. *Mol. Biol. Evol.*, **12**, 180–181.
88. Lu,J. and Clark,A.G. (2010) Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. *Genome Res.*, **20**, 212–227.
89. Petrov,D.A., Fiston-Lavier,A.S., Lipatov,M., Lenkov,K. and Gonzalez,J. (2011) Population genomics of transposable elements in Drosophila melanogaster. *Mol. Biol. Evol.*, **28**, 1633–1644.
90. Kofler,R., Nolte,V. and Schlotterer,C. (2015) Tempo and mode of transposable element activity in Drosophila. *PLoS Genet.*, **11**, e1005406.
91. Harada,K., Yukuhiro,K. and Mukai,T. (1990) Transposition rates of movable genetic elements in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 3248–3252.
92. Maside,X., Bartolome,C., Assimacopoulos,S. and Charlesworth,B. (2001) Rates of movement and distribution of transposable elements in Drosophila melanogaster: in situ hybridization vs Southern blotting data. *Genet Res.*, **78**, 121–136.
93. Lee,Y.C. and Langley,C.H. (2012) Long-term and short-term evolutionary impacts of transposable elements on Drosophila. *Genetics*, **192**, 1411–1432.
94. Dominguez,A. and Albornoz,J. (1996) Rates of movement of transposable elements in Drosophila melanogaster. *Mol. Gen. Genet.*, **251**, 130–138.
95. Vazquez,J.F., Albornoz,J. and Dominguez,A. (2007) Direct determination of the effects of genotype and extreme temperature on the transposition of roo in long-term mutation accumulation lines of Drosophila melanogaster. *Mol. Genet. Genomics*, **278**, 653–664.
96. Eggleston,W.B., Johnson-Schlitz,D.M. and Engels,W.R. (1988) P-M hybrid dysgenesis does not mobilize other transposable element families in D. melanogaster. *Nature*, **331**, 368–370.
97. Maside,X., Assimacopoulos,S. and Charlesworth,B. (2000) Rates of movement of transposable elements on the second chromosome of Drosophila melanogaster. *Genet. Res.*, **75**, 275–284.
98. Nuzhdin,S.V., Pasyukova,E.G. and Mackay,T.F. (1997) Accumulation of transposable elements in laboratory lines of Drosophila melanogaster. *Genetica*, **100**, 167–175.
99. Kelleher,E.S. and Barbash,D.A. (2013) Analysis of piRNA-mediated silencing of active TEs in Drosophila melanogaster suggests limits on the evolution of host genome defense. *Mol. Biol. Evol.*, **30**, 1816–1829.
100. Castillo,D.M., Mell,J.C., Box,K.S. and Blumenstiel,J.P. (2011) Molecular evolution under increasing transposable element burden in Drosophila: a speed limit on the evolutionary arms race. *BMC Evol. Biol.*, **11**, 258.
101. Lee,Y.C. (2015) The role of piRNA-mediated epigenetic silencing in the population dynamics of transposable elements in Drosophila melanogaster. *PLoS Genet.*, **11**, e1005269.
102. Deceliere,G., Charles,S. and Biemont,C. (2005) The dynamics of transposable elements in structured populations. *Genetics*, **169**, 467–474.
103. Daniels,S.B., Peterson,K.R., Strausbaugh,L.D., Kidwell,M.G. and Chovnick,A. (1990) Evidence for horizontal transmission of the P transposable element between Drosophila species. *Genetics*, **124**, 339–355.
104. Simmons,G.M. (1992) Horizontal transfer of hobo transposable elements within the Drosophila melanogaster species complex: evidence from DNA sequencing. *Mol. Biol. Evol.*, **9**, 1050–1060.
105. Lerat,E., Burlet,N., Biemont,C. and Vieira,C. (2011) Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. *Gene*, **473**, 100–109.
106. Bartolome,C., Bello,X. and Maside,X. (2009) Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes. *Genome Biol.*, **10**, R22.
107. Arnault,C. and Dufournel,I. (1994) Genome and stresses: reactions against aggressions, behavior of transposable elements. *Genetica*, **93**, 149–160.
108. Capy,P., Gasperi,G., Biemont,C. and Bazin,C. (2000) Stress and transposable elements: co-evolution or useful parasites? *Heredity (Edinb.)*, **85**, 101–106.
109. Nuzhdin,S.V. (1999) Sure facts, speculations, and open questions about the evolution of transposable element copy number. *Genetica*, **107**, 129–137.
110. Kidwell,M.G. and Lisch,D.R. (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution*, **55**, 1–24.

Rahman, et al.

SUPPLEMENTARY MATERIALS

SUPPLEMENTAL FIGURES AND TABLES LEGENDS

Figure S1. Detailed flowchart of the TIDAL program.

After the split, transposon insertions are detected through the steps on the left, while depletions are detected through the steps on the right. Complete parameter details are described in the Methods section.

Figure S2. Comparison of the overall transposon distributions and genomic architecture between the current Release6/dm6 build and the previous Release5/dm3 build, and functional annotations of TE InDels in cell line genomes.

(**A**) The percentage content of RepeatMasker masked sequences per 50kb window is shown above the banding patterns for each chromosome of the *D. melanogaster* genome. Many of the regions previously absent in the diagrams for Release5/dm3 were not assembled as contiguous portions of the main chromosome arms. (**B**) Proportions of the functional annotations for TE InDels in the various cell line genomes. This panel relates to Fig. 2.

Figure S3. Genomic PCR validation of TIDAL predictions for *D. melanogaster* S2c1 cells (Part 1) and ISO1 sub-strains (Part 2).

Part 1: (**A**) Transposon insertions validation tests for S2c1 cells predictions. Top, schematic of primers designed to flank the TE insertion by pairing to unique genomic sequence. Middle, gels of 15 sites tested by PCR with genomic DNA from S2c1 cells, and ISO1-BL and RAL-362 flies. Bottom, table listing the loci with TE insertions tested in the PCR assays above. (**B**) Transposons depletions validation tests for S2c1 cells predictions. Top, schematic of primers designed to flank the TE depletion by pairing to unique genomic sequence. Middle, gels of 10 sites tested by PCR with genomic DNA from S2c1 cells, and ISO1-BL and RAL-362 flies. Bottom, table listing the loci with TE depletions tested in the PCR assays above. Asterisks mark judgement calls for validations based on the absence of the reference genome allele for TE Insertions and

presence of a smaller band in the TE Depletions as an indication that the PCR result was supportive of the TIDAL prediction. **Part 2:** (**C**) Transposon insertions validation tests for ISO1-BL predictions. Top, schematic of primers designed to flank the TE insertion by pairing to unique genomic sequence. Middle, gels of 12 sites tested by PCR with genomic DNA from RAL-362, ISO1-UC, and ISO1-BL flies. Bottom, table listing the loci with TE insertions tested in the PCR assays above. (**D**) Transposons depletions validation tests for ISO1-BL predictions. Top, schematic of primers designed to flank the TE depletion by pairing to unique genomic sequence (left) and a PCR that uses one reverse primer base-pairing to the TE sequence present in the reference genome sequence (right). The reverse primer against the TE is more prone to generating a TE amplicon artifact. Middle, gels of 12 sites tested by PCR with genomic DNA from RAL-362, ISO1-UC, and ISO1-BL flies. Left gels are PCRs from primers that both flank the TE, while the right gels are PCRs with one reverse primer against TE sequence, which tends to fail more frequently with primer dimers. Bottom, table listing the loci with TE depletions tested in the PCR assays above.

Figure S4. Additional analyses of TLs in Canton-S sub-strains.
(**A**) Euler diagrams showing the overlap of TE InDels between sub-strains from 5 different labs, with all TE InDels in the top diagrams, whereas below are only the TE InDels with CR≥3.0. (**B**) Similar Euler diagrams as in (**A**) except that the CanS-JC sub-strain replaces the CanS-SH sub-strain. These diagrams are to illustrate the notably stronger overlap in TE insertions between the CanS-JC and the CanS-TP sub-strains, which are two lines maintained independently but both originating from the same single lab source. (**C**) Proportions of all of the InDels by TE family between the CanS sub-strains. (**D**) Proportions of InDels with CR≥3.0 by TE family between the CanS sub-strains, which is generally similar to the proportions of all TE InDels in (**C**). (**E**) Proportions of the genomic functional annotations immediately near the sites of TE InDels. (**F**) Scatterplot comparing sequencing depth to numbers of TE InDels detected amongst CanS sub-strain libraries (various combinations of independent sequencing runs). Trend lines show linear relationship between sequencing depth and server

processing time, and logarithmic relationship between depth and numbers of TE InDels detected by TIDAL.

Figure S5. Additional analyses of TLs in ISO1 sub-strains.
(**A**) Euler diagrams comparing the overlap of TE InDels between –BL and –CL sub-strain of ISO1, (**B**) between –BL and –AS sub-strains, and (**C**) between different tissues of the ISO1–AS fly line. Abbreviations: BL, Bloomington; CL, Charles Langley lab; AS, Alan Spradling lab; HEm12, Haploid Embryo samples 1&2; Ova, Ovaries; LSg123, Larval Salivary gland samples 1,2,3; Emb123, Embryos sample 1,2,3. The ISO1-BL library is from adult females from the Bloomington Drosophila Stock Center, the ISO1-CL libraries (whole flies and haploid embryos) are from the Langley lab, and the ISO1-AS libraries (ovaries, embryos, and larval salivary glands) are from the Spradling lab. (**D**) Proportions of InDels by TE family between ISO1 sub-strains. (**E**) Comparison of the InDels proportions by TE family predicted by TIDAL from different combinations of sequencing runs of the ISO1-BL libraries. Greater sequencing depth generally uncovers the rarer TE InDels in the genomic library. (**F**) Scatterplot comparing sequencing depth to numbers of TE InDels detected amongst ISO1 sub-strains. Trend lines show linear relationship between sequencing depth and server processing time, and logarithmic relationship between depth and numbers of TE InDels detected by TIDAL. (**G**) Scatterplot comparing sequencing depth to numbers of TE InDels detected amongst various combinations of independent sequencing runs of the ISO1-BL library. (**H**) Calculations of transposition rates in the ISO1-BL sub-strain for each TE per genomic copy and per generation, assuming that ~450 generations (~15 years between the sequencing of ISO1-BL and the original reference genome isolate, and an average 12 day life cycle).

Figure S6. Additional benchmarking results comparing TIDAL with LnB, CnT, and TEMP programs for TE insertion predictions in DGRP fly strains.
(**A**) For DGRP strains that are only analyzed by TEMP and TIDAL, there is much greater overlap in TE depletion predictions between the two programs (red bars) compared to the TE insertion predictions. (**B**) Table comparing the general design,

distinctions, and limitations of our TIDAL program compared to the other three programs with common outputs for DGRP strains. A diagram of chromosome 4's high density of transposons, from Figure S2, is shown as a reminder for how false positives can be spotted in other programs. (**C**) Euler diagrams of the overlap in TE insertions predicted by all four programs for RAL-109 and RAL-391 strains, with diagrams on the left showing all TE insertions and the diagrams on the right that lack obvious false positives such as Chr4 and Chr#Het predictions, these are frequently repetitive genomic loci that for some reason have not been masked by RepeatMasker. Note the removal of these false positives largely only reduce the predictions unique to TEMP and CnT programs.

Figure S7. Comparison of genomic PCR validation tests between TIDAL, LnB, CnT and TEMP programs for transposon insertions predictions for the RAL-765 strains. (**A**) Transposon insertions only predicted by TIDAL. Top, gels of 12 sites tested by PCR with genomic DNA from RAL-765, RAL-362, RAL-517, and ISO1-BL flies. Bottom, table listing the loci with TE insertions tested in the PCR assays above. The same format of panel A extends to panels B, C, and D. (**B**) Transposon insertions predicted by LnB which also mostly overlap with TIDAL calls. (**C**) Transposon insertions only predicted by the TEMP program. (**D**) Transposon insertions only predicted by the CnT program. Asterisks mark judgement calls for validations based on the absence of the reference genome allele for TE Insertions and presence of a smaller band in the TE Depletions as an indication that the PCR result was supportive of the TIDAL prediction.

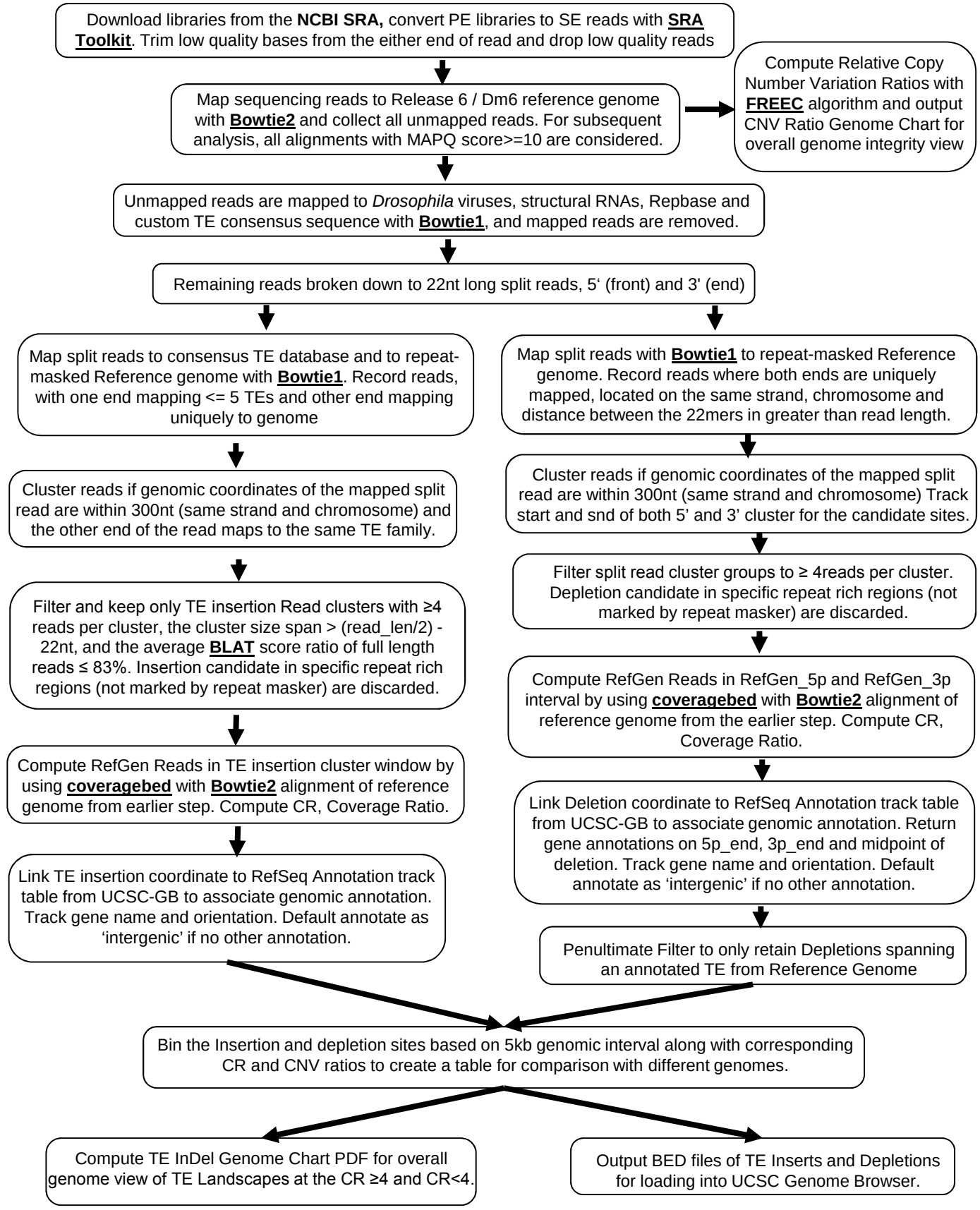
Figure S8. Detailed profiles of transposon depletion landscapes in DGRP and DGN fly strains and pools of flies. (**A**) 45 DGRP Fly strains. (**B**) 46 DGN Fly strains. (**C**) 32 Pools of flies. The top panels show the total counts of transposon insertions, the middle panels show the proportions of each transposon class, and the bottom panels show the proportions of the transposon species with ≥20 insertions (the rest are grouped together at the top). The dashed lines and labels in the bottom panels mark notable transposon species.

Figure S9. Detailed profiles of transposon insertion landscapes in DGRP and DGN fly strains and pools of flies, referring to the abridged profiles in Fig. 5. (**A**) 45 DGRP Fly strains. (**B**) 46 DGN Fly strains. (**C**) 32 Pools of flies. The top panels show the total counts of transposon insertions, the middle panels show the proportions of each transposon class, and the bottom panels show the proportions of the transposon species with ≥20 insertions (the rest are grouped together at the top). The dashed lines and labels in the bottom panels mark notable transposon species.

Table S1. Details and accession numbers for Illumina libraries of *D. melanogaster* cell lines and fly strains analyzed in this study and incorporated into the TIDAL-Fly database. There are 5 tabs in this worksheet. (**A**) *D. melanogaster* cell lines. (**B**) DGRP fly strains. (**C**) DGN fly strains. (**D**) Key for the demographics of DGN fly strains. (**E**) Pools of flies.

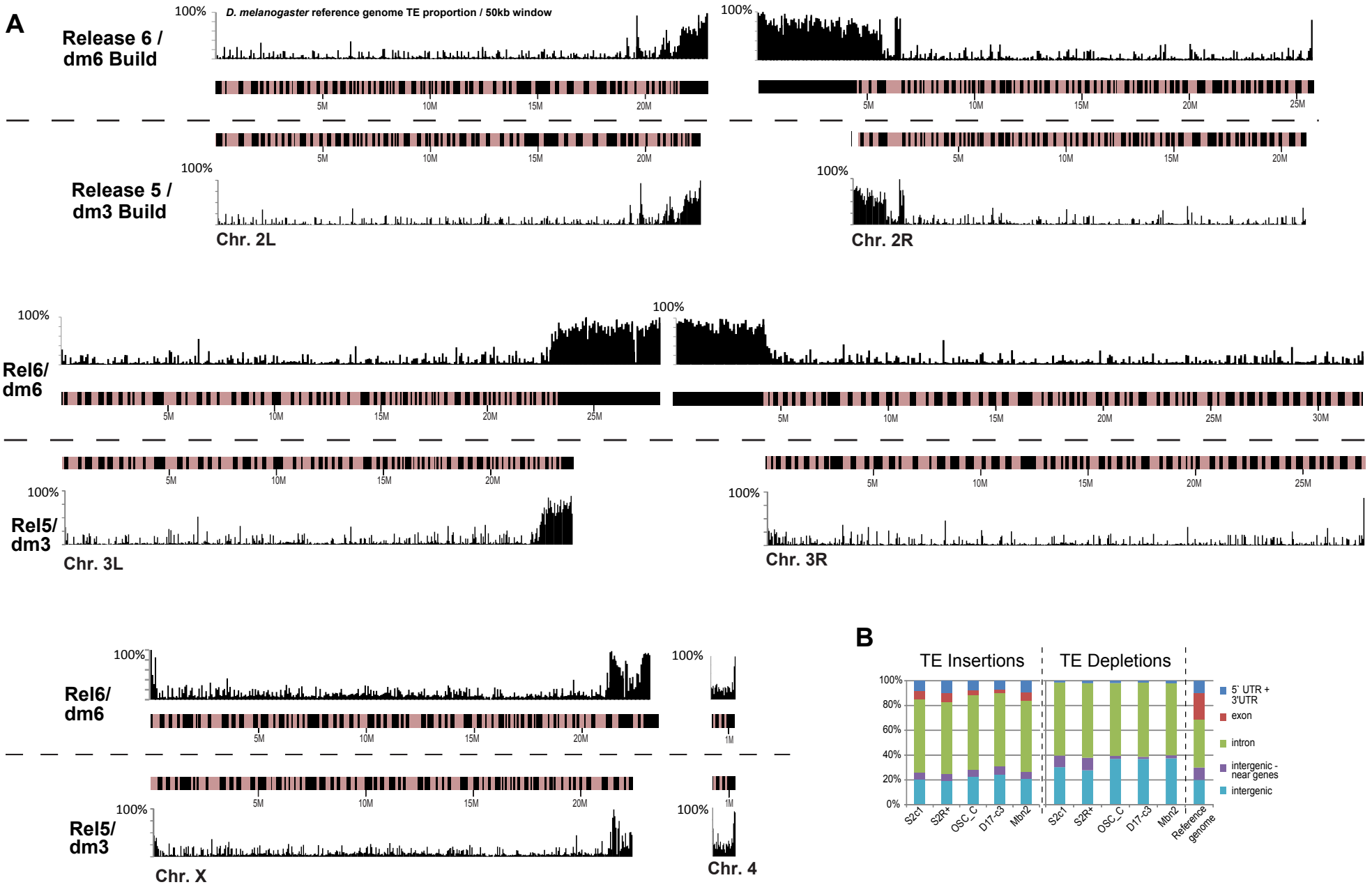
Table S2. Oligonucleotides used in this study.

Rahman, et al.
Figure S1.

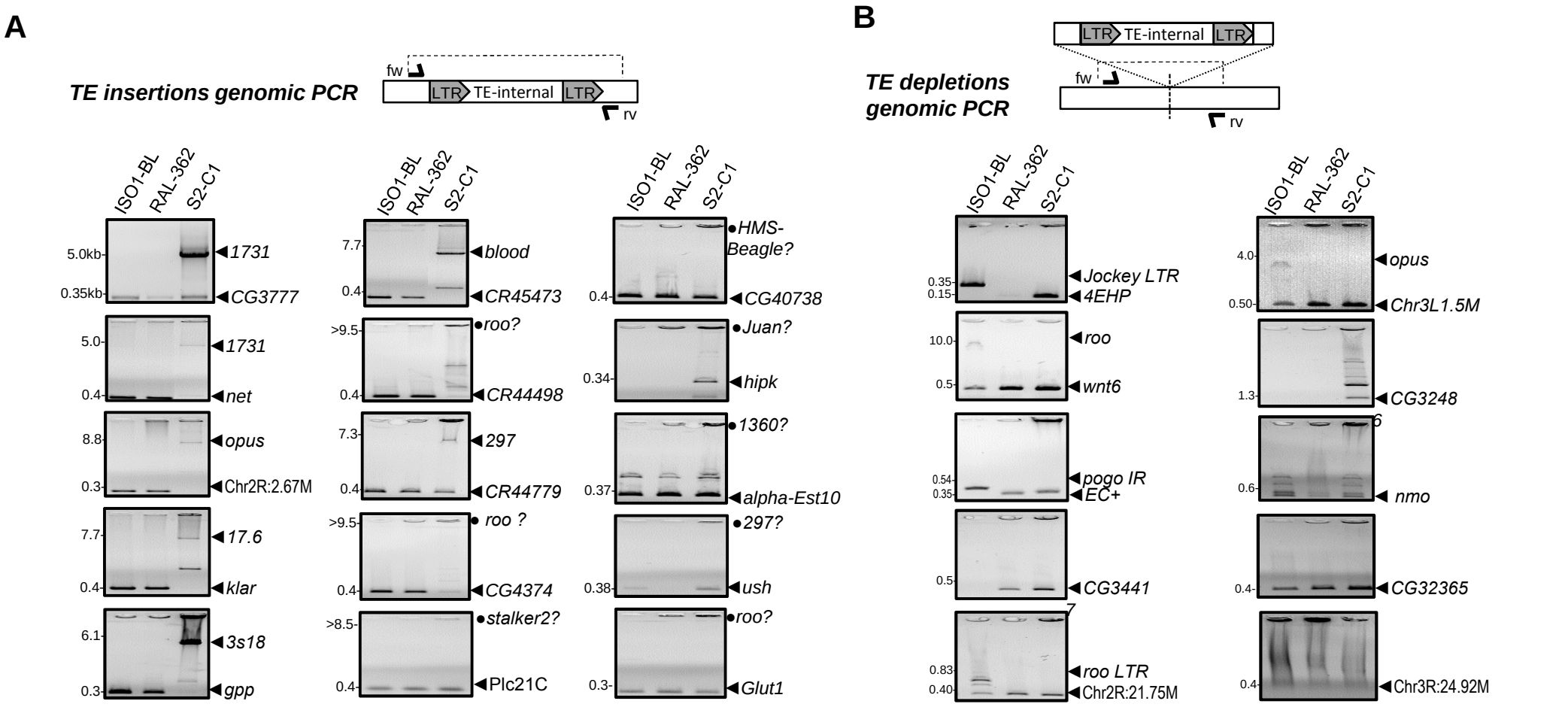


See Methods Text for details of program parameters and equations.

Rahman et al.
Figure S2.



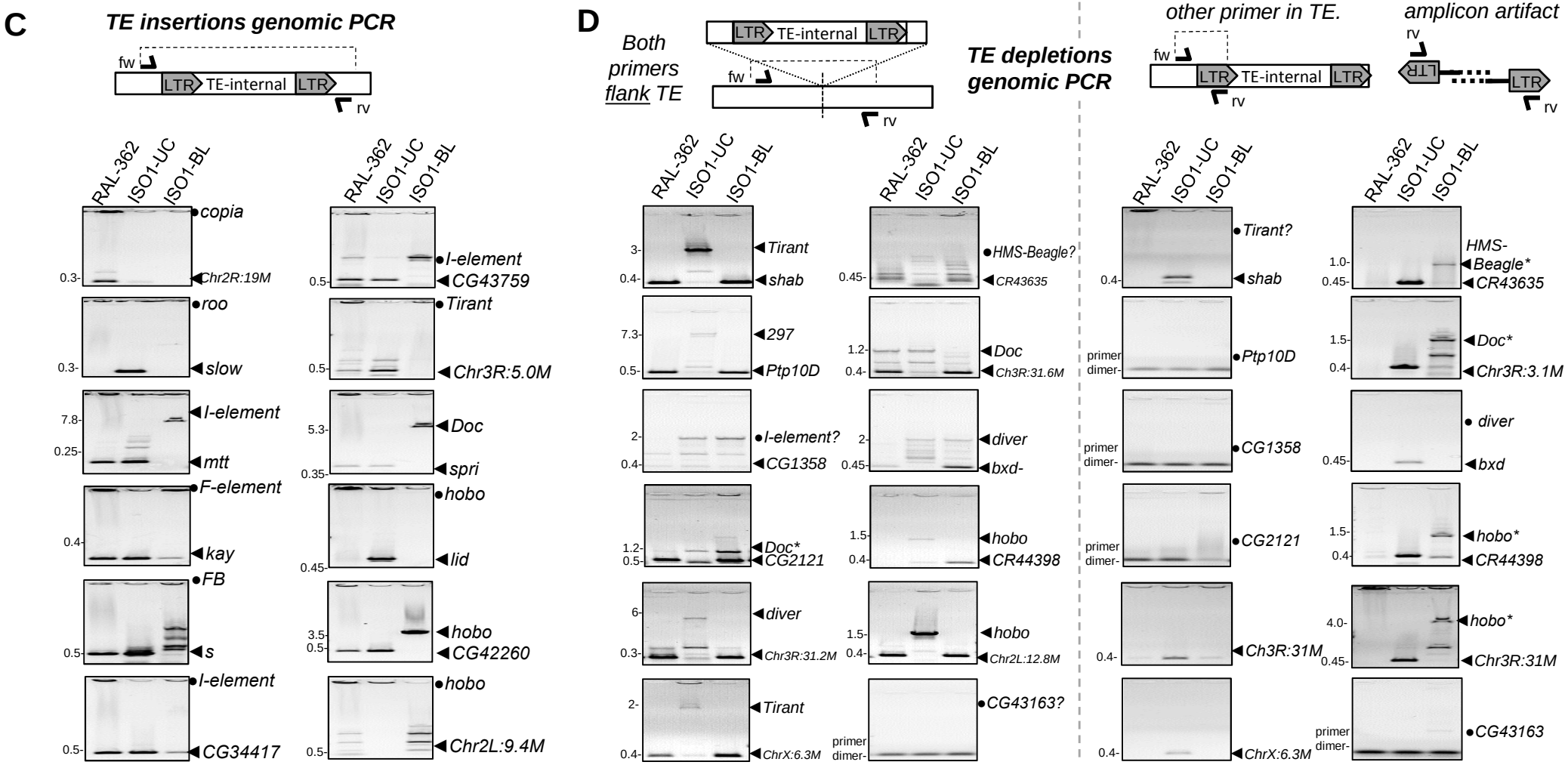
Rahman, et al. Figure S3, PART 1



Site. No.	T.E.	Locus	Nearest Gene	Coverage Ratio	Size w/o insertion	Size with insertion	PCR evidence
1	*1731*	chrX:325934-326139	CG3777 +,CG32816 -	4.5	349	4998	yes
2	*1731*	chr2L:87055-87258	net -	4	404	5053	yes
3	*opus*	chr2R:2675584-2675743	not near genes	3.4	332	8891	yes
4	*17.6*	chr3L:516184-516383	klar -	5.1	349	7788	yes
5	*3S18*	chr3R:6430955-6431163	gpp +	5.6	293	6126	yes
6	*blood*	chrX:141518-141720	CR45473 -	5.6	379	7790	yes
7	*roo*	chrX:6005755-6005959	CR44498 +	5	381	9473	yes*
8	*297*	chrX:1082310-1082488	CR44779 +,CG3655 -	3.7	373	7368	yes
9	*roo*	chr3R:23470563-23470775	CG4374 -	4	367	9459	yes*
10	*Stalker2*	chr2L:312233-312444	Plc21C +	5.9	411	8661	no
11	*HMS-Beagle*	chr2R:544728-544940	CG40378 +	4.9	353	6882	yes*
12	*Juan*	chr3L:552782-552981	hipk +	3.3	344	4580	no
13	*1360*	chr3R:7492347-7492543	alpha-Est10 -	4.7	374	4480	no
14	*297*	chr2L:522270-522480	ush +	4.9	386	7381	no
15	*roo*	chr3L:957937-958157	Glut1 +	7.6	296	9388	no

Site. no.	T.E.	Locus	Nearest Gene	Coverage Ratio	Size with depletion	Size without depletion	PCR evidence
1	jockey	chr3R:24072059-24072628	4EHP -	4.5	150	341	yes
2	roo	chr2L:7343256-7350536	Wnt6 +	4.6	5757	7480	yes
3	pogo	chrX:3842276-3842647	ec +	5.8	355	541	yes
4	Tirant	chrX:6562758-6571399	CG34417 +	5.2	1240	8802	yes
5	roo	chr2R:21753020-21753663	not near genes	5.2	405	833	yes
6	opus	chr3L:15423693-15426700	not near genes	4.6	2642	3160	yes
7	Tirant	chr3L:3061219-3069958	CG32486 -	4.6	1350	8912	yes*
8	Stalker4	chr3L:7886449-7889129	CG32365 +	4.3	1107	2873	no
9	NOF	chr3L:7985303-7987783	nmo +	4.2	620	2650	no
10	Doc	chr3R:24927069-24932010	not near genes	5.7	395	5120	no

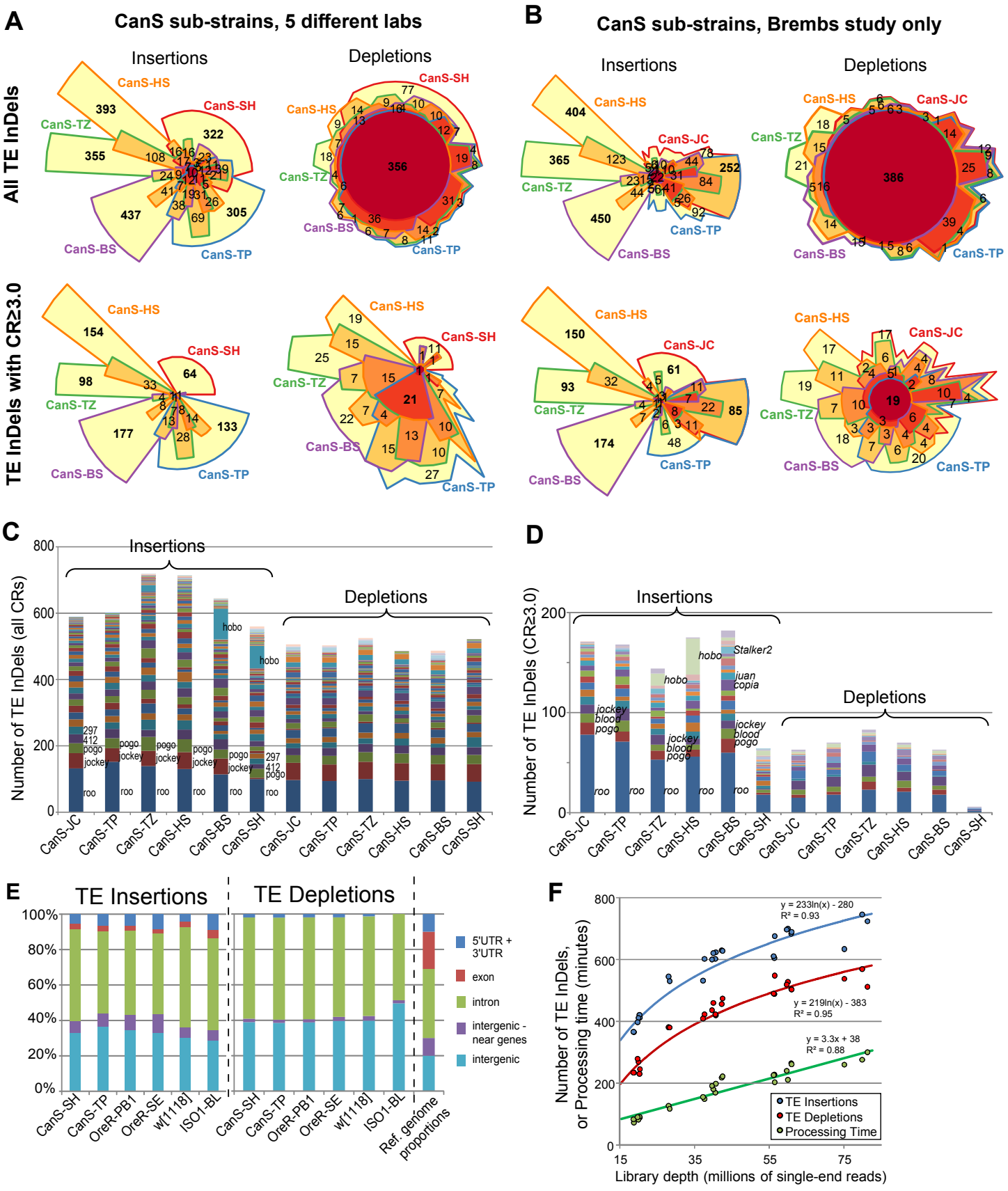
Rahman, et al. Figure S3, PART 2



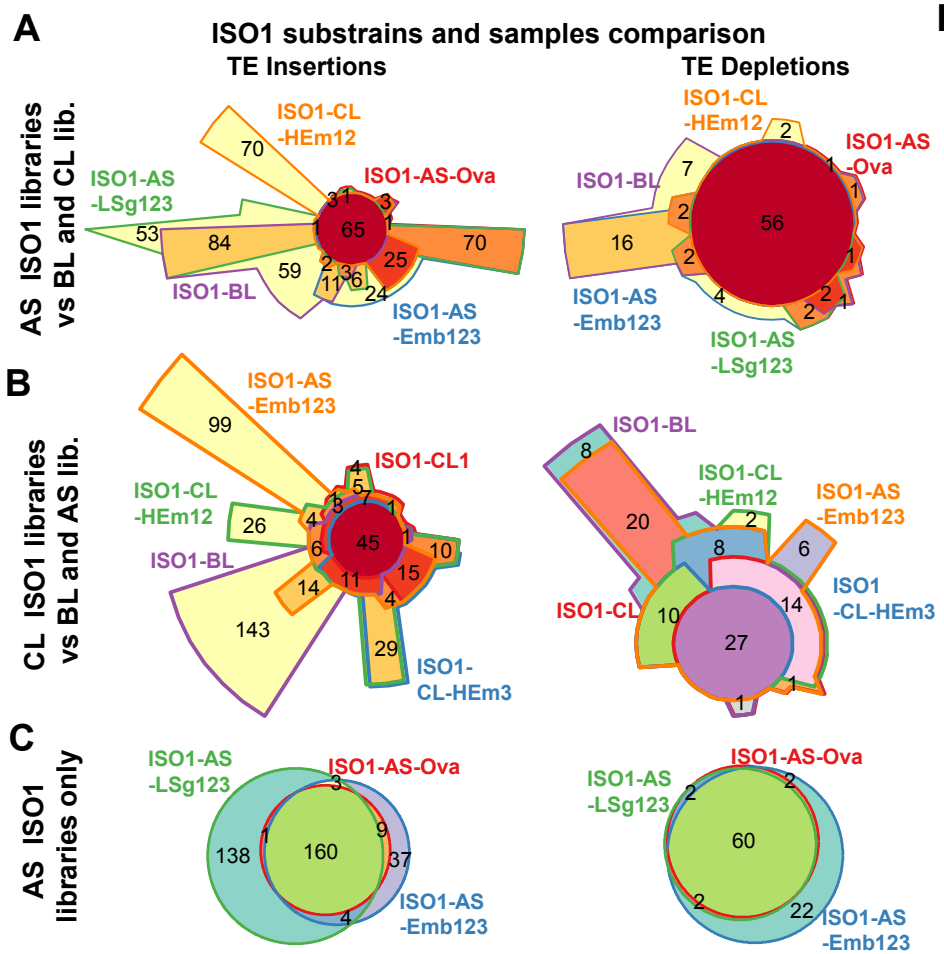
Site No.	Locus	Nearest gene	T.E.	C.R.	size w/o insertion	size with insertion	PCR validation
1	chr2R:19768184-19768307	not near genes	copia	4.5	295bp	5.1kb	yes
2	chr3L:4390337-4390468	slow	roo	4.5	330bp	9kb	yes
3	chr2R:8528862-8528983	mtt	I-element	4.4	273bp	7.8kb	yes
4	chr3R:29777197-29777247	kay	F-element	17.9	414bp	5kb	yes
5	chr2L:1068653-1068778	s	FB	4.2	517bp	2kb	yes
6	chrX:6542904-6543018	CG34417	I-element	4.3	501bp	7.9 kb	yes
7	chrX:19018636-19018748	CG43759	copia	4.7	469bp	5.6kb	yes
8	chr3R:5066592-5066727	not near genes	Tirant	4.4	515bp	9kb	yes
9	chrX:10556247-10556301	spri	Doc	5.7	363bp	5.3kb	yes
10	chr2L:5999659-5999785	lid	hobo	5.6	479bp	3.5kb	yes
11	chr2R:22786248-22786444	CG42260	hobo	5.3	523bp	3.5 kb	yes
12	chr2L:9485261-9485390	not near genes	hobo	4.2	514bp	3.5kb	no

Site No.	Locus	Nearest gene	T.E.	C.R.	size with depletion	size without depletion	PCR validation	
							Flank TE	Gene-T.E.
1	chr3L:2927813-2927871	shab+	Tirant	3.5	393bp	8.5kb	Yes	yes
2	chrX:11633392-11633445	Ptp10D+	297	2.9	497bp	7.3kb	Yes	No
3	chr2R:7608382-7608431	CG1358-	I-element	2.6	509bp	5.4kb	No	No
4	chr2R:8398189-8398240	CG2121-	Doc	3	502bp	5.0kb	Yes*	No
5	chr3R:31279353-31279405	not near genes	diver	2.5	436bp	6.4kb	Yes	yes
6	chrX:6335185-6335241	not near genes	Tirant	4.6	380bp	2.7kb	Yes	yes
7	chr3R:5045818-5045876	CR43635-	HMS-Beagle	4.4	448bp	7.3kb	No	yes*
8	chr3R:31691509-31691562	not near genes	Doc	4.3	449bp	5.0kb	Yes	yes*
9	chr3R:16756294-16756346	bxd-	diver	3.6	436 bp	6.4kb	Yes	yes*
10	chr2L:17951924-17951978	CR44398-	hobo	3.5	381bp	1.7kb	Yes	yes*
11	chr2L:12861671-12861728	not near genes	hobo	3.5	458bp	1.7kb	Yes	yes*
12	chr3L:8481746-8481802	CG43163-	roo	3.3	487bp	9.4kb	No	No

Rahman et al.
Figure S4.

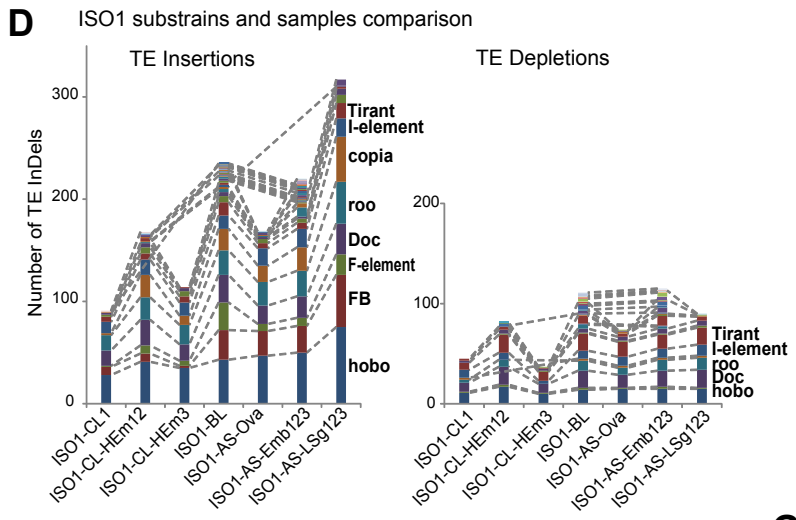


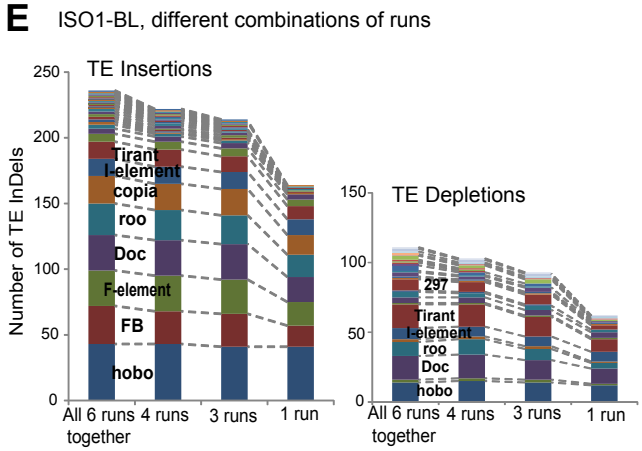
Rahman et al.
Figure S5.

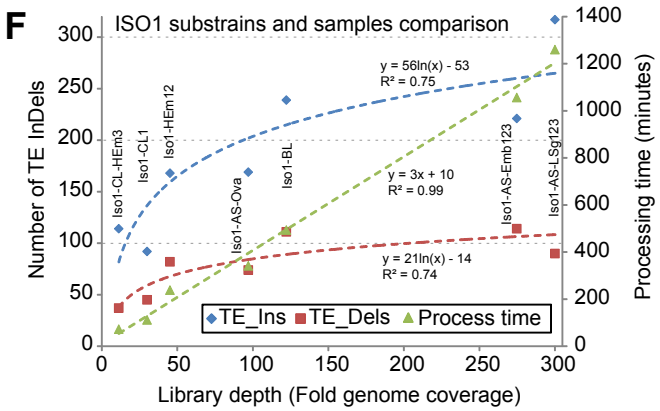


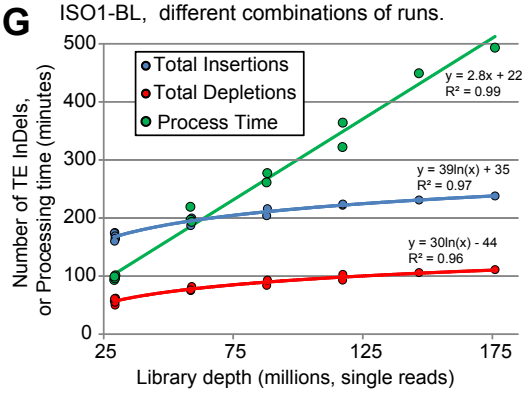
H

New TE insertions in current ISO1-BL genome	# New insertion	# Existing insertions	Rate/copy/ generation
hobo	43	201	4.8E-04
FB	29	329	2.0E-04
Doc	27	276	2.2E-04
F-element	27	471	1.3E-04
roo	24	1000	5.3E-05
copia	21	302	1.5E-04
Tirant	13	127	2.3E-04
I-element	13	235	1.2E-04
HMS-Beagle	6	350	3.8E-05
diver	4	53	1.7E-04
1360	3	836	8.0E-06
Stalker4	2	408	1.1E-05
jockey	2	178	2.5E-05
297	2	451	9.9E-06
accord	2	46	9.7E-05
Quasimodo	2	421	1.1E-05
Idefix	2	623	7.1E-06
1731	1	121	1.8E-05
Chimpo	1	62	3.6E-05
Tc1-2	1	75	3.0E-05
Cr1a	1	1548	1.4E-06
diver2	1	452	4.9E-06
Stalker	1	408	5.4E-06
Max-element	1	342	6.5E-06
rover	1	184	1.2E-05
G3	1	103	2.2E-05
mdg1	1	316	7.0E-06
McClintock	1	65	3.4E-05
invader2	1	209	1.1E-05
invader6	1	112	2.0E-05
Doc3-element	1	325	6.8E-06
Average of each individual TE rate:			7.0E-05

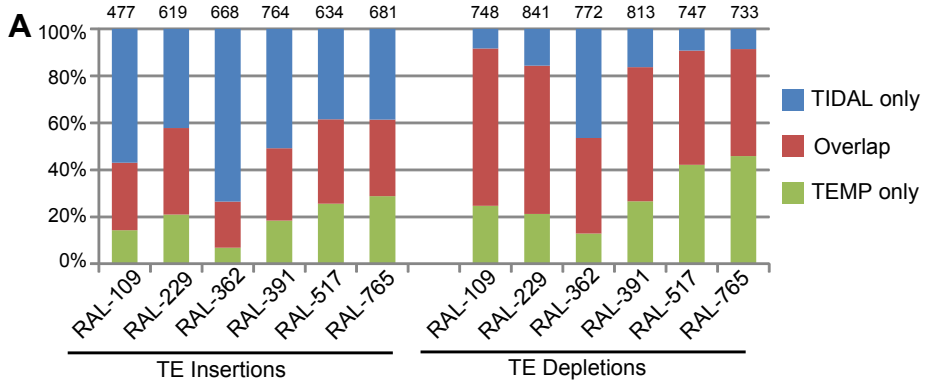






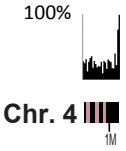


Rahman et al.
Figure S6.

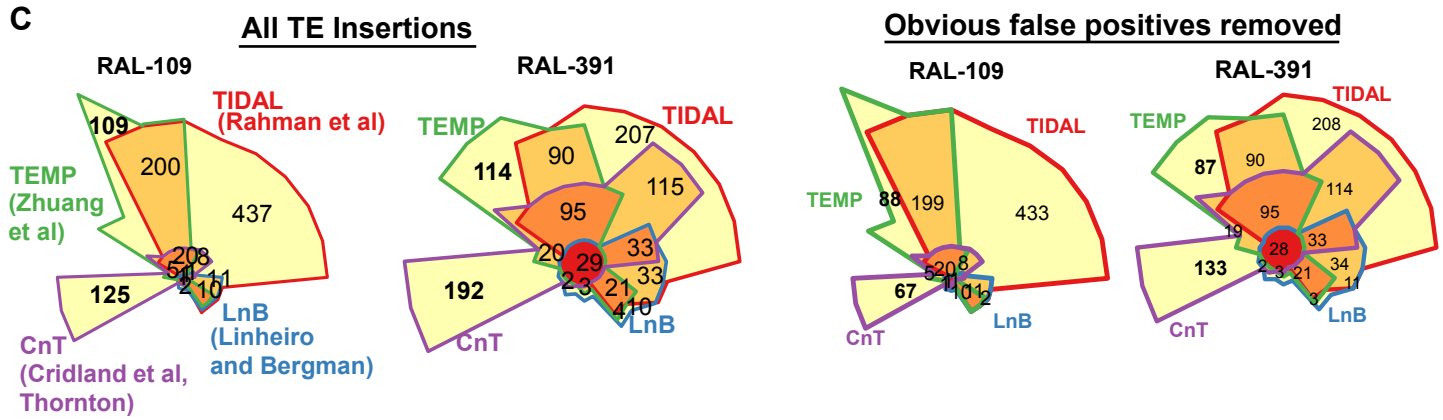


B

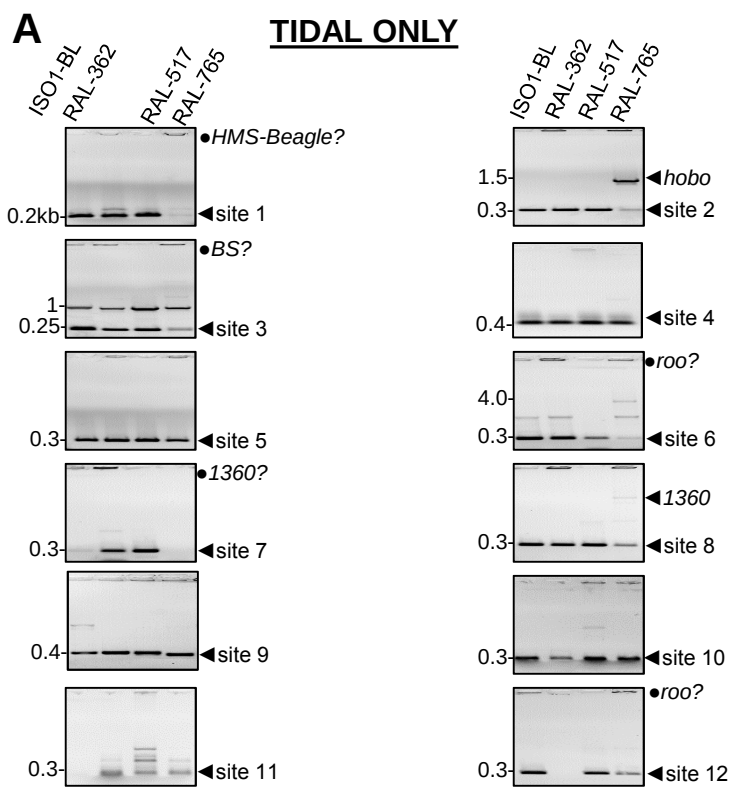
Program	TIDAL-Fly	TEMP	CnT	LnB
Citation	(Rahman et al, this study)	(Zhuang et al, 2014)	(Cridland et al, 2013)	(Linheiro and Bergman, 2012)
Pipeline summary	Bowtie1 aligner for split reads, Bowtie2 aligner for reference genome reads, and BLAT to enforce TE insertion read calls. Only single-end split reads processed for discordant mappings on both sides of InDel.	BWA aligner, full tracking of matched paired ends, comparison to reference supporting reads across strains and computes a population frequency.	BWA aligner, full tracking of matched paired ends, reconstruction of candidate InDels, use of BlastN for validation	BLAT aligner for split reads, first stage identifies discordant reads, second stage selects only unambigous reads on both sides of TE insertion.
Distinctions	Optimization of split short-read approaches to balance sensitivity with specificity. Strain-specific TE landscapes in user-accessible outputs that include rich genomic annotations for Dm6/Rel6 genome release.	Versatility in application being applied to both *Drosophila* and human genomes. Strain-specific DGRP InDels genomic locations can be extracted.	Outputs and analyses focused on population genetics studies for *Drosophila*.	Identifies specific Target Site Duplication sequences. First split-read pipeline applied to both 454-Roche and Illumina datasets from DGRP fly genomes.
Limitations	Excludes any TE InDels in major heterochromatin, given high mapping ambiguity. Cannot automatically determine Target Site Duplication sequences. Currently only optimized for *Drosophila melanogaster* genomes.	Several program-specific predictions in *Drosophila*, some in heterochromatic regions, that have potential to be artifacts. Libraries with multiple read lengths can trip up program.	Many program-specific predictions, some in heterochromatic regions, that have potential to be artifacts. Outputs only available for libraries meeting read depth cutoff. TE Depletion data at broad population level, not strain level.	Pioneering pipeline sacrificed sensitivity by promoting strong specificity of BLAT algorithm. Only detects TE insertions, no capacity to detect TE depletions.

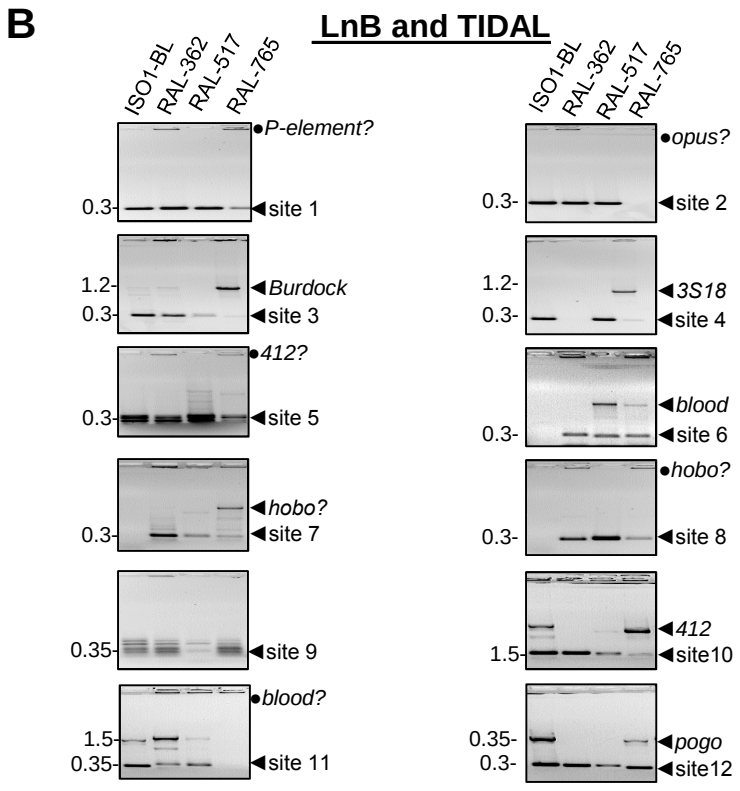


C



Rahman, et al., Figure S7, Part 1

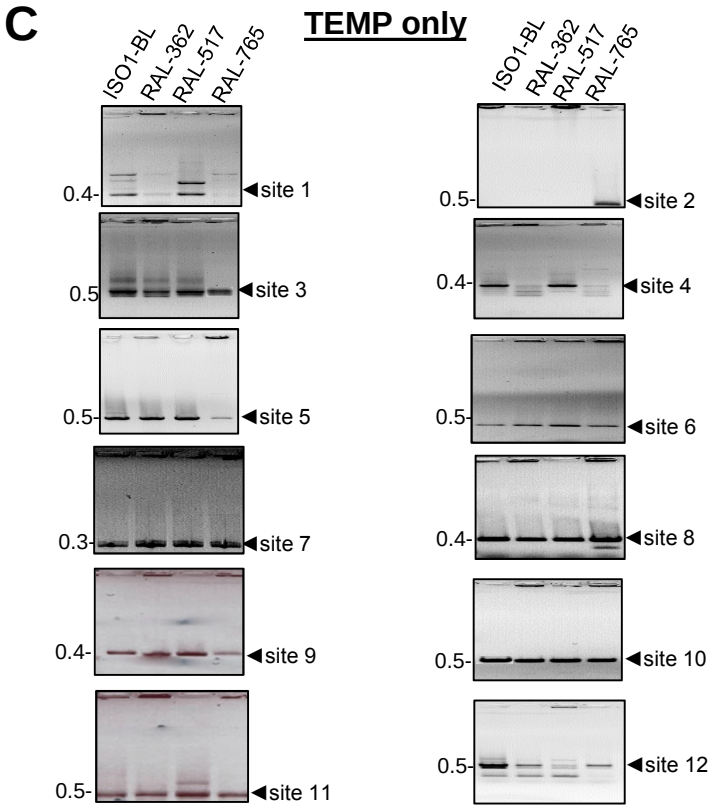


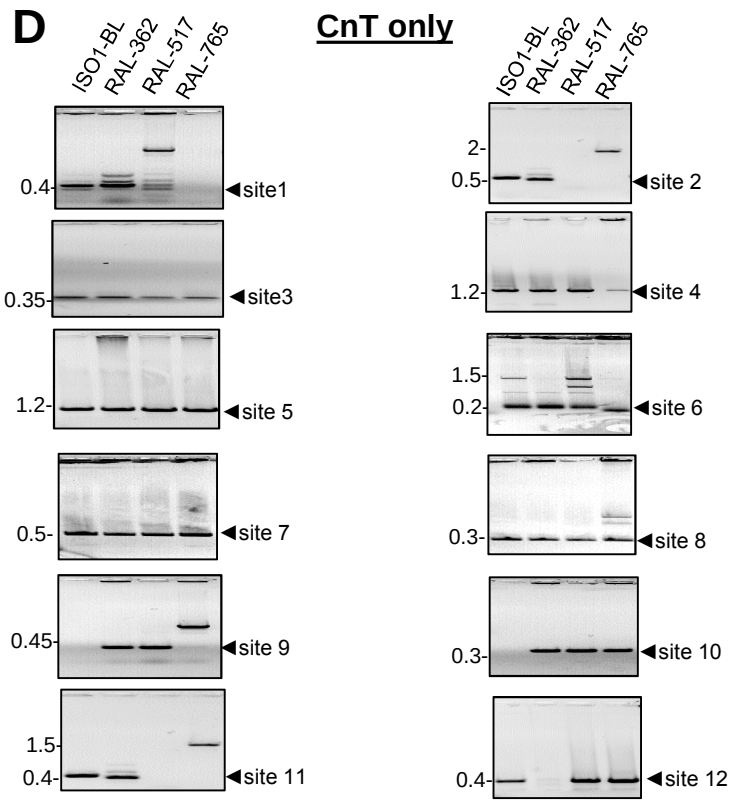


Site. No.	T.E.	Locus	Coverage Ratio	Size w/o insertion	Size with insertion	PCR evidence
1	HMS-Beagle	chr2L8133670::-8133752	2.4	275	7575	yes
2	hobo	chr2L:19834994-19835061	4.7	354	3354	yes
3	BS	chr2L:12227998-12228146	5.3	274	5374	yes
4	pogo	chr2R:12581030-12581105	5	333	2533	no
5	mdg1	chr2R:7621371-7621430	4	328	7628	no
6	roo	chr3L:4481564-4481640	4.1	337	9337	yes
7	1360	chr3L:18394377-18394445	4.2	360	4760	yes
8	1360	chr3R:4885334-4885399	1.7	354	4754	yes
9	jockey	chr3R:20125023-20125087	2.4	360	5360	no
10	roo	chr3R:11089486-11089562	5.4	359	9359	no
11	hobo	chrX:7785488-7785556	3.3	342	3342	no
12	roo	chrX:15391856-15391929	4.1	341	9341	yes

Site. No.	T.E.	Locus	Coverage Ratio	Size w/o insertion	Size with insertion	PCR evidence
1	P-element	chr2L:267478-267547	4.4	314	3214	yes
2	opus	chr2L:1167146-1167232	2.6	386	8386	yes
3	Burdock	chr2L:20800099-20800248	5.8	329	6729	yes
4	3S18	chr2R:9380884-9380961	2.4	365	6465	yes
5	412	chr2R:9442941-9443015	3.9	331	7931	yes
6	blood	chr2R:11578561-11578638	5.6	372	5672	yes
7	hobo	chr3L:934796-934851	8.3	364	3364	yes
8	hobo	chr3L:22566716-22566793	4.3	367	3367	yes
9	Burdock	chr3R:18968019-18968094	3.8	375	6775	no
10	412	chr3R:30626323-30626403	5.2	379	7979	yes
11	blood	chrX:6487781-6487856	4.3	375	5675	yes
12	pogo	chrX:21399401-21399469	3.6	381	2581	yes

Rahman, et al., Figure S7, Part 2

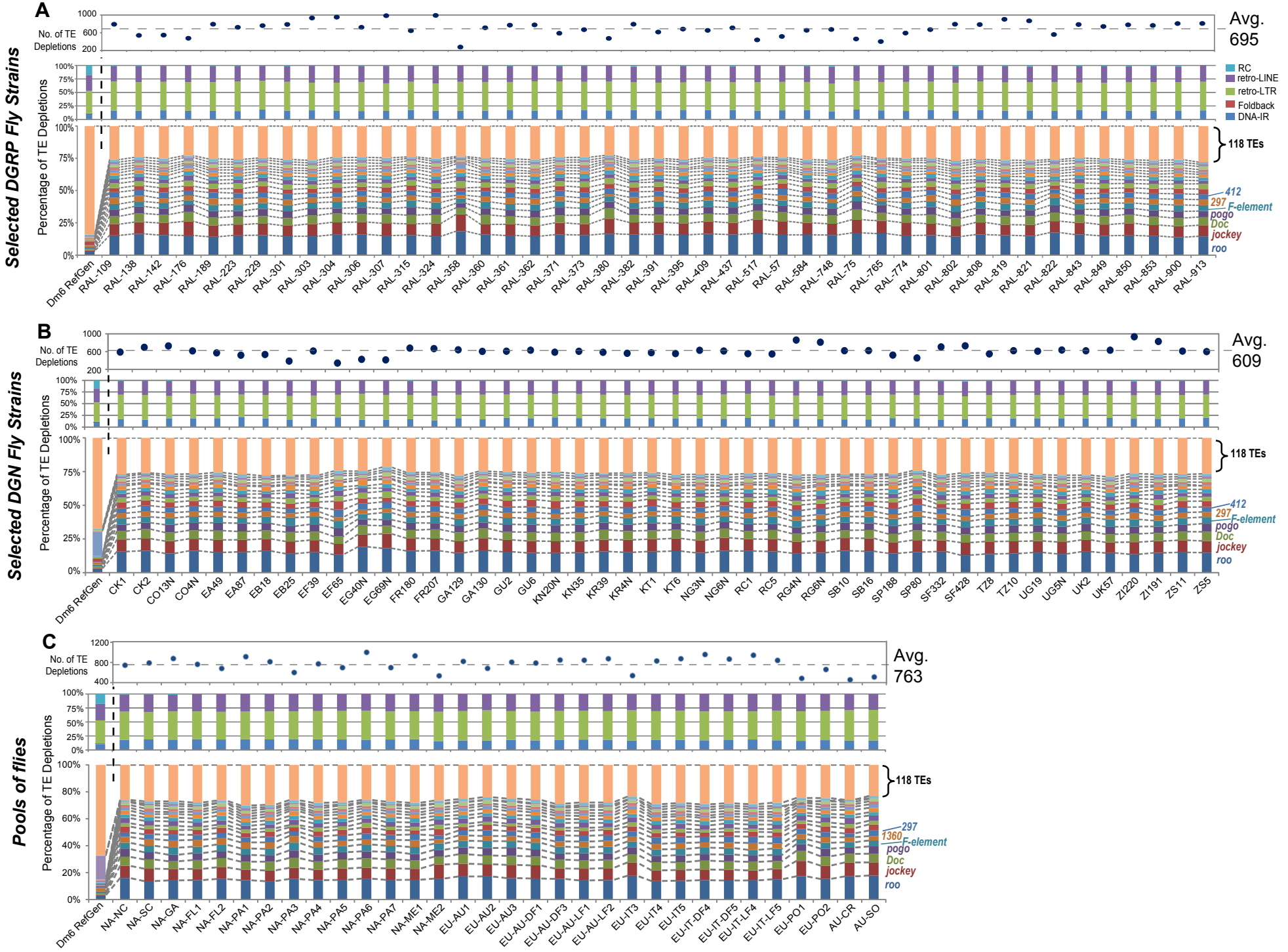




Site. No.	Locus	Amplicon Size	PCR evidence
1	chr2L:9850436-9850443	383	yes
2	chr2L:1558495-1558995	482	no
3	chr2R:5334166-5334666	494	yes
4	chr2R:2095477-2095485	405	no
5	chr3L:9571935-9572435	482	yes
6	chr3L:5056174-5056674	499	no
7	chr3R:5035704-5035739	296	no
8	chr3R:1211275-1211285	398	no
9	chrX:294866-294869	395	yes
10	chrX:5410656-5411156	488	no
11	chrY:203747-204247	487	no
12	chrY:1565254-1565753	645	no

Site. No.	Locus	Amplicon Size*	PCR evidence
1	chrX:4334105-4334327	376	yes
2	chrX:14001033-14007159	6092	yes
3	chr3R:7696150-7703338	7171	no
4	chr3R:14021702-14023463	1734	yes
5	chr3L:4361048-4361901	844	no
6	chr3L:17180645-17180660	394	no
7	chr2R:6948454-6948972	493	no
8	chr2R:13064762-13064861	357	yes
9	chr2L:1809645-1810122	447	yes
10	chr2L:9783398-9783415	394	no
11	chr4:180927-183823	2892	yes
12	chr4:999517-1001425	2008	no

Rahman et al.
Figure S8.



Rahman et al.
Figure S9.

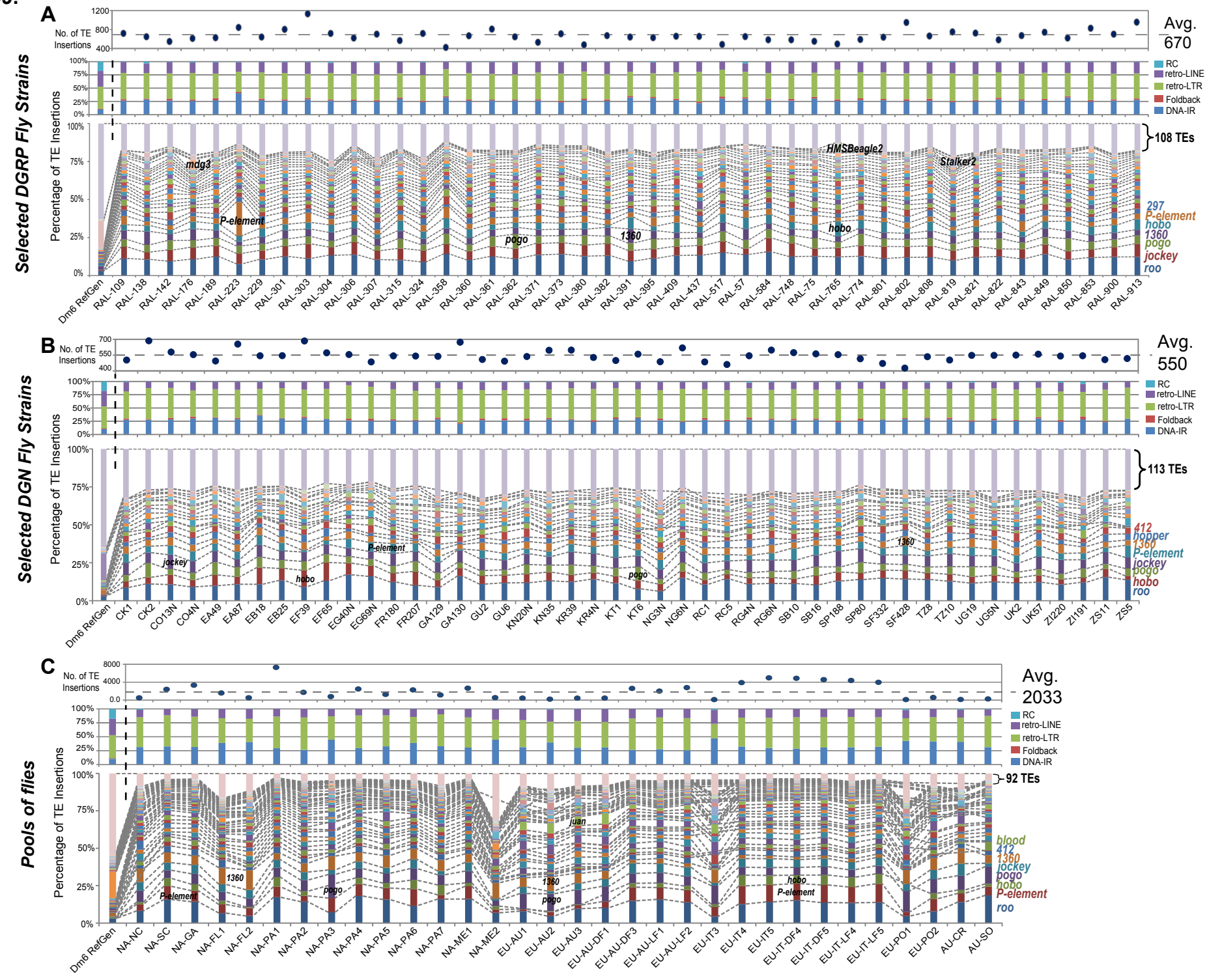


Table S1A. Drosophila Cell Lines in the TIDAL-Fly v1 database

Number	Official name	Tissue origin	Origin genotype	modENCODE Name	Library ID*	SRR Accesion	Min Library Length	Library Size (M)	Read 1 length	Read 2 length	Latest Ref		Other accessions	
1	S2_C1	Embryo	Oregon R	n/a	S2c1	SRR1983913	150	33.8	150	na	This Study			
2	S2R+	Embryo	Oregon R	modENCODE_5492	S2R+	SRR497719	100	7.7	100	100	Lee et al. Genome Bio 2014		SRR497722	2x75
3	S1	Embryo	Oregon R	modENCODE_5491	S1	SRR497713	75	31.8	75	75	Lee et al. Genome Bio 2014			
4	S3	Embryo	Oregon R	modENCODE_5493	S3	SRR497721	100	11.4	100	100	Lee et al. Genome Bio 2014			
5	Sg4	Embryo	Oregon R	modENCODE_5494	Sg4	SRR497720	100	20.5	100	100	Lee et al. Genome Bio 2014			
6	W2	Wing disc	Oregon R	modENCODE_5495	W2	SRR497730	75	30.5	75	75	Lee et al. Genome Bio 2014			
7	CME W1 Cl.8+	L3 wing disc	Oregon R	modENCODE_5480	Cl.8	SRR497726	75	18.8	75	75	Lee et al. Genome Bio 2014			
8	CME L1	L3 leg disc	Oregon R	modENCODE_5489	L1	SRR497712	100	46.9	100	100	Lee et al. Genome Bio 2014			
9	OSS_C	Ovary	?	n/a	OSS_C	SRR1185816	150	33.7	150	na	Sytnikova et al. Genome Res 2014			
10	OSS_E	Ovary	?	n/a	OSS_E	SRR1185793	150	26.6	150	na	Sytnikova et al. Genome Res 2014			
11	OSC_C	Ovary	?	n/a	OSC_C	SRR609664	100	65.7	100	na	Sienski et al. Cell 2013			
12	OSC_E	Ovary	?	n/a	OSC_C	SRR1185771	150	24.6	150	na	Sytnikova et al. Genome Res 2014			
13	ML-DmD16-c3	L3 wing disc	y^1 v^1 f^1 mal^{F1}	modENCODE_5482	D16-c3	SRR497710	100	36.6	100	100	Lee et al. Genome Bio 2014		SRR497715	2x75
14	ML-DmD17-c3	L3 haltere disc	y^1 v^1 f^1 mal^{F1}	modENCODE_5483	D17-c3	SRR497725	100	43.4	100	100	Lee et al. Genome Bio 2014			
15	ML-DmD20-c2	L3 antennal disc	y^1 v^1 f^1 mal^{F1}	modENCODE_5484	D20-c2	SRR497724	75	26.7	75	75	Lee et al. Genome Bio 2014			
16	ML-DmD20-c5	L3 antennal disc	y^1 v^1 f^1 mal^{F1}	modENCODE_5490	D20-c5	SRR497723	100	12.5	100	100	Lee et al. Genome Bio 2014		SRR497718	2x75
17	ML-DmD4-c1	L3 mixed discs	y^1 v^1 f^1 mal^{F1}	modENCODE_5486	D4-c1	SRR497716	75	36.5	75	75	Lee et al. Genome Bio 2014			
18	ML-DmD8	L3 wing disc	y^1 v^1 f^1 mal^{F1}	modENCODE_5487	D8	SRR497729	75	29.4	75	75	Lee et al. Genome Bio 2014			
19	ML-DmD9	L3 wing disc	y^1 v^1 f^1 mal^{F1}	modENCODE_5488	D9	SRR497711	100	30.2	100	100	Lee et al. Genome Bio 2014		SRR497714	2x75
20	1182-4H	Embryo	mh	modENCODE_5479	1182-4H	SRR497717	100	21.1	100	100	Lee et al. Genome Bio 2014			
21	Mbn2	L3 hemocytes	l(2)mbn	modENCODE_5496	Mbn2	SRR497728	75	18.3	75	75	Lee et al. Genome Bio 2014			

Table S1B. Drosophila Lab Strains in the TIDAL-Fly v1 database

Number	Stock ID*	Stock Location	Genome Type	Library ID*	SRA Accession	Min Read Length	PE Reads (M)	Read Length 1	Read Length 2	Data Group	
1	Oregon R	Terry Orr-Weaver	OregonR adult mix male female	OreR-TO1	SRR1538752	125	8.6	125	125	ModEncode	
2	Oregon R	Terry Orr-Weaver	OregonR adult mix male female	OreR-TO2	SRR1538753	85	36.1	85	85	ModEncode	
3	Oregon R	Sarah Elgin	OregonR adult mix male female	OreR-SE	SRR1538754	125	27.7	125	125	ModEncode	
4	Oregon R	Peter Cherbas	OregonR adult mix male female	OreR-PB1	SRR1538751	85	38	125	85	ModEncode	
5	Oregon R	Peter Cherbas	OregonR adult mix male female	OreR-PB2	SRR1538755	125	9.9	125	125	ModEncode	
6	Oregon R	modENCODE_5497_Brent Graveley	OregonR adult virgin female	OreR-BG	SRR497727	75	19.9	75	75	ModEncode	
7	Oregon-R	Dworkin lab	OregonR adult mix male female	OreR-Dw1	SRR1104304, SRR1104305	75	16.7,18	75	75	Dworkin Lab, Genetics paper	
8	DarkFly	Fuse lab	OregonR-S mix	OreR-Dark	DRR001446	48	44.5	48	48	Izutsu Fuse PlosONE paper	
9	Samarkand	Dworkin lab	Samarkand adult mix	Samar-Dw1	SRR1104160, SRR1104161	75	12.9	75	75	Dworkin Lab, Genetics paper	
10	Canton-S substrain JC	Brembs lab, Julien Colomb from Tomas Preat	20 adult females	CanS-JC	ERR744541, ERR744542, ERR744543, ERR744544	100	8.8, 8.7, 8.7, 8.7	100	100	Brembs Lab, F1000 paper and this study	
11	Canton-S substrain BS	Brembs lab, Bruno Swinderen from Ralf Greenspan	20 adult females	CanS-BS	ERR744549, ERR744550, ERR744551, ERR744552	100	7.1, 7.1, 7.2, 7.2	100	100	Brembs Lab, F1000 paper and this study	
12	Canton-S substrain HS	Brembs lab, Henrike Scholz from Ulrike Herberlein	20 adult females	CanS-HS	ERR744545, ERR744546, ERR744547, ERR744548	100	10.3, 10.2, 10.3, 10.2	100	100	Brembs Lab, F1000 paper and this study	
13	Canton-S substrain TP	Brembs lab, Tim Tully lab, Brandeis	20 adult females	CanS-TP	ERR744537, ERR744538, ERR744539, ERR744540	100	9.4, 9.4, 9.5, 9.6	100	100	Brembs Lab, F1000 paper and this study	
14	Canton-S substrain TZ	Brembs lab, Troy Zars from Martin Heisenberg	20 adult females	CanS-TZ	ERR744533, ERR744534, ERR744535, ERR744536	100	10.2, 10.1, 10.0, 10.0	100	100	Brembs Lab, F1000 paper and this study	
15	CanS-SH	Stowers Institute, Scott Hawley lab	Adult mix, iso-2,3,X	CanS-SH	SRR2044310	150	25.5	150	150	Public deposit	
16	w[1118]	Stowers Institute, Scott Hawley lab	Adult mix, iso-2,3,X	w_1118	SRR2044312	150	27.2	150	150	Public deposit	
17	ISO1 - BL	Bergman lab, ISO1 from Bloomington	10 adult males	ISO1-BL	ERR701706, ERR701707, ERR701708, ERR701709, ERR701710, ERR701711	100	15, 14.9, 14.9, 15.1, 14.8, 15.0	100	100	Gutzwiller et al, ArXiv	
18	ISO1-CL1	Langley lab ISO1, mixed adults	mixed adult	ISO1-CL1	SRR350908	75	21.8	100	100	Langley et al (Begun), Genetics 2012	
19	ISO1-CL-HEm3	Langley lab ISO1, haploid embryos	haploid embryos	ISO1-CL-HEm3	SRR306608	100	16.4	100	100	Langley et al (Begun), Genetics 2012	
20	ISO1-CL-HEm12	Langley lab ISO1, haploid embryos	haploid embryos	ISO1-CL-HEm12	SRR097732, SRR306607	75	24.2, 19.7	75	75	Langley et al (Begun), Genetics 2012	
21	ISO1-AS-Ova	Spradling lab ISO1, Ovaries	females	ISO1-AS-Ova	SRR1519054	100	71.5	100	100	Yarosh and Spradline, GenesDev 2014	
22	ISO1-AS-Emb123	Spradling lab, ISO1 embryos	0-18hr embryos	ISO1-AS-Emb123	SRR1516224	100	81.8, 69.1, 50.5	100	100	Yarosh and Spradline, GenesDev 2014	
23	ISO1-AS-LSg123	Spradling lab ISO1, L3 Salivary gland	L3 instar	ISO1-AS-LSg123	SRR1516221, SRR1516223, SRR1516225	100	62.9, 88.4, 68.6	100	100	Yarosh and Spradline, GenesDev 2014	
24	OreR-AS-FC	Spradling lab OreR, stg10-14 follicle cells	females	OreR-AS-FC	SRR1519051	100	43.1	100	100	Yarosh and Spradline, GenesDev 2014	

Table S1C. Drosophila Lab Strains in the TIDAL-Fly v1 database

Genomes previously examined in McKay et al, Nature 2012, and Huang et al, Genome Res, 2014

PHASE 1: First Set of Libraries Analyzed for the paper. The remaining libraries in Phase 2 below are also listed in the TIDAL-Fly database

Library	Stock ID*	Stock Location	Genome Type	SRA Accession	Library ID*	SRA Accession	Min Read Length	Read Length 1	Read Length 2	Data Group	Reads (M)	Raw Read Length:Read Number		Freeze		Bloomington Stock	Synonym	NCBI SRA	Reads	Mapped Reads	Mapped Coverage	Strains
1	DGRP_362	Bloomington-25187	inbred line, mix of	SRX006288	RAL-362	SRR846985	125	merged, use min length	125.0	DGRP-F1 + F2	56.3	45bp:18900352, 75bp:19241828, 12	DGR	F1 + F2	Blo	25187	RAL-362	SRX006288 and	70.4	70,371,466	39.4	1
2	DGRP_517	Bloomington-25197	inbred line, mix of	SRX024363	RAL-517	SRR933579	75	merged, use min length	75.0	DGRP-F1	59.8	45bp:75005702, 75bp:44531158	DGR	F1	Blo	25197	RAL-517	SRX024363 and	115.8	115,773,424	37	2
3	DGRP_765	Bloomington-25204	inbred line, mix of	SRX006169	RAL-765	SRR933595	75	75.0	75.0	DGRP-F1	19.9	75bp:39893442	DGR	F1	Blo	25204	RAL-765	SRX006169 and	38.1	38,107,800	16.3	3
4	DGRP_109	Bloomington-28140	inbred line, mix of	SRX020746	RAL-109_75b	SRR835219	75	merged, use min length	125.0	DGRP-F1	44.8	75bp:28899572, 125bp:60695744	DGR	F1	Blo	28140	RAL-109	SRX020746 and	78.6	78,598,946	48.7	4
5	DGRP_109	Bloomington-28140	inbred line, mix of	SRX020746	RAL-109_125b	SRR835219	125	merged, use min length	125.0	DGRP-F1	44.8	75bp:28899572, 125bp:60695744	DGR	F1	Blo	28140	RAL-109	SRX020746 and	78.6	78,598,946	48.7	4
6	DGRP_229	Bloomington-29653	inbred line, mix of	SRX021052	RAL-229_75b	SRR835221	75	merged, use min length	125.0	DGRP-F1	49.6	75bp:27785294, 125bp:71504164	DGR	F1	Blo	29653	RAL-229	SRX021052	83.9	83,937,442	52.7	5
7	DGRP_229	Bloomington-29653	inbred line, mix of	SRX021052	RAL-229_125b	SRR835221	125	merged, use min length	125.0	DGRP-F1	49.6	75bp:27785294, 125bp:71504164	DGR	F1	Blo	29653	RAL-229	SRX021052	83.9	83,937,442	52.7	5
8	DGRP_301	Bloomington-25175	inbred line, mix of	SRX005978	RAL-301_75b	SRR835223	75	merged, use min length	125.0	DGRP-F2	73.6	45bp:13099368, 64bp:10588312, 75	DGR	F2	Blo	25175	RAL-301	SRX155995 and	133.5	133,504,636	71.7	6
9	DGRP_301	Bloomington-25175	inbred line, mix of	SRX005978	RAL-301_125b	SRR835223	125	merged, use min length	125.0	DGRP-F2	73.6	45bp:13099368, 64bp:10588312, 75	DGR	F2	Blo	25175	RAL-301	SRX155995 and	133.5	133,504,636	71.7	6
10	DGRP_303	Bloomington-25176	inbred line, mix of	SRX005986	RAL-303_100b	SRR835228	100	merged, use min length	100.0	DGRP-F2	161.6	45bp:79536584, 64bp:4167822, 100	DGR	F2	Blo	25176	RAL-303	SRX155978 and	287.4	287,358,990	148	7
11	DGRP_303	Bloomington-25176	inbred line, mix of	SRX005986	RAL-303_125b	SRR835228	125	merged, use min length	100.0	DGRP-F2	161.6	45bp:79536584, 64bp:4167822, 100	DGR	F2	Blo	25176	RAL-303	SRX155978 and	287.4	287,358,990	148	7
12	DGRP_304	Bloomington-25177	inbred line, mix of	SRX005988	RAL-304_75b	SRR835236	75	merged, use min length	100.0	DGRP-F2	142.5	45bp:25081192, 64bp:7071260, 75	DGR	F2	Blo	25177	RAL-304	SRX156009 and	252.1	252,108,056	133.6	8
13	DGRP_304	Bloomington-25177	inbred line, mix of	SRX005988	RAL-304_100b	SRR835236	100	merged, use min length	100.0	DGRP-F2	142.5	45bp:25081192, 64bp:7071260, 75	DGR	F2	Blo	25177	RAL-304	SRX156009 and	252.1	252,108,056	133.6	8
14	DGRP_304	Bloomington-25177	inbred line, mix of	SRX005988	RAL-304_125b	SRR835236	125	merged, use min length	100.0	DGRP-F2	142.5	45bp:25081192, 64bp:7071260, 75	DGR	F2	Blo	25177	RAL-304	SRX156009 and	252.1	252,108,056	133.6	8
15	DGRP_306	Bloomington-37525	inbred line, mix of	SRX006140	RAL-306_75b	SRR835242	75	merged, use min length	125.0	DGRP-F2	63.2	45bp:29368112, 64bp:8510324, 75	DGR	F2	Blo	37525	RAL-306	SRX156007 and	100.4	100,446,332	50.7	9
16	DGRP_306	Bloomington-37525	inbred line, mix of	SRX006140	RAL-306_125b	SRR835242	125	merged, use min length	125.0	DGRP-F2	63.2	45bp:29368112, 64bp:8510324, 75	DGR	F2	Blo	37525	RAL-306	SRX156007 and	100.4	100,446,332	50.7	9
17	DGRP_307	Bloomington-25179	inbred line, mix of	SRX006186	RAL-307_75b	SRR835247	75	merged, use min length	75.0	DGRP-F2	144.8	36bp:9641364, 45bp:17304248, 75	DGR	F2	Blo	25179	RAL-307	SRX156012 and	252.2	252,236,309	142.1	10
18	DGRP_307	Bloomington-25179	inbred line, mix of	SRX006186	RAL-307_100b	SRR835247	100	merged, use min length	100.0	DGRP-F2	144.8	36bp:9641364, 45bp:17304248, 75	DGR	F2	Blo	25179	RAL-307	SRX156012 and	252.2	252,236,309	142.1	10
19	DGRP_307	Bloomington-25179	inbred line, mix of	SRX006186	RAL-307_125b	SRR835247	125	merged, use min length	125.0	DGRP-F2	144.8	36bp:9641364, 45bp:17304248, 75	DGR	F2	Blo	25179	RAL-307	SRX156012 and	252.2	252,236,309	142.1	10
20	DGRP_315	Bloomington-25181	inbred line, mix of	SRX006143	RAL-315_75b	SRR835252	75	merged, use min length	125.0	DGRP-F2	63.1	45bp:48481060, 75bp:20407822, 85	DGR	F2	Blo	25181	RAL-315	SRX156010 and	114.4	114,404,766	51	11
21	DGRP_315	Bloomington-25181	inbred line, mix of	SRX006143	RAL-315_125b	SRR835252	125	merged, use min length	125.0	DGRP-F2	63.1	45bp:48481060, 75bp:20407822, 85	DGR	F2	Blo	25181	RAL-315	SRX156010 and	114.4	114,404,766	51	11
22	DGRP_340	Bloomington-28174	inbred line, mix of	SRX156030	RAL-340_100b	SRR835939	100	merged, use min length	100.0	DGRP-F2	103.7	75bp:41788984, 100bp:165697712	DGR	F2	Blo	28174	RAL-340	SRX156030	130.2	130,158,488	73.8	12
23	DGRP_371	Bloomington-28183	inbred line, mix of	SRX021257	RAL-371_75b	SRR835326	75	merged, use min length	125.0	DGRP-F1 + F2	42.6	75bp:30869276, 125bp:27187181	DGR	F1 + F2	Blo	28183	RAL-371	SRX021257 and	78.6	78,638,404	42.5	13
24	DGRP_371	Bloomington-28183	inbred line, mix of	SRX021257	RAL-371_125b	SRR835326	125	merged, use min length	125.0	DGRP-F1 + F2	42.6	75bp:30869276, 125bp:27187181	DGR	F1 + F2	Blo	28183	RAL-371	SRX021257 and	78.6	78,638,404	42.5	13
25	DGRP_373	Bloomington-28184	inbred line, mix of	SRX023425	RAL-373_95b	SRR835329	95	merged, use min length	125.0	DGRP-F1	42.8	95bp:40216244, 125bp:45373906	DGR	F1	Blo	28184	RAL-373	SRX023425 and	76.7	76,690,462	48.3	14
26	DGRP_373	Bloomington-28184	inbred line, mix of	SRX023425	RAL-373_125b	SRR835329	125	merged, use min length	125.0	DGRP-F1	42.8	95bp:40216244, 125bp:45373906	DGR	F1	Blo	28184	RAL-373	SRX023425 and	76.7	76,690,462	48.3	14
27	DGRP_409	Bloomington-28278	inbred line, mix of	SRX021243	RAL-409_95b	SRR835331	95	merged, use min length	125.0	DGRP-F1	38.7	95bp:33305170, 125bp:44122246	DGR	F1	Blo	28278	RAL-409	SRX021243 and	55.1	55,099,676	35.7	15
28	DGRP_409	Bloomington-28278	inbred line, mix of	SRX021243	RAL-409_125b	SRR835331	125	merged, use min length	125.0	DGRP-F1	38.7	95bp:33305170, 125bp:44122246	DGR	F1	Blo	28278	RAL-409	SRX021243 and	55.1	55,099,676	35.7	15
29	DGRP_584	Bloomington-28212	inbred line, mix of	SRX155987	RAL-584_95b	SRR835942	95	merged, use min length	125.0	DGRP-F2	70	95bp:93885100, 125bp:46038390	DGR	F2	Blo	28212	RAL-584	SRX155987 and	79.7	79,663,962	49.4	16
30	DGRP_584	Bloomington-28212	inbred line, mix of	SRX155987	RAL-584_125b	SRR835942	125	merged, use min length	125.0	DGRP-F2	70	95bp:93885100, 125bp:46038390	DGR	F2	Blo	28212	RAL-584	SRX155987 and	79.7	79,663,962	49.4	16
31	DGRP_748	Bloomington-28224	inbred line, mix of	SRX156019	RAL-748_95b	SRR835336	95	merged, use min length	125.0	DGRP-F2	43.7	95bp:36957084, 125bp:50455402	DGR	F2	Blo	28224	RAL-748	SRX156019 and	57.4	57,410,828	38.3	17
32	DGRP_748	Bloomington-28224	inbred line, mix of	SRX156019	RAL-748_125b	SRR835336	125	merged, use min length	125.0	DGRP-F2	43.7	95bp:36957084, 125bp:50455402	DGR	F2	Blo	28224	RAL-748	SRX156019 and	57.4	57,410,828	38.3	17
33	DGRP_801	Bloomington-28234	inbred line, mix of	SRX021391	RAL-801_95b	SRR835341	95	merged, use min length	125.0	DGRP-F1	43.4	95bp:32922146, 125bp:53899816	DGR	F1	Blo	28234	RAL-801	SRX021391 and	59.3	59,280,892	38.4	18
34	DGRP_801	Bloomington-28234	inbred line, mix of	SRX021391	RAL-801_125b	SRR835341	125	merged, use min length	125.0	DGRP-F1	43.4	95bp:32922146, 125bp:53899816	DGR	F1	Blo	28234	RAL-801	SRX021391 and	59.3	59,280,892	38.4	18
35	DGRP_802	Bloomington-28235	inbred line, mix of	SRX025318	RAL-802_95b	SRR835343	95	merged, use min length	125.0	DGRP-F1	47.1	95bp:27114342, 125bp:67142296	DGR	F1	Blo	28235	RAL-802	SRX025318 and	78.6	78,579,930	52.4	19
36	DGRP_802	Bloomington-28235	inbred line, mix of	SRX025318	RAL-802_125b	SRR835343	125	merged, use min length	125.0	DGRP-F1	47.1	95bp:27114342, 125bp:67142296	DGR	F1	Blo	28235	RAL-802	SRX025318 and	78.6	78,579,930	52.4	19
37	DGRP_808	Bloomington-28238	inbred line, mix of	SRX021402	RAL-808_75b	SRR835345	75	merged, use min length	125.0	DGRP-F1	53.8	75bp:44482188, 125bp:63083538	DGR	F1	Blo	28238	RAL-808	SRX021402 and	83.0	82,954,992	49.3	20
38	DGRP_808	Bloomington-28238	inbred line, mix of	SRX021402	RAL-808_125b	SRR835345	125	merged, use min length	125.0	DGRP-F1	53.8	75bp:44482188, 125bp:63083538	DGR	F1	Blo	28238	RAL-808	SRX021402 and	83.0	82,954,992	49.3	20
39	DGRP_821	Bloomington-28243	inbred line, mix of	SRX155991	RAL-821_100b	SRR835347	100	merged, use min length	100.0	DGRP-F2	108.2	100bp:132570414, 125bp:83747778	DGR	F2	Blo	28243	RAL-821	SRX155990 and	179.3	179,277,520	112	21
40	DGRP_821	Bloomington-28243	inbred line, mix of	SRX155991	RAL-821_125b	SRR835347	125	merged, use min length	100.0	DGRP-F2	108.2	100bp:132570414, 125bp:83747778	DGR	F2	Blo	28243	RAL-821	SRX155990 and	179.3	179,277,520	112	21
41	DGRP_853	Bloomington-28250	inbred line, mix of	SRX021491	RAL-853_110b	SRR835349	110	merged, use min length	125.0	DGRP-F2	55.3	110bp:42021476, 125bp:68611046	DGR	F2	Blo	28250	RAL-853	SRX021491 and	69.8	69,801,922	48.9	22
42	DGRP_853	Bloomington-28250	inbred line, mix of	SRX021491	RAL-853_125b	SRR835349	125	merged, use min length	125.0	DGRP-F2	55.3	110bp:42021476, 125bp:68611046	DGR	F2	Blo	28250	RAL-853	SRX021491 and	69.8	69,801,922	48.9	22
43	DGRP_900	Bloomington-28261	inbred line, mix of	SRX156023	RAL-900_95b	SRR846989	95	merged, use min length	125.0	DGRP-F2	47.6	95bp:30160598, 125bp:65136580	DGR	F2	Blo	28261	RAL-900	SRX156023	77.4	77,355,512	52.3	23
44	DGRP_900	Bloomington-28261	inbred line, mix of	SRX156023	RAL-900_125b	SRR846989	125	merged, use min length	125.0	DGRP-F2	47.6	95bp:30160598, 125bp:65136580	DGR	F2	Blo	28261	RAL-900	SRX156023	77.4	77,355,512	52.3	23
45	DGRP_382	Bloomington-28189	inbred line, mix of	SRX156013	RAL-382	SRR834552	125	125	125	DGRP-F2	36.9	125bp:73812254	DGR	F2	Blo	28189	RAL-382	SRX156013	57.6	57,592,186	41.1	24
46	DGRP_223	Bloomington-28155	inbred line, mix of	SRX155994	RAL-223	SRR834527	125	125	125	DGRP-F2	35.6	125bp:71152512	DGR	F2	Blo	28155	RAL-223	SRX155994	57.2	57,181,894	40.8	25
47	DGRP_850	Bloomington-28249	inbred line, mix of	SRX155993	RAL-850	SRR835061	125	125	125	DGRP-F2	34.8	125bp:69699750	DGR	F2	Blo	28249	RAL-850	SRX155993	61.0	60,987,804	43.6	26
48	DGRP_913	Bloomington-28265	inbred line, mix of	SRX156024	RAL-913	SRR835077	125	125	125	DGRP-F2	34.6	125bp:69250292	DGR	F2	Blo	28265	RAL-913	SRX156024	61.2	61,221,716	43.7	27
49	DGRP_843	Bloomington-28247	inbred line, mix of	SRX156036	RAL-843	SRR835059	125	125	125	DGRP-F2	34.3	125bp:68658714	DGR	F2	Blo	28247	RAL-843	SRX156036	59.2	59,174,050	42.3	28
50	DGRP_361	Bloomington-28180	inbred line, mix of	SRX155984	RAL-361	SRR834553	125	125	125	DGRP-F2	34.1	125bp:68254340	DGR	F2	Blo	28180	RAL-361	SRX155984	56.9	56,897,940	40.6	29
51	DGRP_189	Bloomington-28152	inbred line, mix of	SRX155979	RAL-189	SRR834523	125	125	125	DGRP-F2	31.6	125bp:63289120	DGR	F2	Blo	28152	RAL-189	SRX155979	52.9	52,919,862	37.8	30
52	DGRP_849	Bloomington-28248	inbred line, mix of	SRX156035	RAL-849	SRR835060	125	125	125	DGRP-F2	30.8	125bp:61687178	DGR	F2	Blo	28248	RAL-849	SRX156035	55.8	55,797,340	39.9	31
53	DGRP_142	Bloomington-28144	inbred line, mix of	SRX020759	RAL-142	SRR834551	100	110	110	DGRP-F1	20.6	110bp:41167794	DGR	F1	Blo	28144	RAL-142	SRX020759	31.4	31,409,058	19.7	32
54	DGRP_822	Bloomington-28244	inbred line, mix of	SRX021476	RAL-822	SRR835055	100	110	110	DGRP-F1	20.5	110bp:41079524	DGR	F1	Blo	28244	RAL-822	SRX021476	28.1	28,098,026	17.7	33
55	DGRP_75	Bloomington-28132	inbred line, mix of	SRX021384	RAL-75	SRR835087	100	110	110	DGRP-F1	19.1	110bp:38161744	DGR	F1	Blo	28132	RAL-75	SRX021384	29.5	29,460,700	18.5	34
56	DGRP_395	Bloomington-28194	inbred line, mix of	SRX156015	RAL-395	SRR834521	100	101	101	DGRP-F2	43.6	101bp:87233368	DGR	F2	Blo	28194	RAL-395	SRX156015	81.5	81,549,134	47.1	35
57	DGRP_819	Bloomington-28242	inbred line, mix of	SRX156006	RAL-819	SRR835054	100	100	100	DGRP-F2	75.4	100bp:150745358	DGR	F2	Blo	28242	RAL-819	SRX156006	127.8	127,803,094	73	36
58	DGRP_57	Bloomington-29652	inbred line, mix of	SRX021296	RAL-57	SRR933581	100	100	100	DGRP-F1	32.5	100bp:64966990	DGR	F1	Blo	29652	RAL-57	SRX021296	57.0	57,017,248	32.6	37
59	DGRP_138	Bloomington-28143	inbred line, mix of	SRX021008	RAL-138	SRR932121	100	100	100	DGRP-F1	30.8	100bp:61689820	DGR	F1	Blo	28143	RAL-138	SRX021008	52.6	52,639,192	30.1	38
60	DGRP_324	Bloomington-25182	inbred line, mix of	SRX006145	RAL-324_100	SRR835256	100	merged, use min length	100	DGRP-F2	127.3	45bp:79789940, 100bp:174874844	DGR	F2	Blo	25182	RAL-324	SRX006144 and	233.9	233,891,168	109.2	39
61	DGRP_360	Bloomington-25186	inbred line, mix of	SRX155999	RAL-360_125	SRR835260	125	merged, use min length	125	DGRP-F2	87.3	36bp:59319927, 125bp:55959646	DGR	F2	Blo	25186	RAL-360	SRX155999 and	85.6	85,563,461	38.2	40
62	DGRP_437	Bloomington-25194	inbred line, mix of	SRX006156	RAL-437_125	SRR835333	125	merged, use min length	125	DGRP-F1	46.9	45bp:43208064, 125bp:50502476	DGR	F1	Blo	25194	RAL-437	SRX006156 and	83.9	83,861,754	40.4	41
63	DGRP_774	Bloomington-25205	inbred line, mix of	SRX006170	RAL-774_125	SRR835338	125	merged, use min length	125	DGRP-F1	38.2	45bp:42434168, 125bp:33991884	DGR	F1	Blo	25205	RAL-774	SRX006170 and	68.9	68,894,652	30	42
64	DGRP_391A	Bloomington-25191	inbred line, mix of	SRX006152	RAL-391_75b	SRR018294	75	merged, use min length	75	DGRP-F1 + F2	9.5	45bp:40955170, 75bp:18124720, 95	DGR	F1 + F2	Blo	25191	RAL-391	SRX023452,SR	95.3	95,313,294	38	43
65	DGRP_391B	Bloomington-25191	inbred line, mix of	SRX023452	RAL-391_95b	SRR060098	95	merged, use min length	95	DGRP-F1 + F2	19	45bp:40955170, 75bp:18124720, 95	DGR	F1 + F2	Blo	25191	RAL-391	SRX023452,SR	95.3	95,313,294	38	44
66	DGRP_380	Bloomington-25190	inbred line, mix of	SRX006300	RAL-380	SRR018591, SR	75	merged, use min length	75	DGRP-F1 + F2	10.2, 10.3	45bp:505064, 74bp:20092700, 75b	DGR	F1 + F2	Blo	25190	RAL-380	SRX006303,SR	69.2	69,152,136	29.5	45
67	DGRP_176	Bloomington-28149	inbred line, mix of	SRX020762, SR	RAL-176	SRR048925, SR	75	75	75	DGRP-F1	19.7, 20.3	75bp:79996860	DGR	F1	Blo	28149	RAL-176	SRX020762, SR	73.7	73,734,058	31.6	46

69	DGRP_358	Bloomington-25185	inbred line, mix of	SRX006282, SR	RAL-358	SRR018574, SR	75.0	merged, use min length	75.0	DGRP-F1	6.8, 14.1	45bp:18011532, 75bp:27739857	DGR	F1	Blo	25185	RAL-358	SRX006283 and	42.5	42,499,891	15.2	47	
70	DGRP_399	Bloomington-25192	inbred line, mix of	SRX006153	RAL-399	SRR933575	75.0	merged, use min length	75.0	DGRP-F1	27.8	45bp:41011932, 75bp:14665138	DGR	F1	Blo	25192	RAL-399	SRX006154 and	51.6	51,632,558	15.5	48	
71	DGRP_313	Bloomington-25180	inbred line, mix of	SRX006277	RAL-313	SRR018519	75.0	single end only	75.0	DGRP-F1 + F2	14	36bp:9477876, 45bp:18165034, 75l	DGR	F1 + F2	Blo	25180	RAL-313	SRX006277,SR	39.1	39,060,649	11.7	49	
72	DGRP_375	Bloomington-25188	inbred line, mix of	SRX006148	RAL-375	SRR933572	75.0	merged, use min length	75.0	DGRP-F1	51.8	45bp:52106222, 75bp:51413814	DGR	F1	Blo	25188	RAL-375	SRX006150,SR	99.3	99,258,956	33.8	50	
73	DGRP_555	Bloomington-25198	inbred line, mix of	SRX006159	RAL-555	SRR933580	75	75	75	DGRP-F1	25.1	75bp:50103810	DGR	F1	Blo	25198	RAL-555	SRX006159	44.9	44,883,398	19.2	51	
74	DGRP_705	Bloomington-25744	inbred line, mix of	SRX006162	RAL-705	SRR933585	75	75	75	DGRP-F1	23.5	75bp:47006608	DGR	F1	Blo	25744	RAL-705	SRX006162	38.9	38,920,070	16.7	52	
75	DGRP_707	Bloomington-25200	inbred line, mix of	SRX006163	RAL-707	SRR933586	75	75	75	DGRP-F1	23.3	75bp:46657404	DGR	F1	Blo	25200	RAL-707	SRX006163	41.6	41,568,952	17.8	53	
76	DGRP_712	Bloomington-25201	inbred line, mix of	SRX006164	RAL-712	SRR933587	75	75	75	DGRP-F1	22.3	75bp:44687868	DGR	F1	Blo	25201	RAL-712	SRX006164	37.9	37,949,898	16.3	54	
77	DGRP_714	Bloomington-25745	inbred line, mix of	SRX006166	RAL-714	SRR933588	75.0	75.0	75.0	DGRP-F1	21	75bp:42090922	DGR	F1	Blo	25745	RAL-714	SRX006166 and	40.0	39,955,902	17.1	55	
78	DGRP_732	Bloomington-25203	inbred line, mix of	SRX006167	RAL-732	SRR933591	75	75	75	DGRP-F1	21.1	75bp:42170344	DGR	F1	Blo	25203	RAL-732	SRX006167	38.1	38,082,012	16.3	56	
79	DGRP_379	Bloomington-25189	inbred line, mix of	SRX006293	RAL-379	SRR018583, SR	75.0	single end only	75.0	DGRP-F1	6.1, 12.4	45bp:19006960, 75bp:18496338	DGR	F1	Blo	25189	RAL-379	SRX006293 and	34.0	34,014,707	11.4	57	
80	DGRP_340	Bloomington-28174	inbred line, mix of	SRX156030	RAL-340_75b	SRR835939	75	merged, use min length	100.0	DGRP-F2	103.7	75bp:41788984, 100bp:165697712	DGR	F2	Blo	28174	RAL-340	SRX156030	130.2	130,158,488	73.8	12	This entry repeatedly gen

PHASE 2: Additional DGRP Illumina-sequenced libaries compatible for TIDAL analysis loaded into TIDAL-Fly database but not part of analysis in paper.

Library	Stock ID*	Stock Location	Genome Type	SRA Accession	Library ID*	SRA Accession	Min Read Length	Read Length 1	Read Length 2	Data Group	Reads (M)	Raw Read Length:Read Number		Freeze		Bloomington Stock	Synonym	NCBI SRA	Reads	Mapped Reads	Mapped Coverage	Strains
1	DGRP_365	Bloomington-25445	inbred line, mix of	SRX006290, SR	RAL-365	SRR018580, SRR	75	75.0	75.0	DGRP-F1	10.1	45bp:18394996, 75bp:32783196	DGR	F1	Blo	25445	RAL-365	SRX006291,SR	42.9	42,859,487	15.6	58
3	DGRP_852	Bloomington-25209	inbred line, mix of	SRX006304, SR	RAL-852	SRR018600	75	75.0	75.0	DGRP-F1	11.7	45bp:41788724, 75bp:59375261	DGR	F1	Blo	25209	RAL-852	SRX006304 and	90.6	90,552,271	32.2	59
7	DGRP_730	Bloomington-25202	inbred line, mix of	SRX006308	RAL-730	SRR933590	75	75	75	DGRP-F1	38	75bp:38019145	DGR	F1	Blo	25202	RAL-730	SRX006308	30.9	30,891,608	13.2	60
8	DGRP_738	Bloomington-28223	inbred line, mix of	SRX021383	RAL-738	SRR933593	75	75	75	DGRP-F1	37.9	75bp:75804508	DGR	F1	Blo	28223	RAL-738	SRX021383	63.1	63,119,910	27.1	61
9	DGRP_737	Bloomington-28222	inbred line, mix of	SRX023451	RAL-737	SRR933592	75	75	75	DGRP-F1	37.4	75bp:74740132	DGR	F1	Blo	28222	RAL-737	SRX023451	58.5	58,531,732	25.1	62
10	DGRP_757	Bloomington-28226	inbred line, mix of	SRX021385	RAL-757	SRR933594	75	75	75	DGRP-F1	37.2	75bp:74326240	DGR	F1	Blo	28226	RAL-757	SRX021385	66.2	66,207,766	28.4	63
11	DGRP_727	Bloomington-28221	inbred line, mix of	SRX021382	RAL-727	SRR933589	75	75	75	DGRP-F1	36.9	75bp:73781476	DGR	F1	Blo	28221	RAL-727	SRX021382	64.2	64,202,850	27.5	64
12	DGRP_897	Bloomington-28260	inbred line, mix of	SRX023457	RAL-897	SRR933601	75	75	75	DGRP-F1	35.4	75bp:70892788	DGR	F1	Blo	28260	RAL-897	SRX023457	63.0	63,033,690	27	65
13	DGRP_332	Bloomington-28171	inbred line, mix of	SRX021095	RAL-332	SRR933569	75	75	75	DGRP-F1	32.8	75bp:65583082	DGR	F1	Blo	28171	RAL-332	SRX021095	60.0	59,985,078	25.7	66
14	DGRP_181	Bloomington-28151	inbred line, mix of	SRX020912	RAL-181	SRR933563	75	75	75	DGRP-F1	32	75bp:64093862	DGR	F1	Blo	28151	RAL-181	SRX020912	57.6	57,566,276	24.7	67
15	DGRP_325	Bloomington-28170	inbred line, mix of	SRX021793	RAL-325	SRR933568	75	75	75	DGRP-F1	31.5	75bp:31457046	DGR	F1	Blo	28170	RAL-325	SRX021793	26.3	26,299,389	11.3	68
16	DGRP_381	Bloomington-28188	inbred line, mix of	SRX021112	RAL-381	SRR933573	75	75	75	DGRP-F1	27.2	75bp:54335852	DGR	F1	Blo	28188	RAL-381	SRX021112	48.7	48,708,464	20.9	69
17	DGRP_153	Bloomington-28146	inbred line, mix of	SRX021514	RAL-153	SRR933562	75	75	75	DGRP-F1	26.8	75bp:26821499	DGR	F1	Blo	28146	RAL-153	SRX021514	21.8	21,768,014	9.3	70
18	DGRP_352	Bloomington-28177	inbred line, mix of	SRX021101	RAL-352	SRR834516	75	75	75	DGRP-F1	22.5	75bp:44982388	DGR	F1	Blo	28177	RAL-352	SRX021101	36.3	36,302,672	15.6	71
19	DGRP_882	Bloomington-28255	inbred line, mix of	SRX021496	RAL-882	SRR835067	75	75	75	DGRP-F1	22.4	75bp:44722234	DGR	F1	Blo	28255	RAL-882	SRX021496	40.6	40,585,544	17.4	72
20	DGRP_879	Bloomington-28254	inbred line, mix of	SRX021494	RAL-879	SRR835066	75	75	75	DGRP-F1	21.7	75bp:43485194	DGR	F1	Blo	28254	RAL-879	SRX021494	30.5	30,475,860	13.1	73
21	DGRP_805	Bloomington-28237	inbred line, mix of	SRX021400	RAL-805	SRR835095	75	75	75	DGRP-F1	21.6	75bp:43182102	DGR	F1	Blo	28237	RAL-805	SRX021400	37.7	37,672,090	16.1	74
22	DGRP_356	Bloomington-28178	inbred line, mix of	SRX023833	RAL-356	SRR834537	75	75	75	DGRP-F1	21.5	75bp:42903612	DGR	F1	Blo	28178	RAL-356	SRX023833	36.1	36,113,766	15.5	75
23	DGRP_884	Bloomington-28256	inbred line, mix of	SRX021498	RAL-884	SRR835068	75	75	75	DGRP-F1	21.3	75bp:42658114	DGR	F1	Blo	28256	RAL-884	SRX021498	32.6	32,585,778	14	76
24	DGRP_890	Bloomington-28257	inbred line, mix of	SRX021499	RAL-890	SRR835071	75	75	75	DGRP-F1	21.0	75bp:41954706	DGR	F1	Blo	28257	RAL-890	SRX021499	37.1	37,101,876	15.9	77
25	DGRP_861	Bloomington-28253	inbred line, mix of	SRX021493	RAL-861	SRR835065	75	75	75	DGRP-F1	20.6	75bp:41295320	DGR	F1	Blo	28253	RAL-861	SRX021493	29.5	29,478,054	12.6	78
26	DGRP_83	Bloomington-28134	inbred line, mix of	SRX023456	RAL-83	SRR835058	75	75	75	DGRP-F1	20.5	75bp:41070470	DGR	F1	Blo	28134	RAL-83	SRX023456	38.0	37,982,090	16.3	79
27	DGRP_491	Bloomington-28202	inbred line, mix of	SRX021268	RAL-491	SRR835035	75	75	75	DGRP-F1	20.5	75bp:40944392	DGR	F1	Blo	28202	RAL-491	SRX021268	35.3	35,336,140	15.1	80
28	DGRP_761	Bloomington-28227	inbred line, mix of	SRX021386	RAL-761	SRR835088	75	75	75	DGRP-F1	20.4	75bp:40867250	DGR	F1	Blo	28227	RAL-761	SRX021386	35.6	35,577,354	15.2	81
29	DGRP_149	Bloomington-28145	inbred line, mix of	SRX020760	RAL-149	SRR834550	75	75	75	DGRP-F1	20.3	75bp:40637646	DGR	F1	Blo	28145	RAL-149	SRX020760	33.9	33,914,600	14.5	82
30	DGRP_535	Bloomington-28208	inbred line, mix of	SRX021293	RAL-535	SRR835046	75	75	75	DGRP-F1	20.1	75bp:40234802	DGR	F1	Blo	28208	RAL-535	SRX021293	35.4	35,362,252	15.2	83
31	DGRP_776	Bloomington-28229	inbred line, mix of	SRX021387	RAL-776	SRR835089	75	75	75	DGRP-F1	19.9	75bp:39890986	DGR	F1	Blo	28229	RAL-776	SRX021387	36.4	36,432,512	15.6	84
32	DGRP_787	Bloomington-28231	inbred line, mix of	SRX021388	RAL-787	SRR835091	75	75	75	DGRP-F1	19.9	75bp:39795416	DGR	F1	Blo	28231	RAL-787	SRX021388	35.9	35,896,250	15.4	85
33	DGRP_136	Bloomington-28142	inbred line, mix of	SRX020753	RAL-136	SRR834542	75	75	75	DGRP-F1	19.5	75bp:38970964	DGR	F1	Blo	28142	RAL-136	SRX020753	33.9	33,947,560	14.5	86
34	DGRP_318	Bloomington-28168	inbred line, mix of	SRX021082	RAL-318	SRR834507	75	75	75	DGRP-F1	19.5	75bp:39068236	DGR	F1	Blo	28168	RAL-318	SRX021082	35.4	35,366,996	15.2	87
35	DGRP_804	Bloomington-28236	inbred line, mix of	SRX021399	RAL-804	SRR835094	75	75	75	DGRP-F1	19.5	75bp:38910810	DGR	F1	Blo	28236	RAL-804	SRX021399	34.5	34,482,650	14.8	88
36	DGRP_812	Bloomington-28240	inbred line, mix of	SRX021419	RAL-812	SRR835052	75	75	75	DGRP-F1	19.4	75bp:38719004	DGR	F1	Blo	28240	RAL-812	SRX021419	37.5	37,540,214	16.1	89
37	DGRP_796	Bloomington-28233	inbred line, mix of	SRX021390	RAL-796	SRR835093	75	75	75	DGRP-F1	19.3	75bp:38662002	DGR	F1	Blo	28233	RAL-796	SRX021390	34.7	34,681,678	14.9	90
38	DGRP_509	Bloomington-28206	inbred line, mix of	SRX021273	RAL-509	SRR835041	75	75	75	DGRP-F1	19.0	75bp:38095912	DGR	F1	Blo	28206	RAL-509	SRX021273	35.6	35,615,726	15.3	91
39	DGRP_818	Bloomington-28241	inbred line, mix of	SRX021478	RAL-818	SRR835053	75	75	75	DGRP-F1	18.8	75bp:37531016	DGR	F1	Blo	28241	RAL-818	SRX021478	34.2	34,210,842	14.7	92
40	DGRP_859	Bloomington-25210	inbred line, mix of	SRX006176	RAL-859	SRR933600	75	75	75	DGRP-F1	18.5	75bp:36919488	DGR	F1	Blo	25210	RAL-859	SRX006176	34.0	33,950,076	14.6	93
41	DGRP_810	Bloomington-28239	inbred line, mix of	SRX021418	RAL-810	SRR835051	75	75	75	DGRP-F1	18.5	75bp:36972402	DGR	F1	Blo	28239	RAL-810	SRX021418	36.1	36,075,996	15.5	94
42	DGRP_783	Bloomington-28230	inbred line, mix of	SRX023455	RAL-783	SRR835090	75	75	75	DGRP-F1	18.3	75bp:36692434	DGR	F1	Blo	28230	RAL-783	SRX023455	31.4	31,444,902	13.5	95
43	DGRP_45	Bloomington-28128	inbred line, mix of	SRX021261	RAL-45	SRR835032	75	75	75	DGRP-F1	18.2	75bp:36472946	DGR	F1	Blo	28128	RAL-45	SRX021261	33.0	32,984,934	14.1	96
44	DGRP_59	Bloomington-28129	inbred line, mix of	SRX021327	RAL-59	SRR835081	75	75	75	DGRP-F1	18.2	75bp:36484216	DGR	F1	Blo	28129	RAL-59	SRX021327	34.9	34,871,104	14.9	97
45	DGRP_280	Bloomington-28164	inbred line, mix of	SRX021058	RAL-280	SRR834534	75	75	75	DGRP-F1	17.8	75bp:35581218	DGR	F1	Blo	28164	RAL-280	SRX021058	33.0	32,968,684	14.1	98
46	DGRP_595	Bloomington-28215	inbred line, mix of	SRX021328	RAL-595	SRR835101	75	75	75	DGRP-F1	17.8	75bp:35536106	DGR	F1	Blo	28215	RAL-595	SRX021328	29.2	29,176,032	12.5	99
47	DGRP_93	Bloomington-28137	inbred line, mix of	SRX021504	RAL-93	SRR835080	75	75	75	DGRP-F1	17.6	75bp:35127632	DGR	F1	Blo	28137	RAL-93	SRX021504	33.9	33,871,000	14.5	100
48	DGRP_195	Bloomington-28153	inbred line, mix of	SRX021039	RAL-195	SRR834524	75	75	75	DGRP-F1	17.6	75bp:35114464	DGR	F1	Blo	28153	RAL-195	SRX021039	33.4	33,435,584	14.3	101
49	DGRP_287	Bloomington-28165	inbred line, mix of	SRX021059	RAL-287	SRR834535	75	75	75	DGRP-F1	17.6	75bp:35285772	DGR	F1	Blo	28165	RAL-287	SRX021059	22.8	22,778,884	9.8	102
50	DGRP_101	Bloomington-28138	inbred line, mix of	SRX020747	RAL-101	SRR834536	75	75	75	DGRP-F1	17.4	75bp:34790704	DGR	F1	Blo	28138	RAL-101	SRX020747	28.6	28,625,600	12.3	103
51	DGRP_256	Bloomington-28162	inbred line, mix of	SRX021055	RAL-256	SRR834533	75	75	75	DGRP-F1	17.4	75bp:34717354	DGR	F1	Blo	28162	RAL-256	SRX021055	32.6	32,570,608	14	104
52	DGRP_239	Bloomington-28161	inbred line, mix of	SRX021054	RAL-239	SRR834532	75	75	75	DGRP-F1	17.2	75bp:34456900	DGR	F1	Blo	28161	RAL-239	SRX021054	33.0	32,977,516	14.1	105
53	DGRP_642	Bloomington-28216	inbred line, mix of	SRX021331	RAL-642	SRR835082	75	75	75	DGRP-F1	17.1	75bp:34225882	DGR	F1	Blo	28216	RAL-642	SRX021331	32.7	32,678,408	14	106
54	DGRP_85	Bloomington-28274	inbred line, mix of	SRX021490	RAL-85	SRR835064	75	75	75	DGRP-F1	17.1	75bp:34278668	DGR	F1	Blo	28274	RAL-85	SRX021490	32.8	32,816,544	14.1	107
55	DGRP_309	Bloomington-28166	inbred line, mix of	SRX021060	RAL-309	SRR834504	75	75	75	DGRP-F1	17.0	75bp:34062874	DGR	F1	Blo	28166	RAL-309	SRX021060	31.0	30,967,892	13.3	108
56	DGRP_563	Bloomington-28211	inbred line, mix of	SRX023836	RAL-563	SRR835049	75	75	75	DGRP-F1	17.0	75bp:33949618	DGR	F1	Blo	28211	RAL-563	SRX023836	32.7	32,699,510	14	109
57	DGRP_69	Bloomington-28130	inbred line, mix of	SRX023449	RAL-69	SRR933583	75	75	75	DGRP-F1	16.9	75bp:33850154	DGR	F1	Blo	28130	RAL-69	SRX023449	31.3	31,330,920	13.4	110
58	DGRP_161	Bloomington-28148	inbred line, mix of	SRX020761	RAL-161	SRR834548	75	75	75	DGRP-F1	16.8	75bp:33569140	DGR	F1	Blo	28148	RAL-161	SRX020761	31.9	31,889,608	13.7	111
59	DGRP_721	Bloomington-28220	inbred line, mix of	SRX021381	RAL-721	SRR835099	75	75	75	DGRP-F1	16.8	75bp:33513644	DGR	F1	Blo	28220	RAL-721	SRX021381	29.3	29,343,228	12.6	112
60	DGRP_310	Bloomington-28276	inbred line, mix of	SRX021080	RAL-310	SRR834505	75	75	75	DGRP-F1	16.8	75bp:33591214,	DGR	F1	Blo	28276	RAL-310	SRX021080	29.3	29,276,428	12.5	113
61	DGRP_73	Bloomington-28131	inbred line, mix of	SRX023450	RAL-73	SRR835100	75	75	75	DGRP-F1	16.6	75bp:33134294	DGR	F1	Blo	28131	RAL-73	SRX023450	31.7	31,685,126	13.6	114
62	DGRP_91	Bloomington-28136	inbred line, mix of	SRX021503	RAL-91	SRR835078	75	75	75	DGRP-F1	16.6	75bp:33211530	DGR	F1	Blo	28136	RAL-91	SRX021503	31.5	31,458,484	13.5	115
63	DGRP_367	Bloomington-28181	inbred line, mix of	SRX021103	RAL-367	SRR834544	75	75	75	DGRP-F1	16.4	75bp:32780338	DGR	F1	Blo	28181	RAL-367	SRX021103	30.6	30,576,784	13.1	116
64	DGRP_336	Bloomington-28172	inbred line, mix of	SRX021096	RAL-336	SRR933570	75	75	75	DGRP-F2	16.3	75bp:32559800	DGR	F2	Blo	28172	RAL-336	SRX021096	26.5	26,524,196	11.4	117
65	DGRP_646	Bloomington-28217	inbred line, mix of	SRX021332	RAL-646	SRR835083	75	75	75	DGRP-F1	16.2	75bp:32350764	DGR	F1	Blo	28217	RAL-646	SRX021332	28.7	28,739,440	12.3	118
66	DGRP_88	Bloomington-28135	inbred line, mix of	SRX021495	RAL-88	SRR835070	75	75	75	DGRP-F1	16.0	75bp:32012994	DGR	F1	Blo	28135	RAL-88	SRX021495	30.9	30,899,746	13.2	119
67	DGRP_105	Bloomington-28139	inbred line, mix of	SRX020745	RAL-105	SRR834538	75	75	75	DGRP-F1	15.9	75bp:31797740	DGR	F1	Blo	28139	RAL-105	SRX020745	30.8	30,788,030	13.2	120
68	DGRP_716	Bloomington-28219	inbred line, mix of	SRX021380	RAL-716	SRR835085	75	75	75	DGRP-F1	15.8	75bp:31572156	DGR	F1	Blo	28219	RAL-716	SRX021380	29.2	29,244,674	12.5	121
69	DGRP_217	Bloomington-28154	inbred line, mix of	SRX021041	RAL-217	SRR834525	75	75	75	DGRP-F1	15.7	75bp:31354632	DGR	F1	Blo	28154	RAL-217	SRX021041	29.9	29,857,218	12.8	122
70	DGRP_158	Bloomington-28147	inbred line, mix of	SRX021009	RAL-158	SRR834549	75	75	75	DGRP-F1	15.5	75bp:31050122	DGR	F1	Blo	28147	RAL-158	SRX021009	27.6	27,629,766	11.8	123
71	DGRP_227	Bloomington-28156	inbred line, mix of	SRX021042	RAL-227	SRR834528	75	75	75	DGRP-F1	15.5	75bp:30993282	DGR	F1	Blo	28156	RAL-227	SRX021042	27.8	27,790,122	11.9	124
72	DGRP_233	Bloomington-28159	inbred line, mix of	SRX021061	RAL-233	SRR834530	75	75	75	DGRP-F1	14.9	75bp:29795450	DGR	F1	Blo	28159	RAL-233	SRX021061	28.4	28,358,088	12.2	125

73	DGRP_338	Bloomington-28173	inbred line, mix of	SRX021097	RAL-338	SRR834513	75	75	75	DGRP-F1	14.5	75bp:28907664	DGRP	F1	Blo	28173	RAL-338	SRX021097	26.0	25,964,076	11.1	126
74	DGRP_228	Bloomington-28157	inbred line, mix of	SRX021043	RAL-228	SRR834529	75	75	75	DGRP-F1	14.4	75bp:28874696	DGRP	F1	Blo	28157	RAL-228	SRX021043	27.3	27,339,700	11.7	127
75	DGRP_350	Bloomington-28176	inbred line, mix of	SRX021100	RAL-350	SRR834515	75	75	75	DGRP-F1	14.0	75bp:27970002	DGRP	F1	Blo	28176	RAL-350	SRX021100	25.4	25,355,594	10.9	128
76	DGRP_317	Bloomington-28167	inbred line, mix of	SRX021081	RAL-317	SRR834506	75	75	75	DGRP-F1	13.6	75bp:27246088	DGRP	F1	Blo	28167	RAL-317	SRX021081	23.3	23,252,944	10	129
77	DGRP_129	Bloomington-28141	inbred line, mix of	SRX020748	RAL-129	SRR834540	75	75	75	DGRP-F1	13.4	75bp:26898800	DGRP	F1	Blo	28141	RAL-129	SRX020748	25.8	25,797,252	11.1	130
78	DGRP_26	Bloomington-28123	inbred line, mix of	SRX021056	RAL-26	SRR933566	75	merged, use min length	75.0	DGRP-F1	18.4	44bp:4452211, 75bp:32432947	DGRP	F1	Blo	28123	RAL-26	SRX021056	35.6	35,642,606	14.5	131
79	DGRP_208	Bloomington-25174	inbred line, mix of	SRX005977	RAL-208	SRR933564	75	merged, use min length	75.0	DGRP-F1 + F2	36.3	45bp:11112096, 64bp:8693462, 75l	DGRP	F1 + F2	Blo	25174	RAL-208	SRX005977	69.8	69,801,276	27.5	132
80	DGRP_237	Bloomington-28160	inbred line, mix of	SRX023423	RAL-237	SRR933565	75	merged, use min length	75.0	DGRP-F1	29.1	44bp:15879498, 75bp:34500267	DGRP	F1	Blo	28160	RAL-237	SRX023423 and	45.0	44,995,628	16.7	133
81	DGRP_28	Bloomington-28124	inbred line, mix of	SRX021783	RAL-28	SRR933567	75	merged, use min length	75.0	DGRP-F1	26.4	44bp:15469056, 75bp:26422488	DGRP	F1	Blo	28124	RAL-28	SRX021783 and	39.1	39,097,863	14.1	134
82	DGRP_41	Bloomington-28126	inbred line, mix of	SRX021791	RAL-41	SRR933576	75	merged, use min length	75.0	DGRP-F1	24.4	44bp:14055721, 75bp:24438432	DGRP	F1	Blo	28126	RAL-41	SRX021791 and	36.1	36,082,067	13.1	135
83	DGRP_639	Bloomington-25199	inbred line, mix of	SRX006161	RAL-639	SRR933582	75	75.0	75.0	DGRP-F1	21.1	75bp:42120400	DGRP	F1	Blo	25199	RAL-639	SRX006161 and	39.0	39,017,668	16.7	136
84	DGRP_799	Bloomington-25207	inbred line, mix of	SRX006173	RAL-799	SRR933597	75	75.0	75.0	DGRP-F1	20.7	75bp:41386962	DGRP	F1	Blo	25207	RAL-799	SRX006172 and	38.9	38,942,298	16.7	137
85	DGRP_386	Bloomington-28192	inbred line, mix of	SRX021798	RAL-386	SRR933574	75	merged, use min length	75.0	DGRP-F1	20.5	44bp:11122819, 75bp:20778304	DGRP	F1	Blo	28192	RAL-386	SRX021798 and	29.6	29,626,175	10.8	138
86	DGRP_820	Bloomington-25208	inbred line, mix of	SRX006174	RAL-820	SRR933598	75	75.0	75.0	DGRP-F1	18.8	75bp:37666476	DGRP	F1	Blo	25208	RAL-820	SRX006174 and	33.9	33,910,170	14.5	139
PHASE 3: Remaining DGRP Illumina-sequenced libaries compatible for TIDAL analysis to be loaded into TIDAL-Fly database.																								
87	DGRP_359	Bloomington-28179	inbred line, mix of	SRX023424	RAL-359	SRR834546	95	95.0	95.0	DGRP-F1	18.6	95bp:37271884	DGRP	F1	Blo	28179	RAL-359	SRX023424	37.3	37,271,884	20.2			
88	DGRP_589	Bloomington-28213	inbred line, mix of	SRX023837	RAL-589	SRR060821	75	75.0	75.0	DGRP-F1	16.5	75bp:32998374	DGRP	F1	Blo	28213	RAL-589	SRX023837	190.1	190,072,852	104.4			
89	DGRP_40	Bloomington-29651	inbred line, mix of	SRX021235	RAL-40	SRR835025	95	95	95	DGRP-F1	34.5	95bp:69063428	DGRP	F1	Blo	29651	RAL-40	SRX021235	61.3	61,345,170	33.3			
90	DGRP_321	Bloomington-29655	inbred line, mix of	SRX021094	RAL-321	SRR834511	95	95	95	DGRP-F1	33.7	95bp:67314152	DGRP	F1	Blo	29655	RAL-321	SRX021094	61.7	61,687,556	33.5			
91	DGRP_443	Bloomington-28199	inbred line, mix of	SRX021260	RAL-443	SRR835031	95	95	95	DGRP-F1	28.8	95bp:57567568	DGRP	F1	Blo	28199	RAL-443	SRX021260	52.5	52,492,412	28.5			
92	DGRP_38	Bloomington-28125	inbred line, mix of	SRX025317	RAL-38	SRR834541	95	95	95	DGRP-F1	28.1	95bp:56154204	DGRP	F1	Blo	28125	RAL-38	SRX025317	51.6	51,612,838	28			
93	DGRP_320	Bloomington-29654	inbred line, mix of	SRX021063	RAL-320	SRR834510	95	95	95	DGRP-F1	25.9	95bp:51875680	DGRP	F1	Blo	29654	RAL-320	SRX021063	44.5	44,510,214	24.2			
94	DGRP_406	Bloomington-29657	inbred line, mix of	SRX021254	RAL-406	SRR835024	95	95	95	DGRP-F1	25.9	95bp:51821248	DGRP	F1	Blo	29657	RAL-406	SRX021254	46.0	46,001,604	25			
95	DGRP_392	Bloomington-28194	inbred line, mix of	SRX021157	RAL-392	SRR834520	95	95	95	DGRP-F1	25.6	95bp:51156860	DGRP	F1	Blo	28194	RAL-392	SRX021157	42.8	42,803,392	23.2			
96	DGRP_405	Bloomington-29656	inbred line, mix of	SRX021242	RAL-405	SRR835023	95	95	95	DGRP-F1	25.0	95bp:50080536	DGRP	F1	Blo	29656	RAL-405	SRX021242	42.2	42,196,622	22.9			
97	DGRP_461	Bloomington-28200	inbred line, mix of	SRX021262	RAL-461	SRR835033	95	95	95	DGRP-F1	24.7	95bp:49324528	DGRP	F1	Blo	28200	RAL-461	SRX021262	40.4	40,407,678	21.9			
98	DGRP_177	Bloomington-28150	inbred line, mix of	SRX021026	RAL-177	SRR834547	95	95	95	DGRP-F1	24.6	95bp:49114764	DGRP	F1	Blo	28150	RAL-177	SRX021026	45.3	45,259,994	24.6			
99	DGRP_837	Bloomington-28246	inbred line, mix of	SRX021479	RAL-837	SRR933599	95	95	95	DGRP-F1	23.2	95bp:46411538	DGRP	F1	Blo	28246	RAL-837	SRX021479	38.0	38,040,748	20.7			
100	DGRP_892	Bloomington-28258	inbred line, mix of	SRX023838	RAL-892	SRR835072	95	95	95	DGRP-F1	22.9	95bp:45702226	DGRP	F1	Blo	28258	RAL-892	SRX023838	37.8	37,797,872	20.5			
101	DGRP_439	Bloomington-29658	inbred line, mix of	SRX021244	RAL-439	SRR835028	95	95	95	DGRP-F1	22.4	95bp:44762436	DGRP	F1	Blo	29658	RAL-439	SRX021244	37.5	37,522,394	20.4			
102	DGRP_492	Bloomington-28203	inbred line, mix of	SRX021270	RAL-492	SRR835036	95	95	95	DGRP-F1	22.3	95bp:44580310	DGRP	F1	Blo	28203	RAL-492	SRX021270	40.7	40,704,078	22.1			
103	DGRP_502	Bloomington-28204	inbred line, mix of	SRX021271	RAL-502	SRR835038	95	95	95	DGRP-F1	22.2	95bp:44336646	DGRP	F1	Blo	28204	RAL-502	SRX021271	40.0	39,968,920	21.7			
104	DGRP_370	Bloomington-28182	inbred line, mix of	SRX021104	RAL-370	SRR834539	95	95	95	DGRP-F1	21.9	95bp:43793604	DGRP	F1	Blo	28182	RAL-370	SRX021104	38.4	38,408,446	20.9			
105	DGRP_377	Bloomington-28186	inbred line, mix of	SRX023834	RAL-377	SRR834543	95	95	95	DGRP-F1	21.9	95bp:43796182	DGRP	F1	Blo	28186	RAL-377	SRX023834	40.2	40,225,994	21.8			
106	DGRP_426	Bloomington-28196	inbred line, mix of	SRX021245	RAL-426	SRR835026	95	95	95	DGRP-F1	21.9	95bp:43746634	DGRP	F1	Blo	28196	RAL-426	SRX021245	38.9	38,928,098	21.1			
107	DGRP_887	Bloomington-28279	inbred line, mix of	SRX021527	RAL-887	SRR835069	95	95	95	DGRP-F1	21.8	95bp:43595728	DGRP	F1	Blo	28279	RAL-887	SRX021527	35.9	35,939,248	19.5			
108	DGRP_440	Bloomington-28197	inbred line, mix of	SRX021246	RAL-440	SRR835029	95	95	95	DGRP-F1	21.6	95bp:43161850	DGRP	F1	Blo	28197	RAL-440	SRX021246	31.7	31,659,420	17.2			
109	DGRP_513	Bloomington-29659	inbred line, mix of	SRX021282	RAL-513	SRR835042	95	95	95	DGRP-F1	21.3	95bp:42640722	DGRP	F1	Blo	29659	RAL-513	SRX021282	36.0	36,025,830	19.6			
110	DGRP_508	Bloomington-28205	inbred line, mix of	SRX021272	RAL-508	SRR835040	95	95	95	DGRP-F1	21.2	95bp:42338556	DGRP	F1	Blo	28205	RAL-508	SRX021272	39.1	39,078,690	21.2			
111	DGRP_441	Bloomington-28198	inbred line, mix of	SRX023835	RAL-441	SRR835030	95	95	95	DGRP-F1	21.1	95bp:42278010	DGRP	F1	Blo	28198	RAL-441	SRX023835	34.4	34,448,902	18.7			
112	DGRP_531	Bloomington-28207	inbred line, mix of	SRX021290	RAL-531	SRR835045	95	95	95	DGRP-F1	20.8	95bp:41560152	DGRP	F1	Blo	28207	RAL-531	SRX021290	32.9	32,916,204	17.9			
113	DGRP_383	Bloomington-28190	inbred line, mix of	SRX021113	RAL-383	SRR834554	95	95	95	DGRP-F1	19.9	95bp:39897030	DGRP	F1	Blo	28190	RAL-383	SRX021113	35.1	35,149,446	19.1			
114	DGRP_235	Bloomington-28275	inbred line, mix of	SRX021053	RAL-235	SRR834531	95	95	95	DGRP-F1	19.1	95bp:38296004	DGRP	F1	Blo	28275	RAL-235	SRX021053	33.8	33,829,434	18.4			
115	DGRP_42	Bloomington-28127	inbred line, mix of	SRX021255	RAL-42	SRR835027	95	95	95	DGRP-F1	18.6	95bp:37186556	DGRP	F1	Blo	28127	RAL-42	SRX021255	37.2	37,186,556	20.2			
116	DGRP_21	Bloomington-28122	inbred line, mix of	SRX021040	RAL-21	SRR834526	95	95	95	DGRP-F1	18.5	95bp:37046984	DGRP	F1	Blo	28122	RAL-21	SRX021040	29.2	29,159,002	15.8			
117	DGRP_907	Bloomington-28262	inbred line, mix of	SRX021500	RAL-907	SRR835074	95	95	95	DGRP-F1	18.2	95bp:36385056	DGRP	F1	Blo	28262	RAL-907	SRX021500	32.3	32,273,788	17.5			
118	DGRP_790	Bloomington-28232	inbred line, mix of	SRX021389	RAL-790	SRR835092	95	95	95	DGRP-F1	17.8	95bp:35620658	DGRP	F1	Blo	28232	RAL-790	SRX021389	31.3	31,287,054	17			
119	DGRP_894	Bloomington-28259	inbred line, mix of	SRX021528	RAL-894	SRR835073	95	95	95	DGRP-F1	17.6	95bp:35128536	DGRP	F1	Blo	28259	RAL-894	SRX021528	30.9	30,860,088	16.8			
120	DGRP_832	Bloomington-28245	inbred line, mix of	SRX021477	RAL-832	SRR835057	95	95	95	DGRP-F1	16.9	95bp:33770198	DGRP	F1	Blo	28245	RAL-832	SRX021477	26.0	26,008,378	14.1			
121	DGRP_908	Bloomington-28263	inbred line, mix of	SRX021501	RAL-908	SRR835075	95	95	95	DGRP-F1	16.8	95bp:39111536	DGRP	F1	Blo	28263	RAL-908	SRX021501	36.7	36,702,788	19.9			
122	DGRP_911	Bloomington-28264	inbred line, mix of	SRX021502	RAL-911	SRR835076	95	95	95	DGRP-F1	16.8	95bp:33571736	DGRP	F1	Blo	28264	RAL-911	SRX021502	26.3	26,331,524	14.3			
123	DGRP_703	Bloomington-28218	inbred line, mix of	SRX021508	RAL-703	SRR835084	95	95	95	DGRP-F1	16.5	95bp:32953564	DGRP	F1	Blo	28218	RAL-703	SRX021508	26.3	26,272,930	14.3			
124	DGRP_385	Bloomington-28191	inbred line, mix of	SRX159098	RAL-385	SRR834518	95	95	95	DGRP-F2	14.2	95bp:28408672	DGRP	F2	Blo	28191	RAL-385	SRX159098	21.0	20,956,926	11.4			
125	DGRP_374	Bloomington-28185	inbred line, mix of	SRX023427	RAL-374	SRR835103	95	95	95	DGRP-F1	13.9	95bp:27645755	DGRP	F1	Blo	28185	RAL-374	SRX023427	24.7	24,704,590	13.4			

Table S1D. Drosophila Genome Nexus Strains in the TIDAL-Fly v1 database Genomes previously examined in Lack et al (Pool lab), Genetics 2015

PHASE 1: First Set of Libraries Analyzed for the paper. The remaining libraries in Phase 2 below are also listed in the TIDAL-Fly database

	Stock ID*		Stock Location	Genome Type	SRA Accession	Library ID*	SRA Accession	Min Read Length	Reads (M)	Read Length 1	Read Length 2	Data Group	Focal Genome Repr	Coverage**	Mean Depth**
1	CK1		Langley lab	haploid_embryo	SRX058145	CK1_1-HE	SRR189038	75	35.0	76	76	DPGP2	X,2L,2R,3L,3R	110,870,215	39.4
2	CK2		Langley lab	haploid_embryo	SRX058146	CK2_1-HE	SRR189040	145	20.6	146	146	DPGP2	X,2L,2R,3L,3R	111,567,582	44.7
3	CO13N		Pool Lab	haploid_embryo	SRX058151	CO13N_1-HE	SRR189045	145	18.1	146	146	DPGP2	X,2L,2R,3L,3R	111,264,202	39.8
4	CO15N		Pool Lab	haploid_embryo	SRX058153	CO15N_1-HE	SRR189047	75	38.2	76	76	DPGP2	X,2L,2R,3L,3R	111,058,052	44.2
5	CO4N		Pool Lab	haploid_embryo	SRX058156	CO4N_1-HE	SRR189050	75	32.8	76	76	DPGP2	X,2L,2R,3L,3R	111,025,671	37.7
6	EA119		Pool Lab	haploid_embryo	SRX791698	EA119_1-HE	SRR1686794	100	16.6	100	100	AGES	X,2L,2R,3L,3R	110,802,511	23.9
7	EA49		Pool Lab	haploid_embryo	SRX791713	EA49_1-HE	SRR1686797	100	13.0	100	100	AGES	X,2L,2R,3L,3R	110,446,851	18.2
8	EA87		Pool Lab	haploid_embryo	SRX791850	EA87_1-HE	SRR1686966	100	27.6	100	100	AGES	X,2L,2R,3L,3R	111,259,464	38.9
9	EB148		Langley lab	haploid_embryo	SRX791858	EB148_1-HE	SRR1686971	100	19.1	100	100	AGES	X,2L,2R,3L,3R	110,799,356	27.4
10	EB18		Langley lab	haploid_embryo	SRX792309	EB18_1-HE	SRR1687447	100	16.5	100	100	AGES	X,2L,2R,3L,3R	110,159,072	22.7
11	EB25		Langley lab	haploid_embryo	SRX792329	EB25_1-HE	SRR1687462	100	15.6	100	100	AGES	X,2L,2R,3L,3R	103,098,821	19.4
12	EF39		Pool lab	haploid_embryo	SRX792510	EF39_1-HE	SRR1687668	100	17.7	100	100	AGES	X,2L,2R,3L,3R	110,937,745	24.9
13	EF65		Pool lab	haploid_embryo	SRX792606	EF65_1-HE	SRR1687770	100	12.2	100	100	AGES	X,2L,2R,3L,3R	107,539,051	17.1
14	EF78		Pool lab	haploid_embryo	SRX792639	EF78_1-HE	SRR1687793	100	16.8	100	100	AGES	X,2L,2R,3L,3R	110,871,321	23.8
15	EG40N		Pool lab	inbred_line	SRX792675	EG40N	SRR1687832	100	10.8	100	100	AGES	X,2R	25,659,426	14.3
16	EG69N		Pool lab	inbred_line	SRX793022	EG69N	SRR1688188	100	10.2	100	100	AGES	X,2L,2R,3R	73,335,770	13.6
17	EG73N		Pool lab	inbred_line	SRX793046	EG73N	SRR1688222	100	10.3	100	100	AGES	X,2L,2R,3L,3R	105,554,844	13.8
18	FR180		Pool Lab	haploid_embryo	SRX058184	FR180_2-HE	SRR189090	145	19.1	146	146	DPGP2	X,2L,2R,3L,3R	112,084,373	42.3
19	FR207		Pool Lab	haploid_embryo	SRX058185	FR207_2-HE	SRR189091	145	18.1	146	146	DPGP2	X,2L,2R,3L,3R	112,168,144	38.9
20	FR229		Pool Lab	haploid_embryo	SRX058187	FR229_2-HE	SRR189093	145	19.1	146	146	DPGP2	X,2L,2R,3L,3R	112,159,960	40.6
21	GA129		Pool lab	haploid_embryo	SRX058193	GA129_1-HE	SRR189101	75	41.4	76	76	DPGP2	X,2L,2R,3L,3R	111,099,834	47.9
22	GA130		Pool lab	haploid_embryo	SRX058194	GA130_1-HE	SRR189102	75	43.4	76	76	DPGP2	X,2L,2R,3L,3R	111,771,726	47.6
23	GA141		Pool lab	haploid_embryo	SRX058196	GA141_1-HE	SRR189105	75	40.7	76	76	DPGP2	X,2L,2R,3L,3R	111,146,499	45.1
24	GU10		Langley lab	haploid_embryo	SRX058205	GU10_1-HE	SRR189114	75	33.2	76	76	DPGP2	X,2L,2R,3L,3R	110,887,105	38.9
25	GU2		Langley lab	haploid_embryo	SRX058207	GU2_1-HE	SRR189117	75	33.0	76	76	DPGP2	X,2L,2R,3L,3R	110,886,266	38.4
26	GU6		Langley lab	haploid_embryo	SRX058209	GU6_1-HE	SRR189120	75	33.0	76	76	DPGP2	X,2L,2R,3L,3R	110,933,340	38.1
27	KN133N		Langley lab	haploid_embryo	SRX058256	KN133N_1-HE	SRR189242	75	30.1	76	76	DPGP2	X,2L,2R,3L,3R	111,123,747	35.0
28	KN20N		Langley lab	haploid_embryo	SRX058253	KN20N_1-HE	SRR189239	75	32.0	76	76	DPGP2	X,2L,2R,3L,3R	110,974,588	37.6
29	KN35		Langley lab	haploid_embryo	SRX058258	KN35_1-HE	SRR189245	75	30.2	76	76	DPGP2	X,2L,2R,3L,3R	111,066,184	34.3
30	KR39		Langley lab	haploid_embryo	SRX058267	KR39_1-HE	SRR189254	75	30.2	76	76	DPGP2	X,2L,2R,3L,3R	111,101,981	34.7
31	KR42		Langley lab	haploid_embryo	SRX058268	KR42_1-HE	SRR189255	75	28.5	76	76	DPGP2	X,2L,2R,3L,3R	110,823,311	32.9
32	KR4N		Langley lab	haploid_embryo	SRX058269	KR4N_1-HE	SRR189256	75	29.5	76	76	DPGP2	X,2L,2R,3L,3R	111,045,036	29.6
33	KT1		Langley lab	haploid_embryo	SRX058273	KT1_1-HE	SRR189260	75	25.5	76	76	DPGP2	X,2L,2R,3L,3R	111,229,668	37.6
34	KT6		Langley lab	haploid_embryo	SRX058274	KT6_1-HE	SRR189262	75	27.6	76	76	DPGP2	X,2L,2R,3L,3R	111,074,001	29.6
35	NG10N		Langley lab	haploid_embryo	SRX058275	NG10N_1-HE	SRR189263	75	33.5	76	76	DPGP2	X,2L,2R,3L,3R	111,001,728	38.9
36	NG3N		Pool lab	haploid_embryo	SRX058378	NG3N_1-HE	SRR189414	75	32.3	76	76	DPGP2	X,2L,2R,3L,3R	111,022,313	36.4
37	NG6N		Pool lab	haploid_embryo	SRX058278	NG6N_1-HE	SRR189266	75	33.9	76	76	DPGP2	X,2L,2R,3L,3R	111,483,632	38.4
38	RC1		Langley lab	haploid_embryo	SRX058281	RC1_1-HE	SRR189269	75	26.2	76	76	DPGP2	X,2L,2R,3L,3R	110,659,330	30.3
39	RC5		Langley lab	haploid_embryo	SRX058282	RC5_1-HE	SRR189270	75	29.6	76	76	DPGP2	X,2L,2R,3L,3R	110,723,230	24.9
40	RG33		Pool lab	haploid_embryo	SRX058353	RG33_1-HE	SRR189389	145	37.0	146	146	DPGP2	X,2L,2R,3L,3R	111,689,097	69.8
41	RG4N		Pool lab	haploid_embryo	SRX058362	RG4N_1-HE	SRR306629	145	24.9	146	146	DPGP2	X,2L,2R,3L,3R	111,396,200	54.2
42	RG6N		Langley lab	haploid_embryo	SRX058368	RG6N_1-HE	SRR306624	145	22.0	146	146	DPGP2	X,2L,2R,3L,3R	111,216,270	47.7
43	SB10		Pool lab	haploid_embryo	SRX799453	SB10_1-HE	SRR1696817	100	23.7	100	100	AGES	X,2L,2R,3L,3R	111,238,977	34.2

44	SB16		Pool lab	haploid_embryo	SRX799458	SB16_1-HE	SRR1696818	100	24.1	100	100	AGES	X,2L,2R,3L,3R	111,438,387	32.4
45	SB31		Pool lab	haploid_embryo	SRX799471	SB31_1-HE	SRR1696822	100	21.4	100	100	AGES	X,2L,2R,3L,3R	110,110,096	30.2
46	SF332		Langley lab	haploid_embryo	SRX799661	SF332_1-HE	SRR1696987	100	13.7	100	100	AGES	X,2L,2R,3L,3R	110,582,475	19.9
47	SF428		Langley lab	haploid_embryo	SRX799663	SF428_1-HE	SRR1696988	100	12.6	100	100	AGES	X,2L,2R,3L,3R	110,495,354	16.5
48	SF447		Pool lab	haploid_embryo	SRX799664	SF447_1-HE	SRR1696989	100	15.3	100	100	AGES	X,2L,2R,3L,3R	110,719,070	20.7
49	SP188		Langley lab	haploid_embryo	SRX058287	SP188_1-HE	SRR306632	145	18.4	146	146	DPGP2	X,2L,2R,3L,3R	111,309,908	38.6
50	SP221		Pool lab	haploid_embryo	SRX058288	SP221_1-HE	SRR189277	145	18.9	146	146	DPGP2	X,2L,2R,3L,3R	111,199,509	40.7
51	SP80		Langley lab	haploid_embryo	SRX058292	SP80_1-HE	RR189281	145	18.3	146	146	DPGP2	X,2L,2R,3L,3R	111,128,089	41.6
52	TZ10		Langley lab	haploid_embryo	SRX058283	TZ10_1-HE	SRR189272	75	23.4	76	76	DPGP2	X,2L,2R,3L,3R	110,737,417	26.9
53	TZ14		Langley lab	haploid_embryo	SRX058284	TZ14_1-HE	SRR189273	75	31.7	76	76	DPGP2	X,2L,2R,3L,3R	111,045,117	35.3
54	TZ8		Langley lab	haploid_embryo	SRX058285	TZ8_1-HE	SRR189274	75	31.2	76	76	DPGP2	X,2L,2R,3L,3R	110,920,135	35.9
55	UG19		Pool lab	haploid_embryo	SRX058376	UG19_1-HE	SRR189412	75	33.9	76	76	DPGP2	X,2L,2R,3L,3R	110,998,446	37.2
56	UG28N		Pool lab	haploid_embryo	SRX058277	UG28N_1-HE	SRR189265	75	33.1	76	76	DPGP2	X,2L,2R,3L,3R	110,922,524	38.8
57	UG5N		Pool lab	haploid_embryo	SRX058379	UG5N_1-HE	SRR189415	75	32.1	76	76	DPGP2	X,2L,2R,3L,3R	111,110,172	34.2
58	UK120		Langley lab	haploid_embryo	SRX799667	UK120_1-HE	SRR1696992	100	15.6	100	100	AGES	X,2L,2R,3L,3R	108,541,523	23.2
59	UK2		Langley lab	haploid_embryo	SRX799668	UK2_1-HE	SRR1696993	100	18.3	100	100	AGES	X,2L,2R,3L,3R	111,035,774	25.2
60	UK57		Langley lab	haploid_embryo	SRX799669	UK57_1-HE	SRR1696994	100	20.7	100	100	AGES	X,2L,2R,3L,3R	110,904,117	29.9
61	ZI191		Langley/Pool Lab	haploid_embryo	SRR202128	ZI191-HE	SRR202128	145	23.3	146	146	DPGP3	X,2L,2R,3L,3R	110,845,648	49.2
62	ZI216N		Langley/Pool Lab	haploid_embryo	SRR203328	ZI216N-HE	SRR203328	145	20.4	146	146	DPGP3	X,2L,2R,3L,3R	111,176,278	44.2
63	ZI220		Langley/Pool Lab	haploid_embryo	SRR203067	ZI220-HE	SRR203067	145	23.2	146	146	DPGP3	X,2L,2R,3L,3R	111,180,528	51.3
64	ZI227		Langley/Pool Lab	haploid_embryo	SRR202126	ZI227-HE	SRR202126	145	25.7	146	146	DPGP3	X,2L,2R,3L,3R	111,269,290	54.7
65	ZI254N		Langley/Pool Lab	haploid_embryo	SRR203335	ZI254N-HE	SRR203335	145	18.1	146	146	DPGP3	X,2L,2R,3L,3R	111,099,774	41.2
66	ZI273N		Langley/Pool Lab	haploid_embryo	SRR210786	ZI273N-HE	SRR210786	145	21.1	146	146	DPGP3	X,2L,2R,3L,3R	111,341,925	43.2
67	ZS11		Langley lab	haploid_embryo	SRX058372	ZS11_1-HE	SRR189408	75	32.4	76	76	DPGP2	X,2L,2R,3L,3R	110,736,467	37.7
68	ZS37		Langley lab	haploid_embryo	SRX058373	ZS37_1-HE	SRR189409	75	36.0	76	76	DPGP2	X,2L,2R,3L,3R	110,877,107	41.2
69	ZS5		Langley lab	haploid_embryo	SRX058374	ZS5_1-HE	SRR900425	75	33.4	76	76	DPGP2	X,2L,2R,3L,3R	110,504,765	38.2
70	ZS56		Langley lab	haploid_embryo	SRX058375	ZS56_1-HE	SRR189411	75	30.7	76	76	DPGP2	X,2L,2R,3L,3R	110,737,628	36.0

Table S1E. Key for Two-letter code for the demography of the Drosophila Genome Nexus Strains in the TIDAL-Fly v1 database

									Number of genomes for each euchromatic chromosome				
	Populatio	Country	Locality	Date	Collector(s)	Latitude (I	Longitude	Elevation	X	2	3	Comments	
1	**CK**	Congo	Kisangani	8/2010	J. Kennis	0.51	25.19	400	2	2	2		
2	**CO**	Cameroon	Oku	4/2004	J. Pool	6.25	10.43	2169	10	10	10	3 additional chromosome 3 extractions exist, but derive from str	
3	**EA**	Ethiopia	Gambella	12/2011	R. Corbett-Detig	8.25	34.59	525	4	4	4		
4	**EB**	Ethiopia	Bonga	12/2011	R. Corbett-Detig	7.26	36.25	1725	5	5	5		
5	**ED**	Ethiopia	Dodola	12/2008	J. Pool	6.98	39.18	2492	6	7	6		
6	**EF**	Ethiopia	Fiche	12/2011	R. Corbett-Detig	9.81	38.63	3070	5	5	5		
7	**EG**	Egypt	Cairo	1/2011	J. Atallah	30.10	31.32	25	3	2	2	contains heterozygous intervals (masked in consensus sequenc	
8	**EM**	Ethiopia	Masha	12/2011	R. Corbett-Detig	7.74	35.48	2260	3	3	3		
9	**ER**	Ethiopia	Debre Birha	12/2011	R. Corbett-Detig	9.68	39.53	2840	5	5	5		
10	**EZ**	Ethiopia	Ziway	12/2008	J. Pool	7.93	38.72	1642	5	4	5		
11	**FR**	France	Lyon	7/2010	J. Pool	45.77	4.86	175	9	9	9		
12	**GA**	Gabon	Franceville	3/2002	B. Ballard & S. Charlat	-1.65	13.60	332	10	9	10	GA191 is missing chromosome 2 data due to heterozygosity	
13	**GU**	Guinea	Dondé	6/2005	B. B. Sow	10.70	-12.25	801	5	5	7		
14	**KM**	Kenya	Malindi	7/2002	B. Ballard	-1.43	40.03	78	4	3	1	contains heterozygous intervals (masked in consensus sequenc	
15	**KN**	Kenya	Nyahururu	1/2009	J. Pool	0.04	36.36	2303	6	5	6		
16	**KO**	Kenya	Molo	1/2009	J. Pool	-0.25	35.73	2506	4	0	4		
17	**KR**	Kenya	Marigat	1/2009	J. Pool	0.47	35.98	1062	6	4	6		
18	**KT**	Kenya	Thika	1/2009	J. Pool	-1.04	37.08	1531	2	2	2		
19	**NG**	Nigeria	Maiduguri	9/2004	D. Gwary & B. Sastawa	11.85	13.16	295	6	6	6		
20	**RAL**	United State	Raleigh NC	2003	T. Mackay	35.76	-78.66	91	205	205	205	(sample sizes before heterozygosity and IBD masking)	
21	**RC**	Rwanda	Cyangugu	12/2008	J. Pool	-2.29	28.55	1602	2	2	2		
22	**RG**	Rwanda	Gikongoro	12/2008	J. Pool	-2.49	28.92	1927	27	27	27		
23	**SB**	South Africa	Barkly East	12/2011	J. Pool	-30.97	27.59	1800	5	5	5		
24	**SD**	South Africa	Dullstroom	12/2011	J. Pool	-25.42	30.10	2000	5	4	5	SD82 is missing chromosome 2 data due to heterozygosity	
25	**SE**	South Africa	Port Edward	12/2011	J. Pool	-31.06	30.22	50	3	3	3		
26	**SF**	South Africa	Fouriesburg	12/2011	J. Pool	-28.60	28.05	1800	5	4	5	SF7 is missing chromosome 2 data due to heterozygosity	
27	**SP**	South Africa	Phalaborwa	7/2010	R. Corbett-Detig	-23.94	31.14	350	7	7	7		
28	**TZ**	Tanzania	Uyole	12/2009	L. Nsemwa	-8.89	33.44	1800	3	3	3		
29	**UG**	Uganda	Namulonge	4/2005	J. Ogwang	0.53	32.60	1134	4	4	6		
30	**UK**	Uganda	Kisoro	1/2012	R. Corbett-Detig	-1.28	29.69	1925	5	4	5	UK120 is missing chromosome 2 data due to heterozygosity	
31	**UM**	Uganda	Masindi	7/2010	J. Pool	1.68	31.72	1170	3	3	3		
32	**ZI**	Zambia	Siavonga	7/2010	R. Corbett-Detig	-16.54	28.72	530	196	197	197	ZI382 is missing chromosome X due to heterozygosity	
33	**ZK**	Zimbabwe	Lake Kariba	5/1994	T. Mutangadura	-16.52	28.80	619	3	0	2	contains heterozygous intervals (masked in consensus sequenc	
34	**ZL**	Zambia	Livingstone	7/2010	R. Corbett-Detig	-17.86	25.86	900	1	1	1		
35	**ZO**	Zambia	Solwezi	7/2010	R. Corbett-Detig	-12.18	26.40	1380	2	2	2		
36	**ZS**	Zimbabwe	Sengwa	9/1990	R. Ramey	-18.16	28.22	865	4	4	4		

Table S1F.Pools of flies analyzed in the TIDAL-Fly v1 database

	Fly Pool ID	Fly Pool Location	Genome Type	SRA Accession	Library ID*	SRA Accession	Min Read Length	Reads (M)	Read Length 1	Read Length 2	Reference
1	AU-CI	Australia, Cairns	Pool	SRR1177951	pool-AU-CI	SRR1177951	75	26M	75	0	Reinhardt et al.,2014
2	AU-CR	Australia, Cardwell	Pool	SRR1177952	pool-AU-CR	SRR1177952	75	36M	75	0	Reinhardt et al.,2014
3	AU-MO	Australia, Miller's Orchard	Pool	SRR1177953	pool-AU-MO	SRR1177953	75	24M	75	0	Reinhardt et al.,2014
4	AU-SO	Australia, Sorell	Pool	SRR1177955	pool-AU-SO	SRR1177955	75	38M	75	0	Reinhardt et al.,2014
5	EU-PO1	Europe, Portugal	Pool	SRR188217	pool-EU-PO1	SRR188217	75	19.5M	75	75	Kofler et al., 2012
6	EU-PO2	Europe, Portugal	Pool	SRR189066	pool-EU-PO2	SRR189066	73	60.9M	74	74	Kofler et al., 2012
7	NA-SC	North America, Eutawville, SC	50-100 pooled individuals	SRX661835	pool-NA-SC	SRR1525696	100	82M	100	100	Bergland et al.,2014
8	NA-GA	North America, Hahira, GA	50-100 pooled individuals	SRX661834	pool-NA-GA	SRR1525695	100	97M	100	100	Bergland et al.,2014
9	NA-NC	North America, 90 of the DGRP fly stra	Pool	SRX661836	pool-NA-NC	SRR1525697	90	42M	90	90	Bergland et al.,2014
10	NA-PA1	North America, Linvilla, PA, spring 200	50-100 pooled individuals	SRX661837	pool-NA-PA1	SRR1525768	100	181M,	100	100	Bergland et al.,2014
11	NA-PA2	North America,Linvilla, PA, fall 2009	50-100 pooled individuals	SRX661838	pool-NA-PA2	SRR1525769	100	73.7M	100	100	Bergland et al.,2014
12	NA-PA3	North America,Linvilla, PA, spring 201(	50-100 pooled individuals	SRX661839	pool-NA-PA3	SRR1525770	100	35M	100	100	Bergland et al.,2014
13	NA-PA4	North America,Linvilla, PA, fall 2010	50-100 pooled individuals	SRX661840	pool-NA-PA4	SRR1525771	100	70M	100	100	Bergland et al.,2014
14	NA-PA5	North America,Linvilla, PA, spring 201	50-100 pooled individuals	SRX661841	pool-NA-PA5	SRR1525772	90	78.1M	90	90	Bergland et al.,2014
15	NA-PA6	North America,Linvilla, PA, fall 2011 (p	50-100 pooled individuals	SRX661842	pool-NA-PA6	SRR1525773	100	73.2M	100	100	Bergland et al.,2014
16	NA-PA7	North America,Linvilla, PA, fall 2011 (p	50-100 pooled individuals	SRX661843	pool-NA-PA7	SRR1525774	90	82.1M	90	90	Bergland et al.,2014
17	NA-ME1	North America,Bowdoinham, ME	50-100 pooled individuals	SRX661844	pool-NA-ME1	SRR1525698,	100	95M,	100	100	Bergland et al.,2014
18	NA-ME2	North America,Bowdoinham, ME	50-100 pooled individuals	SRX661845	pool-NA-ME2	SRR1525699, SRR2006283	100	32M, 32M	100	100	Bergland et al.,2014
19	NA-FL1	North America,Homestead, FL	50-100 pooled individuals	SRX661832	pool-NA-FL1	SRR1525685	100	64M,	100	100	Bergland et al.,2014
20	NA-FL2	North America,Homestead, FL	50-100 pooled individuals	SRX661833	pool-NA-FL2	SRR1525694	100	41.6M	100	100	Bergland et al.,2014
21	EU-AU-LF1	Austria, lightest females	Pool, 100 females	ERX149369	pool-EU-AU-LF1	ERR173225	100	86.1	100	100	Bastide et al., 2013
22	EU-AU-LF2	Austria, lightest females	Pool, 100 females	ERX149370	pool-EU-AU-LF2	ERR173226	100	95.0	100	100	Bastide et al., 2013
23	EU-AU-DF1	Austria, darkest females	Pool, 100 females	ERX149371	pool-EU-AU-DF1	ERR173227	100	34.5	100	100	Bastide et al., 2013
24	EU-AU-DF2	Austria, darkest females	Pool, 100 females	ERX149372	pool-EU-AU-DF2	ERR173228	150	39.7	150	85	Bastide et al., 2013
25	EU-AU-DF3	Austria, darkest females	Pool, 100 females	ERX149373	pool-EU-AU-DF3	ERR173229	100	95.0	100	100	Bastide et al., 2013
26	EU-AU1	Austria/Vienna reference	Pool, 100-150 females	ERX149374	pool-EU-AU1	ERR173230	100	41.0	100	100	Bastide et al., 2013
27	EU-AU2	Austria/Vienna reference	Pool, 100-150 females	ERX149375	pool-EU-AU2	ERR173231	150	35.3	150	80	Bastide et al., 2013
28	EU-AU3	Austria/Vienna reference	Pool, 100-150 females	ERX149376	pool-EU-AU3	ERR173232	100	38.9M	100	100	Bastide et al., 2013
29	EU-AU4	Austria/Vienna reference	Pool, 100-150 females	ERX149377	pool-EU-AU4	ERR173233	150	40.7	150	80	Bastide et al., 2013
30	EU-AU5	Austria/Vienna reference	Pool, 100-150 females	ERX149378	pool-EU-AU5	ERR173234	100	41.7	100	100	Bastide et al., 2013
31	EU-AU6	Austria/Vienna reference	Pool, 100-150 females	ERX149379	pool-EU-AU6	ERR173235	100	41.9	100	100	Bastide et al., 2013
32	EU-AU7	Austria/Vienna reference	Pool, 100-150 females	ERX149380	pool-EU-AU7	ERR173236	100	40.8	100	100	Bastide et al., 2013
33	EU-AU8	Austria/Vienna reference	Pool, 100-150 females	ERX149381	pool-EU-AU8	ERR173237	100	40.2	100	100	Bastide et al., 2013
34	EU-AU9	Austria/Vienna reference	Pool, 100-150 females	ERX149382	pool-EU-AU9	ERR173238	100	40.6M	100	100	Bastide et al., 2013
35	EU-AU10	Austria/Vienna reference	Pool, 100-150 females	ERX149383	pool-EU-AU10	ERR173239	150	40.9M	150	80	Bastide et al., 2013
36	EU-AU11	Austria/Vienna reference	Pool, 100-150 females	ERX149384	pool-EU-AU11	ERR173240	100	40	100	100	Bastide et al., 2013
37	EU-AU12	Austria/Vienna reference	Pool, 100-150 females	ERX149385	pool-EU-AU12	ERR173241	150	40.1	150	80	Bastide et al., 2013
38	EU-IT-LF1	Italy, lightest females	Pool, 100 females	ERX149386	pool-EU-IT-LF1	ERR173242	90	14.6	90	90	Bastide et al., 2013
39	EU-IT-LF2	Italy, lightest females	Pool, 100 females	ERX149387	pool-EU-IT-LF2	ERR173243	90	15.6	90	90	Bastide et al., 2013
40	EU-IT-LF3	Italy, lightest females	Pool, 100 females	ERX149388	pool-EU-IT-LF3	ERR173244	90	15.6	90	90	Bastide et al., 2013
41	EU-IT-LF4	Italy, lightest females	Pool, 100 females	ERX149389	pool-EU-IT-LF4	ERR173245	100	138.8	100	100	Bastide et al., 2013
42	EU-IT-LF5	Italy, lightest females	Pool, 100 females	ERX149390	pool-EU-IT-LF5	ERR173246	100	131.7	100	100	Bastide et al., 2013
43	EU-IT-DF1	Italy, darkest females	Pool, 100 females	ERX149391	pool-EU-IT-DF1	ERR173247	92	11.1	92	92	Bastide et al., 2013
44	EU-IT-DF2	Italy, darkest females	Pool, 100 females	ERX149392	pool-EU-IT-DF2	ERR173248	92	10.3	92	92	Bastide et al., 2013
45	EU-IT-DF3	Italy, darkest females	Pool, 100 females	ERX149393	pool-EU-IT-DF3	ERR173249	92	9.5	92	92	Bastide et al., 2013
46	EU-IT-DF4	Italy, darkest females	Pool, 100 females	ERX149394	pool-EU-IT-DF4	ERR173250	100	138.4	100	100	Bastide et al., 2013
47	EU-IT-DF5	Italy, darkest females	Pool, 100 females	ERX149395	pool-EU-IT-DF5	ERR173251	100	138.4	100	100	Bastide et al., 2013
48	EU-IT1	Italy, reference	Pool, 100-150 females	ERX149396	pool-EU-IT1	ERR173252	94	21.6	94	94	Bastide et al., 2013
49	EU-IT2	Italy, reference	Pool, 100-150 females	ERX149397	pool-EU-IT2	ERR173253	94	17.4	94	94	Bastide et al., 2013
50	EU-IT3	Italy, reference	Pool, 100-150 females	ERX149398	pool-EU-IT3	ERR173254	94	15.5	94	94	Bastide et al., 2013
51	EU-IT4	Italy, reference	Pool, 100-150 females	ERX149399	pool-EU-IT4	ERR173255	150	132M	150	150	Bastide et al., 2013
52	EU-IT5	Italy, reference	Pool, 100-150 females	ERX149400	pool-EU-IT5	ERR173256	150	153M	150	150	Bastide et al., 2013

Table S2 Primer Lists

Oligos for gDNA libraries construction	PE_Tdot_common_C* linker	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
	Barcoded linker PE_tdot_CAGCACTA	/5Phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGcagcactaATCTCGTATGCCGTCTTCTGCTTG
	Barcode linker PE_tdot_AGAGATGC	/5Phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGagagatgcATCTCGTATGCCGTCTTCTGCTTG
	Barcode linker PE_tdot_TCACGTGT	/5Phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGtcacgtgtATCTCGTATGCCGTCTTCTGCTTG
	PE-POSTPCR_1	AATGATACGGCGACCACCGAGA
	PE-POSTPCR_2	CAAGCAGAAGACGGCATACGAG

Table S2 Primer Lists

PRIMER SEQUENCE FOR S2c1 INSERTS

Sr. no.	SV#	Chr	Chr_coord_	Chr_coord_	TE	Forward Prmer Sequence	Forward primer Length	Reverse Primer Sequence	Reverse Primer length	Length w/c	Length with TE
1	14	chr2L	312233	312444	Stalker2	ACTGGCAAGACCTCGATTGTC	21	CGGATTGTCCGACGTATTAGA	21	411	8661
2	5	chr2L	87055	87258	1731	GGTGCCGCACAATTCGCAGGC	21	GCTCCGGTCAGAAGATTTTAC	22	404	5053
3	621	chr2R	544728	544940	HMS-Beagle	ATGGGGAACACAGTCAAAATG	21	CATATTATTCAAAGATTGGAT	21	353	6882
4	630	chr2R	2675584	2675743	opus	GAAGAATGAGACGGGCTACCG	21	CTGTAATTGTGTACACAGTCA	21	332	8891
5	1123	chr3L	516184	516383	17.6	AGTCCTCCCAACCGAGATGG	20	GATTGAATTTCAGTTGGCCGG	21	349	7788
6	1131	chr3L	552782	552981	Juan	TTTATAATAATGTTGCTGGAC	21	GGCATTTCTATAGCCCAACCC	21	344	4580
7	1743	chr3R	6430955	6431163	3S18	ATTGAGGCTTTTATTAATCAG	21	GTTCCTATATGCTTGTGATGA	21	293	6126
8	1762	chr3R	7492347	7492543	1360	AATACAAGATATTGATACATC	21	CGAAAGTTCATACAAATGTATA	22	374	4480
9	2406	chrX	141518	141720	blood	TTAATGAAGCCGTTGGGCTAGG	22	CTGTGGTGAATTTGCAGCCTT	21	379	7790
10	2413	chrX	325934	326139	1731	TTTTACACTAAGAGTACCGGC	21	CGACCTAATAGTGCGGAAGCT	21	349	4998
11	2427	chrX	1082310	1082488	297	AGATAAGACGCCTATTAAGTG	21	CATTGTATGTTGCATCTGTAG	21	373	7368
12	32	chr2L	522270	522480	297	ATTTCGAATACGTATCGAATG	21	CATTTCCCTGGCAACTTTGCA	21	386	7381
13	**2165**	**chr3R**	**2.3E+07**	**2.3E+07**	**roo**	**ATCGCTTTTCCAATAAATTGCAT**	**23**	**CAGAGACAGCGATTTCAAAT**	**21**	**367**	**9459**
14	**2531**	**chrX**	**6005755**	**6005959**	**roo**	**ACCTTCACGCAGTCCAGCTTCG**	**22**	**GCACCTGGCAGGGGATTTTAATT**	**23**	**381**	**9473**
15	**1162**	**chr3L**	**957937**	**958157**	**roo**	**TATACAACCACGAGTTCTTGGAAG**	**24**	**ACAACATAAATAATTGCAGAGGC**	**23**	**296**	**9388**

PRIMER SEQUENCE FOR S2c1 DEPLETION

Sr. No.	Chr_cc	Chr_3p	Chr_coord_	Chr_coord_	repName	Forward Primer Sequence	Forward Primer Length	Reverse Primer Sequence	Reverse PrimerLength	Length witl	Length withTE depletic
1		chr2L	7350447	7350536	roo	CTCGCGGTGAGTCAACCCCAG	21	GAGAATTTCGGTGGTTACGCCC	22	7480	5757
2		chrX	6571315	6571399	Tirant	CCATATTTAGACAATTATTC	20	CAGTCCCACAAGTGCGAGCGG	23	8802	1240
3		chr2R	21753581	21753663	roo	GAACACTTTAAGCCCAGTTTG	21	GAACTTCAACAATATCCGAT	20	833	405
4		chr3L	3069878	3069958	Tirant	GACTCCTATGACGCAATACAG	21	CATTCTGTATACAAACAAATC	22	8912	1350
5		chr3L	7889043	7889129	Stalker4	TATGCCTATAATCAATAAATC	21	CTTCATGTTGCTACATAAATCAT	23	2873	1107
6		chr3L	7987706	7987783	NOF	TCCCGAATAGCACCGAATTTC	21	CAATTAATGTGTGAATCTCAAG	22	2650	620
7		chr3L	15426613	15426700	opus	TTTTGGTCGAGTTCTGTTCTG	21	GAAGTCTTTATAAGTCAAGAA	21	3160	2642
8		chr3R	24072571	24072628	jockey	CATTTCTTTGCGCCTGCCATG	21	GACAGCACTTGGATGATCCA	20	341	150
9		chrX	3842569	3842647	pogo	GATCCTTGTCCCGGATAAGTC	21	GATTTTCCCGTTCTGGAGTAA	21	541	355
10		chr3R	24931935	24932010	Doc	TGCAGAATTGTACAAGAGATG	21	CAGATATGTACAATTCCGGACA	22	5120	395

Table S2 Primer Lists

TIDAL_only_TEIns_sites

Site Number	Chr	Chr_coord_5p	Chr_coord_3p	TE	Coverage_Ratio	RAL-765_Inserts_Annotated.Insert_Code	Space	FW primer name	FW primer sequence	Fw Length	Tm	RV primer name	RV primer sequence	Rv Length	Tm	Predicted Amplicon Size
1	chr2L	8133670	8133752	HMS-Beagl	2.4	chr2L_17_8130001	82	TeIns_TID_1_Fw	GATCAAGTCGAACGAACAATT	21	56.8	TeIns_TID_1_Rv	CGCATGCATATAGGCGGATGC	21	67.7	275
2	chr2L	19834994	19835061	hobo	4.7	chr2L_20_19830001	67	TeIns_TID_2_Fw	TAAATATTTATCATTGTGAGCTG	23	51.9	TeIns_TID_2_Rv	AGTTTTATGGCATTTATGTCG	21	54.6	354
3	chr2L	12227998	12228146	BS	5.3	chr2L_10_12225001	148	TeIns_TID_3_Fw	CCAGCTGGCACAAATGAATG	20	62.6	TeIns_TID_3_Rv	CGTTTTAAGCCCACATTGCTG	21	62.7	274
4	chr2R	12581030	12581105	pogo	5	chr2R_14_12580001	75	TeIns_TID_4_Fw	ATTGAAAGAAAAATAGTCATGC	22	53.1	TeIns_TID_4_Rv	CCACTAAGCGATGACGAACGCG	22	69.1	333
5	chr2R	7621371	7621430	mdg1	4	chr2R_10_7620001	59	TeIns_TID_5_Fw	ATGTCAAGAATCTGAAGTTACG	22	53.8	TeIns_TID_5_Rv	CGTTTGTTTATGGTCCAAATGC	22	61.4	328
6	chr3L	4481564	4481640	roo	4.1	chr3L_14_4480001	76	TeIns_TID_6_Fw	GAGTGCAGACCCTACTATTAG	21	51	TeIns_TID_6_Rv	CACGTTCTTATGTGTATAATTA	22	47.8	337
7	chr3L	18394377	18394445	1360	4.2	chr3L_11_18390001	68	TeIns_TID_7_Fw	TCAAACTATGTGTTTTAAGAGTAT	24	50	TeIns_TID_7_Rv	CAGGGTGGATTCGCTATCAGG	21	64	360
8	chr3R	4885334	4885399	1360	1.7	chr3R_13_4885001	65	TeIns_TID_8_Fw	ATTTATGCCAGTCATGGGCC	20	62	TeIns_TID_8_Rv	CGAGCAAGAATGTACCTCAAGT	22	58.5	354
9	chr3R	20125023	20125087	jockey	2.4	chr3R_12_20125001	64	TeIns_TID_9_Fw	AATTGACTTGCCACGATGTGC	21	63.3	TeIns_TID_9_Rv	CCGAGTGCAATGGGAGTGGGC	21	71.3	360
10	chr3R	11089486	11089562	roo	5.4	chr3R_15_11085001	76	TeIns_TID_10_Fw	TTTGAACACAATTTGGAAGTG	21	56.2	TeIns_TID_10_Rv	CAGTCCGCCCAAGAAACCAAA	21	66.5	359
11	chrX	7785488	7785556	hobo	3.3	chrX_10_7785001	68	TeIns_TID_11_Fw	AAGTTTTATGACCTCCATATG	21	51.1	TeIns_TID_11_Rv	CATTAATTTCGCCAGCACAGC	21	62.9	342
12	chrX	15391856	15391929	roo	4.1	chrX_13_15390001	73	TeIns_TID_12_Fw	TATATATTTCGGTAAAACCCTG	22	53.3	TeIns_TID_12_Rv	CTTTGCATGATAACATTAAGT	21	49.9	341

Table S2 Primer Lists

TIDAL and LnB shared sites

Site Number	Chr	Chr_coord_5p	Chr_coord_3p	TE	Coverage_Ratio	Tidal Insert Code	LnB765	TIDAL	CnT	TEMP	Space	FW primer name	FW primer sequence	Fw Length	Tm	RV primer name	RV primer sequence	Rv Length	Tm	Predicted Amplicon Size
1	chr2L	267478	267547	P-element	4.4	chr2L_18_	chr2L_265	chr2L_265		chr2L_2650	69	TeIns_LnB_1_Fw	ACATACACAGAGACAGGAAAGC	22	56.1	TeIns_LnB_1_Rv	CGAATTTATGGAATTCAGCATT	22	58.2	314
2	chr2L	1167146	1167232	opus	2.6	chr2L_21_	chr2L_116	chr2L_116	chr2L_1		86	TeIns_LnB_2_Fw	GTCCACTATAACGTTACCCAAC	22	55.3	TeIns_LnB_2_Rv	CAAGAGACCATTGCGCAGGT	20	63.6	386
3	chr2L	20800099	20800248	Burdock	5.8	chr2L_22_	chr2L_208	chr2L_208		chr2L_2080	149	TeIns_LnB_3_Fw	CCTCCTGTCCCTACCAACTGG	21	63.1	TeIns_LnB_3_Rv	GACCTCTCCTGTCTCTTATTT	21	51.8	329
4	chr2R	9380884	9380961	3S18	2.4	chr2R_16_	chr2R_938	chr2R_938	chr2R_9	chr2R_9380	77	TeIns_LnB_4_Fw	GGGGAGAAAATGGGAGAAAGC	21	63.3	TeIns_LnB_4_Rv	CCAAGTCGAAAATGTTGCTTC	21	59.7	365
5	chr2R	9442941	9443015	412	3.9	chr2R_13_	chr2R_944	chr2R_944	chr2R_9		74	TeIns_LnB_5_Fw	TGTCAAGAATCTGAAGTTACG	21	53.1	TeIns_LnB_5_Rv	GATTCGTTTGTTTATGGTCCA	21	57.5	331
6	chr2R	11578561	11578638	blood	5.6	chr2R_14_	chr2R_115	chr2R_115	chr2R_1	chr2R_1152	77	TeIns_LnB_6_Fw	CATTTTTGAGTGCAGACCCTAC	22	58.8	TeIns_LnB_6_Rv	TGTCAAGTGTATATTGCTTCT	21	50.1	372
7	chr3L	934796	934851	hobo	8.3	chr3L_15_	chr3L_930	chr3L_930			55	TeIns_LnB_7_Fw	ATACGGAATTCAAACTATGTG	21	52	TeIns_LnB_7_Rv	TGGATTCGCTATCAGGATGTG	21	61	364
8	chr3L	22566716	22566793	hobo	4.3	chr3L_32_	chr3L_225	chr3L_225			77	TeIns_LnB_8_Fw	TTGCACTGCTATTTGCAGAAG	21	59.3	TeIns_LnB_8_Rv	CAATCGTCTCCATTTTGTATG	21	55.7	367
9	chr3R	18968019	18968094	Burdock	3.8	chr3R_19_	chr3R_189	chr3R_189		chr3R_1896	75	TeIns_LnB_9_Fw	AGAGTGTTATCTATGGGAGAG	21	50.2	TeIns_LnB_9_Rv	GTCCATTTGTTTGTCCGCTCAC	22	63.9	375
10	chr3R	30626323	30626403	412	5.2	chr3R_22_	chr3R_306	chr3R_306		chr3R_3062	80	TeIns_LnB_10_Fw	GAAATCCAAATGCCCGTTTCATG	23	66.5	TeIns_LnB_10_Rv	CTAGATTAGTGGCCATATTAAAG	23	52.3	379
11	chrX	6487781	6487856	blood	4.3	chrX_18_6	chrX_6485	chrX_6485	chrX_64		75	TeIns_LnB_11_Fw	TTTCATTCAAAATGCTTATGC	21	55.6	TeIns_LnB_11_Rv	CATGTAATGCGTGCATACATG	21	58.5	375
12	chrX	21399401	21399469	pogo	3.6	chrX_12_2	chrX_2139	chrX_2139		chrX_21395	68	TeIns_LnB_12_Fw	CGTGGAGGTAAGTGAAATTTC	21	56.4	TeIns_LnB_12_Rv	CGATGCTGCAACAAAACAAAT	21	61	321

Table S2 Primer Lists

CnT Only sites

Site Number	Insert_code	Chr	Start	Stop	Space	FW primer name	FW primer sequence	Fw Length	Tm	RV primer name	RV primer sequence	Rv Length	Tm	Predicted Amplicon Size
1	chrX_4330001	chrX	4334105	4334327	222	TeIns_CnT_1_Fw	CTAGGCTACTCCTAAAAGATATC	23	50.8	TeIns_CnT_1_Rv	ACGAGGCTAAAATAAAGGCCC	21	60.8	376
2	chrX_14000001	chrX	14001033	14007159	6126	TeIns_CnT_2_Fw	AACAAGAGAGAATGCTATAGTCG	23	54.8	TeIns_CnT_2_Rv	GATGAATTTTCGCAATTCACA	21	58.6	6092
3	chr3R_7695001	chr3R	7696150	7703338	7188	TeIns_CnT_3_Fw	CCACACTTATATTA ATTGGCA	22	53.1	TeIns_CnT_3_Rv	GATCTCTCGCCGCTGTCACTG	21	65.8	7171
4	chr3R_14020001	chr3R	14021702	14023463	1761	TeIns_CnT_4_Fw	CAGTTTGTCAAGAAACTGTTTAC	23	53.6	TeIns_CnT_4_Rv	CAGTTTGTCAAGAAACTGTTT	21	52.2	1734
5	chr3L_4360001	chr3L	4361048	4361901	853	TeIns_CnT_5_Fw	AGCTCAAAGAAGCTGGGGTCG	21	65.2	TeIns_CnT_5_Rv	CCATGAAATGAAACAATGATTAA	23	56.7	844
6	chr3L_17180001	chr3L	17180645	17180660	15	TeIns_CnT_6_Fw	ACGATTTTGCCGCTCACGC	19	66.8	TeIns_CnT_6_Rv	GCCGAAATTTGTAGTTAAAAT	21	52.2	394
7	chr2R_6945001	chr2R	6948454	6948972	518	TeIns_CnT_7_Fw	TCTAAAGTCGTGTTCCCTGAC	21	56.9	TeIns_CnT_7_Rv	CGTAAATTAAATTGGTTGAGGA	22	56	493
8	chr2R_13060001	chr2R	13064762	13064861	99	TeIns_CnT_8_Fw	TATATTTATGCCAGTCATGGG	21	55.2	TeIns_CnT_8_Rv	CGAGCAAGAATGTACCTCAAGT	22	58.5	357
9	chr2L_1805001	chr2L	1809645	1810122	477	TeIns_CnT_9_Fw	TTCAGGATTAGTTTGTCTAACAA	23	53.6	TeIns_CnT_9_Rv	ATTCCGAGAAATTTATTCCGA	21	57.7	447
10	chr2L_9780001	chr2L	9783398	9783415	17	TeIns_CnT_10_Fw	ATACGGGGTGCTGGTAAACTCC	22	63.4	TeIns_CnT_10_Rv	CGTCTTTGTCAACGAATCAATTG	23	61.7	394
11	chr4_180001	chr4	180927	183823	2896	TeIns_CnT_11_Fw	CTTTCTTATACGCATACATTG	21	50.6	TeIns_CnT_11_Rv	CGGCCATACACATTGGTCTTG	21	63.5	2892
12	chr4_1000001	chr4	999517	1001425	1908	TeIns_CnT_12_Fw	GTTCGTATTTACACTAGCTAAGG	23	52.4	TeIns_CnT_12_Rv	GTTACTCAGCAAAACGTAGTG	21	52.9	2008

Table S2 Primer Lists

TEMP Only Sites

Site Number	Chr	Start	Stop	rounder	InsertCode	Space	FW primer name	FW primer sequence	Fw Length	Tm	RV primer name	RV primer sequence	Rv Length	Tm	Predicted Amplicon Size
1	chr2L	9850436	9850443	9850001	chr2L_9850001	7	TeIns_TEM_1_Fw	GGTTGTCGTTGTCCTTCCTG	20	60.5	TeIns_TEM_1_Rv	GGTTCAGGTCAGGCAGGCAG	20	62	383
2	chr2L	1558495	1558995	1555001	chr2L_1555001	500	TeIns_TEM_2_Fw	CAGACAAAAGTAATGTGAAAC	21	50	TeIns_TEM_2_Rv	CAACCATTAATTGCTATATCA	21	50.7	482
3	chr2R	5334166	5334666	5330001	chr2R_5330001	500	TeIns_TEM_3_Fw	CCGTGTAGGCTTACAAGCAATAG	23	60.2	TeIns_TEM_3_Rv	CTTGTGGTGCAAAGTCAAATG	21	59.2	494
4	chr2R	2095477	2095485	2095001	chr2R_2095001	8	TeIns_TEM_4_Fw	GTCGACTATGTATGGAGGCTGCG	23	65.2	TeIns_TEM_4_Rv	CTTACAATTAGATTTAATGCTTT	23	50	405
5	chr3L	9571935	9572435	9570001	chr3L_9570001	500	TeIns_TEM_5_Fw	TCCGCAAATGTAAGGTATAGTG	22	57	TeIns_TEM_5_Rv	GTTTCGATGGATACGAAAACA	21	57.6	482
6	chr3L	5056174	5056674	5055001	chr3L_5055001	500	TeIns_TEM_6_Fw	CAATCCGCACTCAGACACTCAC	22	63.2	TeIns_TEM_6_Rv	GGCTCTAATCGTTTAGTTATAAA	25	52.1	499
7	chr3R	5035704	5035739	5035001	chr3R_5035001	35	TeIns_TEM_7_Fw	TGAATTTTACATTAAAGGCGG	21	57	TeIns_TEM_7_Rv	AATCCGAACCCGAATCGGAA	20	65.8	296
8	chr3R	1211275	1211285	1210001	chr3R_1210001	10	TeIns_TEM_8_Fw	TAGAACTATAGATCTAGCTAAGGG	24	50.8	TeIns_TEM_8_Rv	GTGATTGCCTTAGTTTTGTTT	21	53.8	398
9	chrX	294866	294869	290001	chrX_290001	3	TeIns_TEM_9_Fw	ATTGTATGCATAATTTACCACTT	23	52.6	TeIns_TEM_9_Rv	CGCATTCAGCTTCGGGTCGAC	21	61.8	395
10	chrX	5410656	5411156	5410001	chrX_5410001	500	TeIns_TEM_10_Fw	GTTAATAGTAATTAGCACATTGCA	24	53.5	TeIns_TEM_10_Rv	GAATCTCCGATATCATCAACGA	22	59	488
11	chrY	203747	204247	200001	chrY_200001	500	TeIns_TEM_11_Fw	CGTATCAGACAATCTAGTAGCAG	23	53.8	TeIns_TEM_11_Rv	GGACAACACCATTTCATAAT	20	52.3	487
12	chrY	1565254	1565753	1565001	chrY_1565001	499	TeIns_TEM_12_Fw	TCATCAAATCGGAGACACTCC	21	60.1	TeIns_TEM_12_Rv	ATGAATTTGTTCTATTTGAACTC	23	52.2	645

Primers for TE InDels In the ISO1-BL Line

TE Insertions

Site Number	Chr	Chr_coord_5p	Chr_coord_3p	TE	Coverage_Ratio	FW primer name	FW primer sequence	Fw Length	Tm	RV primer name	RV primer sequence	Rv Length	Tm	Amplicon Size
1	chr2R	2E+07	19768307	copia	4.5	ISO1_BL_FI1	TGCCTAGAATTTGCTATGCCGC	22	65.5	ISO1_BL_RI1	GTTTAAGTAAGGATCGATTTTAG	23	51.5	295bp
2	chr3L	4390337	4390468	roo	4.5	ISO1_BL_FI2	GTTGGACAAAGGAATCGGTAATG	23	61.8	ISO1_BL_RI2	GCATTCTGTACTGTTCCCAGCA	22	62.5	330bp
3	chr2R	8528862	8528983	I-element	4.4	ISO1_BL_FI3	CCATCAAAACGCACATATTGC	21	61.3	ISO1_BL_RI3	TTTCCGCTTTCCCACTTTC	19	60.1	273bp
4	chr3R	3E+07	29777247	F-element	17.9	ISO1_BL_FI4	TGGCGCTACGCTAATGAAATC	21	62.4	ISO1_BL_RI4	GCGGTCCTGATAAACGTATTT	21	58.2	414bp
5	chr2L	1068653	1068778	FB	4.2	ISO1_BL_FI5	CTGATAAGCGGCAGCATAAGCAG	23	65.3	ISO1_BL_RI5	GAGCCAAGCTTAGGCTGCAATT	22	64.1	517bp
6	chrX	6542904	6543018	I-element	4.3	ISO1_BL_FI6	TTGAAGTTGGCCATTTACGGG	21	64.1	ISO1_BL_RI6	GGACGCCAATGGAGCTTATTG	21	63.9	501bp
7	chrX	1.9E+07	19018748	copia	4.7	ISO1_BL_FI7	AAGATTTAGAGACACTCCACC	21	52.6	ISO1_BL_RI7	GCTTAAGGCCGCCGAACAA	19	65.6	469bp
8	chr3R	5066592	5066727	Tirant	4.4	ISO1_BL_FI8	TTATTACCACAGCTAAGAGTC	21	49.1	ISO1_BL_RI8	GATATGCTAATTCGGCGCG	19	61.6	515bp
9	chrX	1.1E+07	10556301	Doc	5.7	ISO1_BL_FI9	AGCCAAGATCTGTCATTTGTC	21	56.8	ISO1_BL_RI9	CAAACGACACAGCGAAGCGTA	21	64.6	363bp
10	chr2L	5999659	5999785	hobo	5.6	ISO1_BL_FI10	GTCGAAGTGCCGGAAACGT	19	63.5	ISO1_BL_RI10	CAAATTCTTAAAGCCAAGGCTT	22	59	479bp
11	chr2R	2.3E+07	22786444	hobo	5.3	ISO1_BL_FI11	AAGAATTTCTTTTCCCAAGCTC	22	58.1	ISO1_BL_RI11	GGTATACTTACCTTTTATGCAG	22	50.6	523bp
12	chr2L	9485261	9485390	hobo	4.2	ISO1_BL_FI12	GAGTGTCGCAGAGTCCGTCTG	21	63.6	ISO1_BL_RI12	CTAGTAGAATGATTTTGGGGCAA	23	59.5	514bp

TE Depletions

Site Number	Chr	Chr_coord_5p	Chr_coord_3p	TE	Coverage_Ratio	FW primer name	FW primer sequence	Fw Length	Tm	RV primer name	RV primer sequence	Rv Length	Tm	Predicted Amplicon Size	RV primer name	RV primer sequence	Rv Length	Tm	Predicted Amplicon Size
1	chrX	6335185	6335241	Tirant	4.6	ISO1_BL_FD1	GCTGATTTACGAATCACTTGAAG	23	58.1	ISO1_BL_RD1	GGCTAGCTGGGGACAATGTTG	21	64	380bp	ISO1_BL_RD1.1	CGTCATTGTATCGCGCTTATA	21	58.8	2757b.p.
2	chr3R	5045818	5045876	HMS-Beagle	4.4	ISO1_BL_FD2	GGTTACACAAGAGCGTGCAGC	21	63.2	ISO1_BL_RD2	CTAAGTCCCTAGCAATCAAGTGA	23	57.7	448bp	ISO1_BL_RD2.1	GCCACTTATTTATAAATGTCA	22	50	7393
3	chr3R	3.2E+07	31691562	Doc	4.3	ISO1_BL_FD3	AAGAGCGACTGAATAACGAAG	21	56.4	ISO1_BL_RD3	CCAGAACAATATGGAATAATT	21	51.1	449bp	ISO1_BL_RD3.1	GTCCATATCTGCTGCCACGC	20	64	5068
4	chr3R	1.7E+07	16756346	diver	3.6	ISO1_BL_FD4	CAATCAAACAGCGCGGCAC	19	64.1	ISO1_BL_RD4	CTGAACTGTATTCTTTCTTTCGC	23	57.4	436 bp	ISO1_BL_RD4.1	GGCCACCTCGGAGTGGCTC	19	65	6493
5	chr2L	1.3E+07	12861728	hobo	3.5	ISO1_BL_FD5	ATTTATCTTATCCTGTTATCCCA	24	53.9	ISO1_BL_RD5	CGACGCAAAACACCGTATTGA	21	64.1	458bp	ISO1_BL_RD5.1	TACTCCAGAATGTGCGTGGAG	21	61.2	1720
6	chr2L	1.8E+07	17951978	hobo	3.5	ISO1_BL_FD6	TAGAGCTAAGCCAGCCAGGAC	21	61	ISO1_BL_RD6	CACTCGAGTATTTTGTGTTGCC	22	59.7	381bp	ISO1_BL_RD6.1	AACTGTTACGAGATGATACAG	21	50	1785
7	chr3L	2927813	2927871	Tirant	3.5	ISO1_BL_FD7	GTCAGAGATAAAGAAGAAACTAA	23	50.5	ISO1_BL_RD7	GACTCGAGTAGCCACTCTCTG	21	55.9	393bp	ISO1_BL_RD7.1	TGTTGTTTTGGTGGTGTATACAA	23	57.9	8527
8	chrX	1.2E+07	11633445	297	2.9	ISO1_BL_FD8B	AACACAGATCAGTGTGTGAGT	21	52.8	ISO1_BL_RD8B	TCTCCTCAATCCAATTTGCAT	21	59.5	430bp	ISO1_BL_RD8.1	CAGAATGGAAATAAATTTGTT	21	51.4	7371
9	chr3L	8481746	8481802	roo	3.3	ISO1_BL_FD9B	AGTAGCTGTGGATCTGTGGC	20	57.9	ISO1_BL_RD9B	ATTTACTTAGGCCTCTGCGTA	21	56	408bp	ISO1_BL_RD9.1	TTGCGTTCTGCATCTGTCAGA	21	63.1	9454
10	chr2R	7608382	7608431	I-element	2.6	ISO1_BL_FD10B	ATTCAAGCGTATTGTTTATTG	21	52.4	ISO1_BL_RD10B	CATAACAAGCCAGCAATTAGTT	22	56.3	317bp	ISO1_BL_RD10.1	TGATCGAAATGGTGTTAAAGTCG	23	61.2	5489
11	chr2R	8398189	8398240	Doc	3	ISO1_BL_FD11B	TCGGAGACGGTGGATAGGTAG	21	61.4	ISO1_BL_RD11B	GTAGGTGCACTTGGAGCACC	20	57.2	399bp	ISO1_BL_RD11.1	GAGGCACGAACTGCTGGCT	19	65	5061
12	chr3R	3.1E+07	31279405	diver	2.5	ISO1_BL_FD12B	AGAGATCCAAATTCAAATATG	21	51.1	ISO1_BL_RD12B	ATCACGTCTGATCGCTATGAC	21	57.8	301bp	ISO1_BL_RD12.1	CTTGGCTAAAACCCACACAC	20	58.1	6402