SUPPLEMENTARY METHODS

Genome assembly and reconciliation

For the reconciliation process, we generated two heterochromatin and two whole genome *de novo* assemblies with Falcon and Canu, independently. For the whole genome *de novo* assemblies, we used Canu v1.2 on all genomic reads with the parameters "genomeSize=160m useGrid=false errorRate=0.035" (Canu 1 assembly) and Falcon v0.3 (Falcon 1 assembly; configuration file is Supplementary text 1). We also generated *de novo* assemblies from the heterochromatin-enriched reads (see Methods) with Canu v1.3 (Canu 2 assembly) and Falcon v0.5 (Falcon 2 assembly; Supplementary text 2 for configuration file). To determine the best parameters for the heterochromatin-enriched Canu 2 assembly, we experimented with assembly conditions by creating *de novo* assemblies for all combinations of bogart em and ee between 0.025 and 0.06 (step size 0.005) for both the default Canu parameters and our repeat-sensitive parameters ("genomeSize=30m stopOnReadQuality=false corMinCoverage=0 corOutCoverage=100 ovlMerSize=31"). The assembly parameters that maximized N50, produced the longest total assembly size, and the longest contig length was bogart em and ee = 0.045. We therefore chose this assembly to represent Canu 2 for subsequent reconciliation steps. For the Falcon 2 assembly, we made assemblies by varying the minimal overlap length in the string graph (fc_ovlp_to_graph min_len 1000 and 6000) and chose min_len 1000 to represent the Falcon 2 assembly. In the next steps, we combined our *de novo* total and heterochromatin-enriched assemblies with reference assemblies from CHAKRABORTY *et al.* 2016 (ISO_merged assembly) and release 6 (HOSKINS *et al.* 2015).

We corrected any assembly errors manually. Our manual curation was primarily in detecting misassemblies in genic and intergenic regions according to the gene order in R6 using 154 heterochromatic and telomeric genes as our BLAST reference. After each reconciliation step, we split contigs with incorrect gene structures or genes from different chromosomal arms, as these likely were inappropriately merged by quickmerge or assemblers (CHAKRABORTY *et al.* 2016). We first reconciled Falcon 1 and Canu 2 using Canu 2 as the reference (Merged 1). Merged 1 was reconciled with Falcon 2 using Merged 1 as the reference (Merged 2). We combined Merged 2 with the major chromosome arms in R6 (2L, 2R, 3L, 3R, 4, and X) using `cat` to create the Merged 3 assembly. To fill the gaps in Merged 3, we reconciled Merged 3 and ISO_merged (CHAKRABORTY *et al.* 2016). using Merged 3 as the reference (Merged 4). Finally, the Merged 4 was reconciled with Canu 1 using Merged 4 as the reference (Final Merged). We corrected remaining assembly errors in the Final Merged assembly base on BLAST results and previous studies (see Methods). We polished the resulting final assembly with quiver and Pilon and used this version of the assembly for all subsequent analyses. We determined the order for the reconciliation process by: *1)* using combinations that improved the contiguity while retaining completeness; *2)* avoiding large-scale misassemblies due to the reconciliation process; and *3)* the ability to fill gaps (*e.g.* Canu 1 was useful for filling some gaps left in Merged 4).

Estimating Y-linked gene conversion rates

Because Y-linked gene families do not undergo crossing over, we expect gene conversion to be the primary mechanism homogenizing different gene copies. We assume that there are a total of n copies of a gene, where x genes have the variant site that differentiates the copies, and for simplicity, any of the n-1 gene copies can convert a gene with equal probability. We also assume that there is no change in copy number. The fraction of differences between two gene copies at any generation n is given by d_n.

$$d_n = x(n-x)$$

The effect of each gene conversion event will happen between copies with different SNPs or without SNPs. After the gene conversion, the divergence will be

$$d_{n'} = \frac{x(n-x)(x+1)(n-x-1)}{n(n-1)} + \frac{x(n-x)(x-1)(n-x+1)}{n(n-1)} + \left(1 - \frac{2x(n-x)}{n(n-1)}\right) x(n-x)$$

We can calculate the expected effect of each gene conversion on divergence.

$$E(\Delta d) = d_{n'} - d_n = -2\frac{x(n-x)}{n(n-1)} = -\pi$$

We assume parameter c is the rate at which a pair of gene copies homogenize each other per generation, and corresponds to Ohta's α (OHTA 1982). The divergence between copies is originated from point mutation with rate, u. If the divergence of gene family is only affected by gene conversion and mutations and the current divergence is under the gene conversion and mutation balance, we can derive,

$$E(\Delta d) \times c/2 + u \times (n-1) = 0$$

$$c = \frac{2u(n-1)}{\pi}$$

We can show that equation is equivalent to Rozen's equation (ROZEN *et al.* 2003) when n =2 and Ohta's equation (OHTA 1982).

Here c is the rate of homogenized effect between 2 sequences by gene conversions. This rate is twice the rate that gene conversion happens. In addition, we need to consider the gene conversion tract length—we assumed that Y chromosome has the similar gene conversion tract length as other *D. melanogaster* chromosomes and normalize c based on 400 bp tract length of a single event (c_g) (MILLER *et al.* 2012; MILLER *et al.* 2016).

$$c_g = \frac{c}{400 \times 2}$$

Supplementary text 1. Falcon 1 configuration

```
[General]
input_fofn = input.fofn
input_type = raw
length_cutoff = 5000
```

length_cutoff_pr = 5000
jobqueue = production
job_type = local
sge_option_da = -pe smp 8 -q %(jobqueue)s
sge_option_la = -pe smp 2 -q %(jobqueue)s
sge_option_pda = -pe smp 8 -q %(jobqueue)s
sge_option_pla = -pe smp 2 -q %(jobqueue)s
sge_option_fc = -pe smp 24 -q %(jobqueue)s
sge_option_cns = -pe smp 8 -q %(jobqueue)s
pa_concurrent_jobs = 8
ovlp_concurrent_jobs = 8
pa_HPCdaligner_option = -v -dal128 -t8 -e.70 -l1000 -s1000 -M16
ovlp_HPCdaligner_option = -v -dal128 -t8 -h60 -e.96 -l500 -s1000 -M16
pa_DBsplit_option = -x500 -s400
ovlp_DBsplit_option = -x500 -s400
falcon_sense_option = --output_multi --min_idt 0.70 --min_cov 4 --
local_match_count_threshold 2 --max_n_read 200 --n_core 8 --output_dformatq
overlap_filtering_setting = --max_diff 100 --max_cov 100 --min_cov 1 --bestn 10 --
n_core 8

Supplementary text 2. Falcon 2 configuration

[General]
input_fofn = input_fal.fofn
input_type = raw
length_cutoff = -1
seed_coverage = 50
genome_size = 15000000
length_cutoff_pr = 1000
jobqueue = production
job_type = local
sge_option_da = -pe smp 8 -q %(jobqueue)s
sge_option_la = -pe smp 2 -q %(jobqueue)s
sge_option_pda = -pe smp 8 -q %(jobqueue)s
sge_option_pla = -pe smp 2 -q %(jobqueue)s
sge_option_fc = -pe smp 24 -q %(jobqueue)s
sge_option_cns = -pe smp 8 -q %(jobqueue)s
pa_concurrent_jobs = 9
ovlp_concurrent_jobs = 9
pa_HPCdaligner_option = -v -dal128 -t20 -H15000 -e.70 -k18 -w8 -l1000 -s100 -
M24 -b
ovlp_HPCdaligner_option = -v -dal128 -t40 -M24 -k24 -h60 -e.95 -l500 -s100 -
H15000 -b
pa_DBsplit_option = -x500 -s400
ovlp_DBsplit_option = -x500 -s400

falcon_sense_option = --output_multi --min_idt 0.70 --min_cov 1 --max_n_read 200 --n_core 12
overlap_filtering_setting = --max_diff 100 --max_cov 100 --min_cov 1 --bestn 10 --n_core 12

REFERENCES

Chakraborty, M., J. G. Baldwin-Brown, A. D. Long and J. J. Emerson, 2016 Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res 44**:** e147.

Hoskins, R. A., J. W. Carlson, K. H. Wan, S. Park, I. Mendez *et al.*, 2015 The Release 6 reference sequence of the *Drosophila melanogaster* genome. Genome Res 25**:** 445-458.

Miller, D. E., C. B. Smith, N. Y. Kazemi, A. J. Cockrell, A. V. Arvanitakas *et al.*, 2016 Whole-Genome Analysis of Individual Meiotic Events in Drosophila melanogaster Reveals That Noncrossover Gene Conversions Are Insensitive to Interference and the Centromere Effect. Genetics 203**:** 159-171.

Miller, D. E., S. Takeo, K. Nandanan, A. Paulson, M. M. Gogol *et al.*, 2012 A Whole-Chromosome Analysis of Meiotic Recombination in *Drosophila melanogaster*. G3 (Bethesda) 2**:** 249-260.

Ohta, T., 1982 Allelic and nonallelic homology of a supergene family. Proc Natl Acad Sci U S A 79**:** 3251-3254.

Rozen, S., H. Skaletsky, J. D. Marszalek, P. J. Minx, H. S. Cordum *et al.*, 2003 Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423**:** 873-876.