######################

#SNP calling and CMH test

######################

1. The raw reads for each samples of founder population (F0) can be downloaded using the SRA run accession numbers provided in S7A Table in the manuscript. Trimming, mapping and filtering of reads was performed as described in Barghi et al. (Barghi N, Tobler R, Nolte V, Schlötterer C. *Drosophila simulans* : A Species with Improved Resolution in Evolve and Resequence Studies. G3 (Bethesda). 2017;7:2337–43).

2. A mpileup file was made for bam files of all 10 libraries of F0 and converted to sync file with minimum quality of 40 (bam file nomenclature is the same as in S7A Table). Samtools version version 1.2 was used. *popoolation2* software is from Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011;27(24):3435–6.

*samtools mpileup --max-depth 8000 -6 -B -Q 0 -f reference\_genome.fa Dsim\_Fl\_Base\_1.bam Dsim\_Fl\_Base\_2.bam Dsim\_Fl\_Base\_3.bam Dsim\_Fl\_Base\_4.bam Dsim\_Fl\_Base\_5.bam Dsim\_Fl\_Base\_6.bam Dsim\_Fl\_Base\_7.bam Dsim\_Fl\_Base\_8.bam Dsim\_Fl\_Base\_9.bam Dsim\_Fl\_Base\_10.bam | java -Xmx20g -jar /popoolation2\_1201/mpileup2sync.jar --input /dev/stdin --output F0\_Q40.sync --fastq-type sanger --min-qual 40 --threads 12*

3. SNPs can be called using call\_SNP.py

*python call\_SNP.py --input F0\_Q40.sync --output F0\_Q40\_SNPs.sync*

4. Identify indels and mask indels, TEs and Y-translocated genes

4a. The raw reads for each samples of founder and evolved populations can be downloaded using the SRA run accession numbers provided in S7A Table in the manuscript. Trimming, mapping and filtering of reads was performed as described in Barghi et al. (Barghi N, Tobler R, Nolte V, Schlötterer C. *Drosophila simulans* : A Species with Improved Resolution in Evolve and Resequence Studies. G3 (Bethesda). 2017;7:2337–43). Make a mpileup file of all libraries (10 replicates and 7 timepoints):

*samtools mpileup --max-depth 8000 -6 -B -Q 0 -f reference\_genome.fa Dsim\_Fl\_Base\_1.bam Dsim\_Fl\_Base\_2.bam Dsim\_Fl\_Base\_3.bam Dsim\_Fl\_Base\_4.bam Dsim\_Fl\_Base\_5.bam Dsim\_Fl\_Base\_6.bam Dsim\_Fl\_Base\_7.bam Dsim\_Fl\_Base\_8.bam Dsim\_Fl\_Base\_9.bam Dsim\_Fl\_Base\_10.bam Dsim\_Fl\_Hot\_F10\_1.bam Dsim\_Fl\_Hot\_F10\_2.bam Dsim\_Fl\_Hot\_F10\_3.bam Dsim\_Fl\_Hot\_F10\_4.bam Dsim\_Fl\_Hot\_F10\_5.bam Dsim\_Fl\_Hot\_F10\_6.bam Dsim\_Fl\_Hot\_F10\_7.bam Dsim\_Fl\_Hot\_F10\_8.bam Dsim\_Fl\_Hot\_F10\_9.bam Dsim\_Fl\_Hot\_F10\_10.bam Dsim\_Fl\_Hot\_F20\_1.bam Dsim\_Fl\_Hot\_F20\_2.bam Dsim\_Fl\_Hot\_F20\_3.bam Dsim\_Fl\_Hot\_F20\_4.bam Dsim\_Fl\_Hot\_F20\_5.bam Dsim\_Fl\_Hot\_F20\_6.bam Dsim\_Fl\_Hot\_F20\_7.bam Dsim\_Fl\_Hot\_F20\_8.bam Dsim\_Fl\_Hot\_F20\_9.bam Dsim\_Fl\_Hot\_F20\_10.bam Dsim\_Fl\_Hot\_F30\_1.bam Dsim\_Fl\_Hot\_F30\_2.bam Dsim\_Fl\_Hot\_F30\_3.bam Dsim\_Fl\_Hot\_F30\_4.bam Dsim\_Fl\_Hot\_F30\_5.bam Dsim\_Fl\_Hot\_F30\_6.bam Dsim\_Fl\_Hot\_F30\_7.bam Dsim\_Fl\_Hot\_F30\_8.bam Dsim\_Fl\_Hot\_F30\_9.bam Dsim\_Fl\_Hot\_F30\_10.bam Dsim\_Fl\_Hot\_F40\_1.bam Dsim\_Fl\_Hot\_F40\_2.bam Dsim\_Fl\_Hot\_F40\_3.bam Dsim\_Fl\_Hot\_F40\_4.bam Dsim\_Fl\_Hot\_F40\_5.bam Dsim\_Fl\_Hot\_F40\_6.bam Dsim\_Fl\_Hot\_F40\_7.bam Dsim\_Fl\_Hot\_F40\_8.bam Dsim\_Fl\_Hot\_F40\_9.bam Dsim\_Fl\_Hot\_F40\_10.bam Dsim\_Fl\_Hot\_F50\_1.bam Dsim\_Fl\_Hot\_F50\_2.bam Dsim\_Fl\_Hot\_F50\_3.bam Dsim\_Fl\_Hot\_F50\_4.bam Dsim\_Fl\_Hot\_F50\_5.bam Dsim\_Fl\_Hot\_F50\_6.bam Dsim\_Fl\_Hot\_F50\_7.bam Dsim\_Fl\_Hot\_F50\_8.bam Dsim\_Fl\_Hot\_F50\_9.bam Dsim\_Fl\_Hot\_F50\_10.bam Dsim\_Fl\_Hot\_F60\_1.bam Dsim\_Fl\_Hot\_F60\_2.bam Dsim\_Fl\_Hot\_F60\_3.bam Dsim\_Fl\_Hot\_F60\_4.bam Dsim\_Fl\_Hot\_F60\_5.bam Dsim\_Fl\_Hot\_F60\_6.bam Dsim\_Fl\_Hot\_F60\_7.bam Dsim\_Fl\_Hot\_F60\_8.bam Dsim\_Fl\_Hot\_F60\_9.bam Dsim\_Fl\_Hot\_F60\_10.bam > F0-F60.mpileup*

4b. Identify indels

*perl /popoolation\_1.2.2/basic-pipeline/identify-genomic-indel-regions.pl --input F0-F60.mpileup --output F0-F60\_indelregions.gtf --indel-window 5 --min-count 167*

4c. combine indels, TEs and Y-translocated genes into a file. *TEfinal.gff* was annotated using the pipeline described in Kofler et al. (Kofler R, Nolte V, Schlötterer C. Tempo and Mode of Transposable Element Activity in Drosophila. Plos Genetics. 2015; 117: e1005406) and *Ytransloc\_regions\_200bp.gff* contains 200-bp flanking the SNPs specific to autosomal genes translocated to the Y chromosome in Tobler et al. (Tobler R, Nolte V, Schlötterer C. High rate of translocation-based gene birth on the *Drosophila* Y chromosome. Proc Natl Acad Sci USA. 2017; 114:44: 11721-11726).

*cat F0-F60\_indelregions.gtf TEfinal.gff Ytransloc\_regions\_200bp.gff > F0-F60\_indel\_TE\_Ytransloc.gtf*

4d. mask indels, TEs and Y-translocated genes

*perl /popoolation2\_1201/indel\_filtering/filter-sync-by-gtf.pl --gtf F0-F60\_indel\_TE\_Ytransloc.gtf --input F0\_Q40\_SNPs.sync --output F0\_Q40\_SNPs\_indel\_TE\_masked.sync*

5. Extract the SNPs for all samples

5a. A mpileup file was made for bam files of all libraries (10 replicates and 7 timepoints) and converted to sync file as follows (bam file nomenclature is the same as in S7A Table):

*samtools mpileup --max-depth 8000 -6 -B -Q 0 -f reference\_genome.fa Dsim\_Fl\_Base\_1.bam Dsim\_Fl\_Base\_2.bam Dsim\_Fl\_Base\_3.bam Dsim\_Fl\_Base\_4.bam Dsim\_Fl\_Base\_5.bam Dsim\_Fl\_Base\_6.bam Dsim\_Fl\_Base\_7.bam Dsim\_Fl\_Base\_8.bam Dsim\_Fl\_Base\_9.bam Dsim\_Fl\_Base\_10.bam Dsim\_Fl\_Hot\_F10\_1.bam Dsim\_Fl\_Hot\_F10\_2.bam Dsim\_Fl\_Hot\_F10\_3.bam Dsim\_Fl\_Hot\_F10\_4.bam Dsim\_Fl\_Hot\_F10\_5.bam Dsim\_Fl\_Hot\_F10\_6.bam Dsim\_Fl\_Hot\_F10\_7.bam Dsim\_Fl\_Hot\_F10\_8.bam Dsim\_Fl\_Hot\_F10\_9.bam Dsim\_Fl\_Hot\_F10\_10.bam Dsim\_Fl\_Hot\_F20\_1.bam Dsim\_Fl\_Hot\_F20\_2.bam Dsim\_Fl\_Hot\_F20\_3.bam Dsim\_Fl\_Hot\_F20\_4.bam Dsim\_Fl\_Hot\_F20\_5.bam Dsim\_Fl\_Hot\_F20\_6.bam Dsim\_Fl\_Hot\_F20\_7.bam Dsim\_Fl\_Hot\_F20\_8.bam Dsim\_Fl\_Hot\_F20\_9.bam Dsim\_Fl\_Hot\_F20\_10.bam Dsim\_Fl\_Hot\_F30\_1.bam Dsim\_Fl\_Hot\_F30\_2.bam Dsim\_Fl\_Hot\_F30\_3.bam Dsim\_Fl\_Hot\_F30\_4.bam Dsim\_Fl\_Hot\_F30\_5.bam Dsim\_Fl\_Hot\_F30\_6.bam Dsim\_Fl\_Hot\_F30\_7.bam Dsim\_Fl\_Hot\_F30\_8.bam Dsim\_Fl\_Hot\_F30\_9.bam Dsim\_Fl\_Hot\_F30\_10.bam Dsim\_Fl\_Hot\_F40\_1.bam Dsim\_Fl\_Hot\_F40\_2.bam Dsim\_Fl\_Hot\_F40\_3.bam Dsim\_Fl\_Hot\_F40\_4.bam Dsim\_Fl\_Hot\_F40\_5.bam Dsim\_Fl\_Hot\_F40\_6.bam Dsim\_Fl\_Hot\_F40\_7.bam Dsim\_Fl\_Hot\_F40\_8.bam Dsim\_Fl\_Hot\_F40\_9.bam Dsim\_Fl\_Hot\_F40\_10.bam Dsim\_Fl\_Hot\_F50\_1.bam Dsim\_Fl\_Hot\_F50\_2.bam Dsim\_Fl\_Hot\_F50\_3.bam Dsim\_Fl\_Hot\_F50\_4.bam Dsim\_Fl\_Hot\_F50\_5.bam Dsim\_Fl\_Hot\_F50\_6.bam Dsim\_Fl\_Hot\_F50\_7.bam Dsim\_Fl\_Hot\_F50\_8.bam Dsim\_Fl\_Hot\_F50\_9.bam Dsim\_Fl\_Hot\_F50\_10.bam Dsim\_Fl\_Hot\_F60\_1.bam Dsim\_Fl\_Hot\_F60\_2.bam Dsim\_Fl\_Hot\_F60\_3.bam Dsim\_Fl\_Hot\_F60\_4.bam Dsim\_Fl\_Hot\_F60\_5.bam Dsim\_Fl\_Hot\_F60\_6.bam Dsim\_Fl\_Hot\_F60\_7.bam Dsim\_Fl\_Hot\_F60\_8.bam Dsim\_Fl\_Hot\_F60\_9.bam Dsim\_Fl\_Hot\_F60\_10.bam | java -Xmx20g -jar /popoolation2\_1201/mpileup2sync.jar --input /dev/stdin --output F0-F60\_Q20.sync --fastq-type sanger --min-qual 20 --threads 12*

5b. Extract the called SNPs (at stage 4d) for all samples

*awk 'BEGIN{OFS="\t"} NR==FNR {f1[$1$2] = $0; next} ($1$2 in f1) {print f1[$1$2], $0}' F0\_Q40\_SNPs\_indel\_TE\_masked.sync F0-F60\_Q20.sync | cut -f 1-3,17-86 > F0-F60SNP.sync*

6. Run CMH test on F0 and F60 samples

6a. Separate F0 and F60 samples

*cut -f 1-13,64-73 F0-F60SNP.sync > F0F60SNP.sync*

6b. Run CMH test

*perl /popoolation2/cmh-test.pl --input F0F60SNP.sync --output F0F60SNP.sync.cmh --min-count 10 --min-coverage 30 --max-coverage 423 --population 1-11,2-12,3-13,4-14,5-15,6-16,7-17,8-18,9-19,10-20 --remove-temp*

F0F60SNP.sync.cmh is already filtered for low and high coverage SNPs and is the **final SNP set**. The SNP set and -log10-transformed p-values of CMH test is provided in F0-F60SNP\_CMH\_FET\_blockID.sync.zip file (Dryad Digital Repository: https://doi.org/10.5061/dryad.rr137kn) columns 1 to 74.