

Supporting Information – Genome-wide tests for  
introgression between cactophilic *Drosophila* implicate a role  
of inversions during speciation

Konrad Lohse<sup>1</sup>, Magnus Clarke<sup>1</sup>, Michael G. Ritchie<sup>2</sup>, William J. Etges<sup>3</sup>

<sup>1</sup> Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

<sup>3</sup> School of Biology, University of St. Andrews, St. Andrews KY16 9TH, UK

<sup>3</sup> Program in Ecology and Evolutionary Biology, Department of Biological Sciences, University of  
Arkansas, Fayetteville, Arkansas 72701, USA

Table S1: Origins of the three populations of *Drosophila mojavensis* and *D. arizonae* in this study and numbers of flies used to establish laboratory populations. Flies were collected over banana baits in nature unless otherwise noted.

Species	Population	Stock #	Latitude	Longitude	# of founders	mean cov.
<i>D. arizonae</i>	Ejido Puerto Arturo, Sonora	A989	28°25'08.04"N	111°24'38.62"W	3	49.1
<i>D. mojavensis</i>	Las Bocas, Sonora	LB09	26°42'57.73"N	109°17'55.42" W	1264 <sup>1</sup>	25.4
<i>D. mojavensis</i>	Punta Onah, Sonora	PO88	29°5'23.15"N	112°10'15.59" W	2426	28.9
<i>D. mojavensis</i>	Bahía de Concepción, Baja C.	A975	26°32'03.72"N	111°44'03.72" W	15	23.6
<i>D. mojavensis</i>	Santiago, Baja C. Sur	A976	23°28'07.91"N	109°40'45.8" W	45 <sup>2</sup>	24.9

<sup>1</sup> includes 2298 adults reared from 10 *Stenocereus gummosus* cactus rots.

<sup>2</sup> includes 2 adults reared from a *Stenocereus gummosus* cactus rot.

Table S2: Summary of scaffolds analysed: Composition (% exon), total length of mapped reads before and after filtering and average mapping quality (MQ) of *D. arizonae* reads mapped against the *D. mojavensis* reference genome.

Chrom.	scaffolds	MQ	Length (Mb)	% exon	filtered Length (Mb)	%	Inversions
2	6540	85.2	34.1	25.6	26.4	77.4	2q, 2s, 2r
3	6500	82.9	32.4	20.9	21.7	67.0	3d, 3p2
4	6680	86.0	24.8	24.8	18.9	76.3	
5	6496	85.6	26.9	26.0	20.5	76.2	
X	6473, 6328, 6308, 6359	82.2	29.3	17.2	21.7	74.2	Xe
Total			147.4		109.3	74	

Table S3: Breakpoint coordinates of inversions fixed between *D. mojavensis* and *D. arizonae*.

Inversion	Flanking proteins	BP location	Reference
2q	CG31528	10,420,224 – 10,422,204	(Guillen & Ruiz, 2012)
	CG1213	31,254,883 – 31,255,399	
2r	Hsp68b	7,230,145 – 7,321,956	
	Hel89B	15,160,462 – 15,162,581	
2s	CG10214	13,149,238 – 13,149,496	
	CG34135	25,966,954 – 25,968,814	
Xe		12,078,196 (+/- 6 bp)	(Runcie & Noor, 2009)
		1,298,611 (+/- 6 bp)	

Table S4: Mean pairwise divergence for exons, introns and intergenic regions. \* indicates chromosomes with fixed inversion differences between *D. mojavensis* and *D. arizonae*.

Chrom.	<i>Dariz/Dmoj, sym</i>			<i>Dariz/Dmoj, allo</i>			<i>Dmoj/Dmoj</i>		
	exon	intron	interg.	exon	intron	interg.	exon	intron	interg.
2*	0.0196	0.0318	0.0254	0.0195	0.0311	0.0255	0.0050	0.0100	0.00774
3*	0.0220	0.0380	0.0296	0.0218	0.0369	0.0294	0.0052	0.0113	0.00944
4	0.0164	0.0258	0.0220	0.0162	0.0252	0.0213	0.0067	0.0118	0.00931
5	0.0163	0.0285	0.0215	0.0161	0.0277	0.0215	0.0065	0.0133	0.00921
X*	0.0151	0.0313	0.0246	0.0150	0.0306	0.0252	0.0047	0.0111	0.00815

Table S5: Counts of sites uniquely shared between *D. mojavensis* and *D. arizonae* in sympatry or allopatry at colinear autosomes. Sites were sampled at least 100kb apart to avoid linkage effects (p-values, binomial sign test); the bottom half shows counts across all sites.

Chromosome	<i>Dariz=Dmoj-LB09</i> $\neq$ <i>Dmoj-A975</i>	<i>Dariz=Dmoj-A975</i> $\neq$ <i>Dmoj-LB09</i>	p-value
4	115	133	0.14
5	144	124	0.12
Total	259	257	0.48
4	76,320	88,006	n/a
5	89,273	79,597	n/a
Total	164,326	177,279	n/a

Table S6: Maximum likelihood estimates of parameters under the IIM model estimated from 250 base intergenic blocks without constraints, i.e.  $M$  and  $\tau$  parameters are free to vary between colinear autosomes, chromosome 2 and chromosome 3. Marginal support for  $\tau_0$  estimates is shown in figure S4.

Chrom.	$M$	$\tau_1$	$\tau_0$
<i>Dariz/Dmoj-LB09, sym</i>			
2	0.365	313 KY	1,580 KY
3	0.300	411 KY	1,810 KY
4 & 5	0.483	245 KY	1,690 KY
<i>Dariz/Dmoj-A975, allo</i>			
2	0.363	326 KY	1,600 KY
3	0.298	412 KY	1,850 KY
4 & 5	0.467	228 KY	1,690 KY

Table S7: Maximum likelihood estimates of parameters under the simplest, supported model of speciation estimated from 500bp intergenic blocks. For comparisons between *D. arizonae* and *D. mojavenis* this is the IIM model with different  $M$  for colinear and each rearranged autosome, for the pairwise analysis of the two *D. mojavenis* populations the IM model.

Comparison	$M_2$	$M_3$	$M_{4\&5}$	$\tau_1$	$\tau_0$
<i>Dariz/Dmoj-LB09</i> , sym	0.38	0.20	0.68	381 KY	1,320 KY
<i>Dariz/Dmoj-A975</i> , allo	0.36	0.19	0.90	337 KY	1,100 KY
<i>Dmoj-LB09/Dmoj-A975</i>	0.19	0.19	0.19	0	259 KY

Table S8: Maximum likelihood estimates of parameters under a model of isolation with initial migration (IIM) which differs between rearranged and colinear autosomes.

Comparison	$M_2$	$M_3$	$M_{4\&5}$	$\tau_1$	$\tau_0$
<i>Dariz/Dmoj-LB09</i> , sym	0.47	0.25	0.89	272 KY	1,290 KY
<i>Dariz/Dmoj-A975</i> , allo	0.45	0.25	0.98	260 KY	1,240 KY
<i>Dariz/Dmoj-PO88</i> , sym	0.92	0.29	0.96	271 KY	1,300 KY
<i>Dariz/Dmoj-A976</i> , allo	0.35	0.30	0.66	272 KY	1,560 KY

Scaled parameters are given in brackets.

Marginal support around these point estimates is shown in figure S4.

Table S9: Mean chromosome-wide divergence between *D. mojavensis* and *D. arizonae* in sympatry (Sonora) and allopatry (Baja) for replicate lines PO88 and A976.

Chrom.	<i>Dariz/Dmoj-PO88</i> , sym	<i>Dariz/Dmoj-A976</i> , allo	<i>Dmoj-PO88/Dmoj-A976</i>
2*	0.0285	0.0283	0.0084
3*	0.0333	0.0331	0.0101
4	0.0235	0.0232	0.0109
5	0.0249	0.0247	0.0118
X	0.0278	0.0277	0.0098

\* chromosomes with fixed inversion differences between *D. arizonae* and *D. mojavensis*.

Figure S1: The effect of filtering on mean chromosome-wide divergence between *D. arizonae* and (allopatric) *D. mojavensis*; the filtering thresholds used are shown as dashed lines. Top) Increasing minimum and maximum sequence coverage (DPlow and DPhigh) has almost no effect on the relative difference in gene divergence observed between colinear (grey=chromosome 4) and rearranged (black=chromosome 2) chromosomes. Bottom) Overfiltering for genotype quality (GQ) artificially reduces divergence.

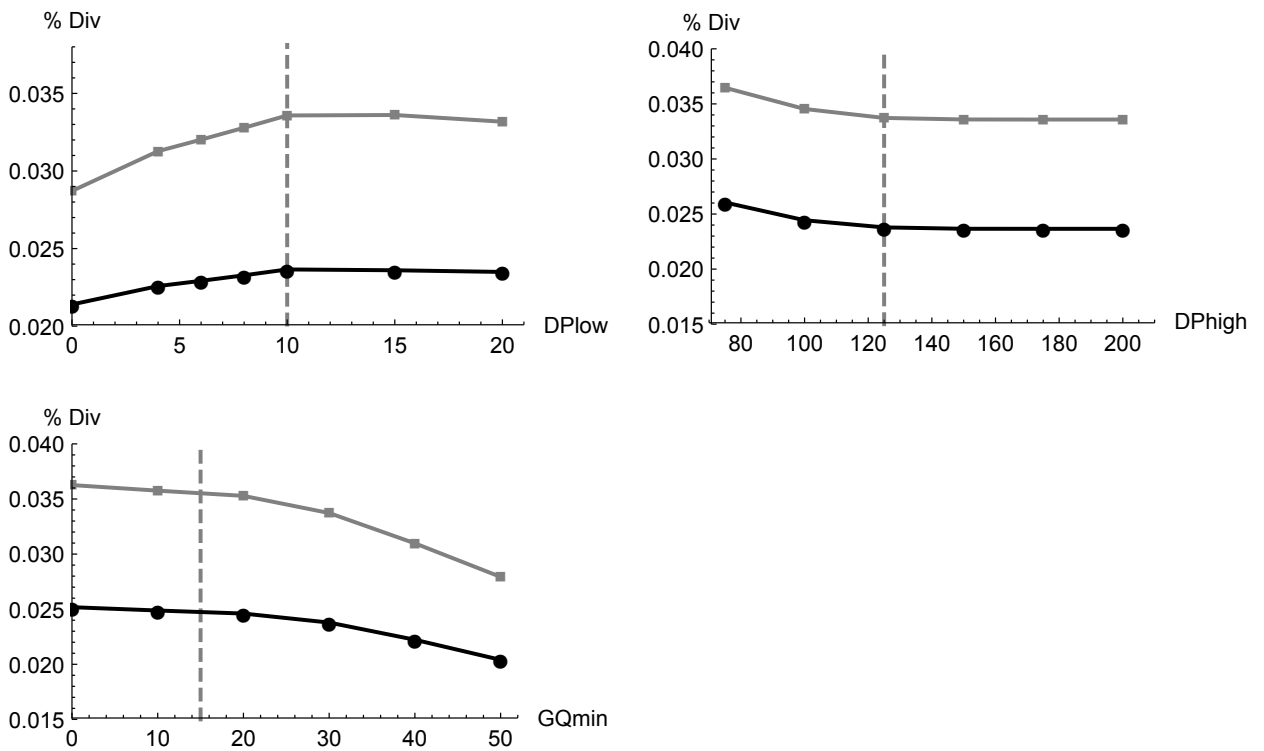


Figure S2: Example IGV screenshot of *D. arizonae* reads mapped to the *D. mojavensis* reference genome. Read pairs mapped to a colinear part of the reference are to the left and right of the unsequenced part of each fragment, i.e. the default read orientation is LR (grey boxes). The coloured boxes show read pairs around either side of the distal breakpoint of inversion 2s whose orientation is reversed (green=LL; Blue=RR).

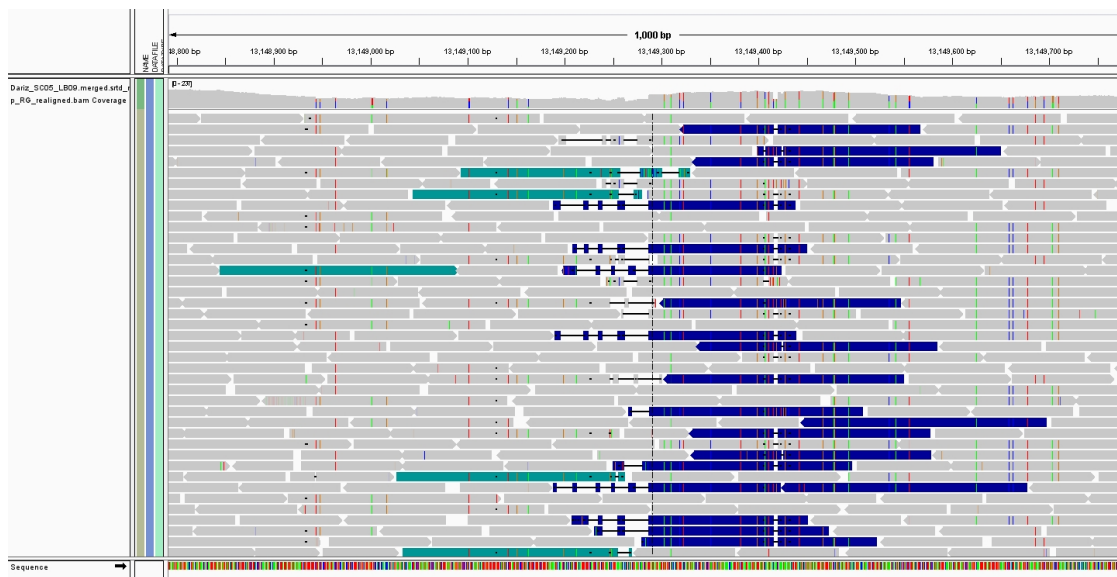


Figure S3: Mean correlation coefficient for the number of divergent sites between *D. mojavensis* (LB09) and *D. arizonae* for pairs of 250 bp intergenic blocks plotted against distance (i.e. # of successive blocks apart). The mean number of blocks in 100kb of sequence is 18.6 (dashed line) and this value was used to correct likelihood analyses for the effect of linkage between blocks (see Methods). For pairs of blocks further apart, the correlation falls off very slowly.

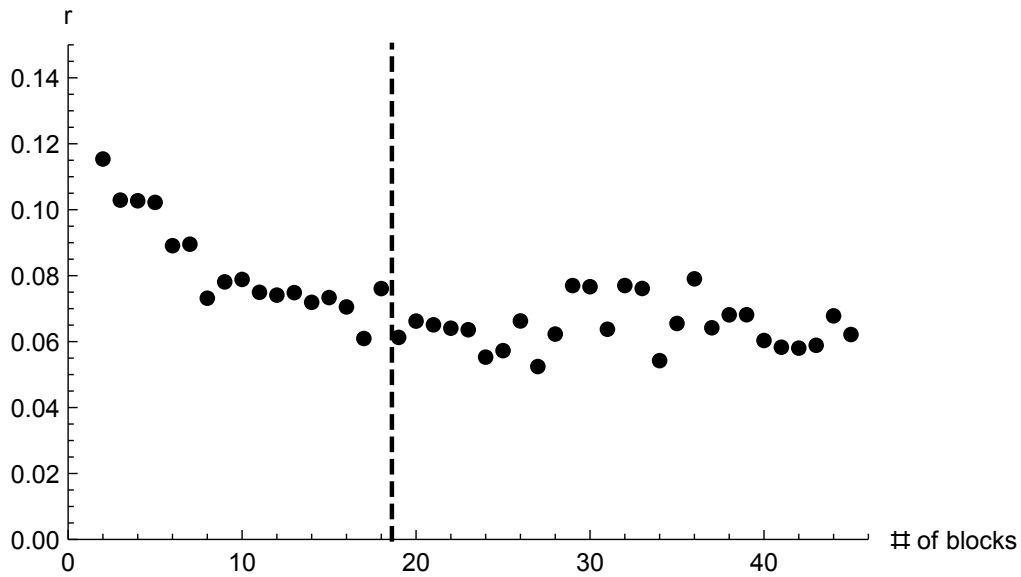
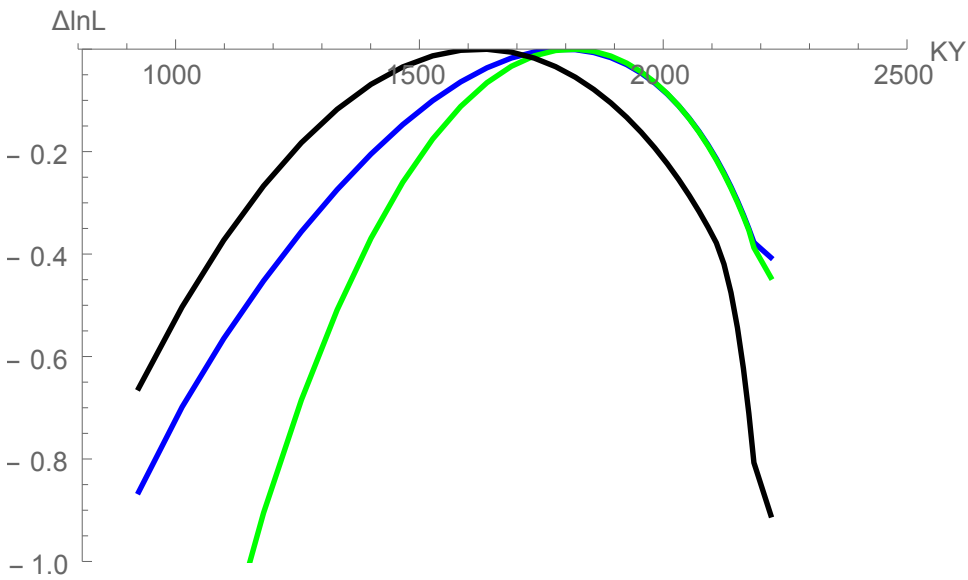




Figure S4: Marginal support ( $\Delta \ln L$ ) for  $\tau_0$  estimated independently for chromosome 2 (blue), 3 (green) and 4& 5 combined (black) (point estimates in Table S6). The broad overlap, agrees with the model comparison results which finds no decrease in support for a model in which  $\tau_0$  is shared by all autosomes.



## References

Guillen, Y. & Ruiz, A. (2012). Gene alterations at *Drosophila* inversion breakpoints provide *prima facie* evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genomics*, 13(1), 53. ISSN 1471-2164. doi:10.1186/1471-2164-13-53.

Runcie, D.E. & Noor, M.A.F. (2009). Sequence signatures of a recent chromosomal rearrangement in *Drosophila mojavensis*. *Genetica*, 136(1), 5–11.