Supplemental methods

Copepod collection and culturing

Copepods were collected with fine-mesh dip nets and large plastic pipettes in the spring of 2018 then transported back to Scripps Institution of Oceanography in 1 L plastic bottles containing water collected from the tidepools. Collections were split into approximately fifteen 200 mL laboratory cultures that were established in 400 mL glass beakers, and held across four incubators for laboratory acclimations. Acclimation conditions (20 °C, 36 ppt and 12:12 light:dark) were maintained for at least one month (approximately one generation) prior to the start of all experiments. Copepods consumed natural algal growth in the cultures as well as a mixture of ground fish flakes and powdered Spirulina that were fed to each culture ad lib.

Inter-population crosses

Prior to mating, male *T. californicus* clasp juvenile females forming a breeding pair (1). Females mate only once, and thus separation of breeding pairs allows the isolation of virgin females for experimental crosses (1). Two sets of reciprocal crosses were made between SD and SC copepods. First, for assessment of variation in ATP synthesis rates (see below), 40 pairs of each population were gently teased apart with a needle (e.g., 2), and males of one population were combined with females of the other population in 10 cm petri dishes containing ~60 mL of filtered seawater. Copepods were allowed to pair, and the dishes were monitored for the appearance of gravid females, which, when observed, were moved to a new dish. These females were allowed to produce multiple egg sacs each in the new dish, and were removed once F_1 offspring were visible. F_1 offspring matured and haphazardly formed breeding pairs. Gravid

females were again moved to a new dish, and were monitored until mature (red) F_2 egg sacs were observed. Second, for isolation of DNA for Pool-seq (see below), 120 pairs of each population were separated, and reciprocal F_2 hybrid egg sacs were obtained as described above. Throughout the experimental crosses holding conditions and feeding routines were the same as those for the initial laboratory acclimations.

Inter-individual variation in developmental time

Variation in developmental rate among individuals was assessed for both parental and F_2 hybrid copepods by measurement of time to metamorphosis (e.g., 3). *T. californicus* development consists of 6 naupliar stages, 5 copepodid stages and the final adult stage (4). The majority of stages are visually cryptic; however, there is an easily distinguished metamorphosis between the final naupliar stage and initial copepodid stage (i.e., copepodid stage I), which can be observed through a microscope at 10X magnification. To score inter-individual differences in developmental rate, gravid females with red egg sacs were pipetted onto filter paper, egg sacs were removed with a fine needle, and dissected egg sacs were placed in filtered seawater in 6-well plates ($\leq$ 4 per well). This procedure synchronizes hatching as dissected mature egg sacs hatch overnight. Offspring were fed Spirulina, and were monitored daily for the appearance of copepodids. Days post hatch (dph) to metamorphosis was scored for all individuals, and copepodids were moved to fresh petri dishes after scoring. In total, offspring from 68 SD egg sacs, 58 SC egg sacs, 352 F_2 SD♀xSC♂ egg sacs (205 for ATP assays and 147 for Pool-seq) and 314 F_2 SC♀xSD♂ egg sacs (115 for ATP assays and 199 for Pool-seq) were scored.

ATP synthesis rates

F$_2$ hybrid copepodids were divided into three developmental time groups: 8-10, 11-13 and ≥17 dph to metamorphosis. Development was allowed to continue, and adults from each group were used for assessment of maximal mitochondrial ATP synthesis rates as in Harada et al. (3). In brief, for each reciprocal cross, 6 pools of 6 adults from each developmental group were moved to petri dishes with fresh filtered seawater and no food overnight. Each pool of copepods was then homogenized in 800 μL of ice-cold homogenization buffer (400 mM sucrose, 100 mM KCl, 70 mM HEPES, 3 mM EDTA, 6 mM EGTA, 1% BSA, pH 7.6) in 1 mL teflon-on-glass homogenizers. Homogenates were transferred to 1.5 mL microcentrifuge tubes, and centrifuged at 1,000 g for 5 min at 4 °C. Supernatants were pipetted to new 1.5 mL tubes, which were then centrifuged at 11,000 g for 10 min at 4 °C. After removal of the supernatants, mitochondrial pellets were resuspended in 55 μL of assay buffer (560 mM sucrose, 100 mM KCl, 70 mM HEPES, 10 mM KH$_2$PO$_4$, pH 7.6). For the ATP synthesis assays, 5 μL of a complex I substrate cocktail (final assay substrate concentrations: 5 mM pyruvate, 2 mM malate and 1 mM ADP) was added to 25 μL of each sample in 0.2 mL strip tubes. This was done twice for each sample: once for the initial ATP concentration determinations and once for the ATP synthesis reactions. For initial ATP measurements, CellTiter-Glo (Promega, Madison, WI), which is used for ATP quantification and prevents additional ATP synthesis, was immediately added to one tube for each sample after substrate additions. For synthesis reactions, the second tube for each sample was incubated at 20 °C for 10 min prior to the addition of CellTiter-Glo. All samples were incubated with CellTiter-Glo at room temperature in the dark for 10 min prior to reading luminescence with a Fluoroskan Ascent® FL (Thermo Fisher Scientific, Waltham, MA). Sample luminescence was

compared to an ATP standard curve, and ATP synthesis rate was calculated by subtracting initial ATP concentrations from final ATP concentrations. Protein content in each sample was measured with NanoOrange™ Protein Quantification Kits according to the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA), and was used for ATP synthesis rate normalization.

Genomic sequencing and allele frequency determination

Two developmental groups of F_2 hybrid copepodids for each reciprocal cross were allowed to develop to adulthood: those that metamorphosed 8-12 dph ("fast developers") and those that metamorphosed >22 dph ("slow developers"). For each group, 180 adults (approximately equal numbers of females and males) were pooled for DNA isolation by phenol-chloroform extraction (5). Briefly, copepods were rinsed with deionized water and homogenized by hand in 150 µL of Bender buffer (200 mM sucrose, 100 mM NaCl, 100 mM Tris-HCl pH 9.1, 50 mM EDTA, 0.5% SDS) in 1.5 mL microcentrifuge tubes. An additional 250 µL of Bender buffer was added to each sample followed by 100 µg of Proteinase K (Thermo Fisher Scientific, Waltham, MA). Samples were incubated at 56 °C overnight then cooled to room temperature for ~15 min. 25 µg of RNase A (Thermo Fisher Scientific, Waltham, MA) was added to each sample prior to a 37 °C incubation for 30 min, which was followed by addition of 200 µL of 5 M potassium acetate and a 10 min incubation on ice. Samples were then centrifuged at 13,000 g for 10 min at 4 °C, supernatants were transferred to 2.0 mL microcentrifuge tubes, and 400 µL of UltraPure™ Buffer-Saturated Phenol (Thermo Fisher Scientific, Waltham, MA) and 400 µL of OmniPur® Chloroform (EMD Millipore Corporation, Darmstadt, Germany) were added to each supernatant. The phenol-chloroform mixtures

4

were then gently mixed for 1 min, and centrifuged at 20,000 g for 5 min at 4 °C. Aqueous phases were transferred to new 2.0 mL tubes, and organic phases were back-extracted as above to maximize DNA yield. 400 µL of chloroform was again added to each aqueous phase for re-extraction: samples were centrifuged at 15,000 g for 1 min at 4 °C, aqueous phases were transferred to new 2.0 mL tubes, and again organic phases were back-extracted repeating the above procedure. 1,200 µL of ice-cold 95% ethanol was added to each aqueous phase; tubes were incubated at -20 °C for 1 h to facilitate DNA precipitation, and then centrifuged at 16,000 g for 20 min at 4 °C. 95% ethanol was removed by pipette, and all pellets for a sample (from back-extractions etc.) were combined in 1,000 µL of ice-cold 75% ethanol. Samples were then centrifuged at 16,000 g for 5 min at 4 °C. 75% ethanol was removed by pipette followed by an additional 16,000 g centrifugation for 1 min at 4 °C. Any remaining ethanol was removed and samples were dried in air for 20 min then resuspended in UltraPure™ Distilled Water (Thermo Fisher Scientific, Waltham, MA). DNA isolations were quantified with a Qubit® 2.0 Fluorometer and a dsDNA HS assay kit according to the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA).

Approximately 1 µg of genomic DNA for each pool was sent to Novogene Co., Ltd. (Sacramento, CA) for whole-genome 150 bp paired-end sequencing on a NovaSeq 6000 (Illumina Inc., San Diego, CA). Between 59,765,048 and 74,464,410 paired reads were obtained for each sample, which were trimmed to remove adapter sequences and base pairs with Phred scores less than 25. After trimming, reads with less than 50 bp remaining were removed. BWA MEM v0.7.12 (6) was used to align the filtered reads to the SD *T. californicus* reference genome v2.1 (7) and an updated SC reference genome,

which was prepared as described in Barreto et al. (7) and Lima et al. (8). Prior to read

mapping the references were equalized such that any "N" position in one reference was

also an "N" in the other reference. Mapping hybrid sample reads to both parental

references allows calculation of average allele frequency estimates between the

mappings, which accounts for mapping biases between matched and mismatched allelic

reads (9). Read mappings with MAPQ scores less than 20 were discarded, resulting in

average genome coverage values between 66X and 83X for all sample-to-reference

combinations (SI Appendix, Tables S1 and S2). 2,890,130 biallelic single nucleotide

polymorphisms (SNPs) that are fixed between the SD and SC populations were identified

using the previously published methods (8, 9) with population-specific sequencing reads

obtained from Barreto et al. (7). Briefly, population-specific reads were mapped to the

other population's reference genome, and variant loci with minor allele frequencies of 0

in both mappings were kept as fixed inter-population SNPs. Sample allele frequencies at

these SNPs were determined using PoPoolation2 (10) for all sites that had a minimum

coverage of at least 50 in the mappings to both parental reference genomes and a

minimum minor allele read count of 4 (as in 8). 1,910,010 of the fixed SNPs met these

filtering cutoffs. Estimated allele frequencies for each sample were averaged between the

two mappings to account for mapping biases, and mean allele frequencies were calculated

for non-overlapping 250 kb windows along each chromosome, which reduces noise in

allele frequency estimates as a single generation of recombination between SD and SC

chromosomes (which occurs only in males in *T. californicus* [2]) is not expected to break

apart large chromosomal blocks in F_2 hybrids (9).

Deviations in F_2 allele frequencies were detected as in Huang et al. (11). For each

comparison between the sequencing pools, Z-statistics and *P*-values were calculated for all SNPs with ≥80X coverage (42,502 SNPs) followed by false-discovery rate correction with the Benjamini-Hochberg method (12; $\alpha = 0.01$). Note this increase in coverage threshold improves the accuracy of Pool-seq estimates for the allele frequencies at individual loci (10). Allele frequency deviations between pools of F_2 hybrids are expected to be small (i.e., ±0.167; 8, 9), and therefore statistical tests using reads counts for individual genomic loci have low power to detect even strong deviations unless the allele frequencies deviate from 0.5 in opposite directions between the pools (e.g., 9). As a result, we also performed secondary analyses using Kolmogorov-Smirnov (KS) tests with the averaged allele frequencies for 250 kb chromosomal windows (as in 8, 9). To minimize the influence of linked loci (i.e., non-independent estimates), we used only a subset of windows per chromosome that were separated by 2 Mb (7-9 windows per chromosome). Correction for multiple tests was performed by Bonferroni correction of $\alpha = 0.05$. Although these secondary analyses improve the power to detect potential variation in allele frequencies due to mitonuclear interactions in our study, this approach is unlikely to completely account for linkage between loci across the chromosomes. Thus, the false-positive rate of our KS tests may be higher than expected, and these results should be interpreted both with some caution and in combination with our individual-SNP-based tests. In all cases, significant excesses of maternal alleles in fast developers are consistent with selection for coevolved (i.e., compatible) mitonuclear genotypes.

Supplemental references

1. R. S. Burton, Mating system of the intertidal copepod *Tigriopus californicus*. *Mar. Biol.* **86**, 247-252 (1985).

2. R. S. Burton, M. W. Feldman, S. G. Swisher, Linkage relationships among five enzyme-coding gene loci in the copepod *Tigriopus californicus*: a genetic confirmation of achiasmatic meiosis. *Biochem. Genet.* **19**, 1237-1245 (1981).

3. A. E. Harada, T. M. Healy, R. S. Burton, Variation in thermal tolerance and its relationship to mitochondrial function across populations of *Tigriopus californicus*. *Front. Physiol.* **10**, 213 (2019).

4. S. Tsuboko-Ishii, R. S. Burton, Individual culturing of *Tigriopus* copepods and quantitative analysis of their mate-guarding behavior. *J. Vis. Exp.* **139**, e58378 (2018).

5. J. Sambrook, D. W. Russell, Purification of nucleic acids by extraction with phenol:chloroform. *Cold Spring Harb. Protoc.* **2006**, pdb-rot4455 (2006).

6. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **25**, 1754-1760 (2009).

7. F. S. Barreto *et al.*, Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. *Nat. Ecol. Evol.* **2**, 1250-1257 (2018).

8. T. G. Lima, R. S. Burton, C. S. Willett, Genomic scans reveal multiple mito-nuclear incompatibilies in population crosses of the copepod *Tigriopus californicus*. *Evolution.* **73**, 609-620 (2019).

9. T. G. Lima, C. S. Willett, Using Pool-seq to search for genomic regions affected by hybrid inviability in the copepod *T. californicus*. *J. Hered.* **109**, 469-476 (2018).

10. R. Kofler, R. V. Pandey, C. Schlötterer, PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics.* **27**, 3435-3436 (2011).

11. W. Huang *et al.*, Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc. Natl. Acad. Sci. USA.* **109**, 15553-15559 (2012).

12. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B.* **57**, 289-300 (1995).
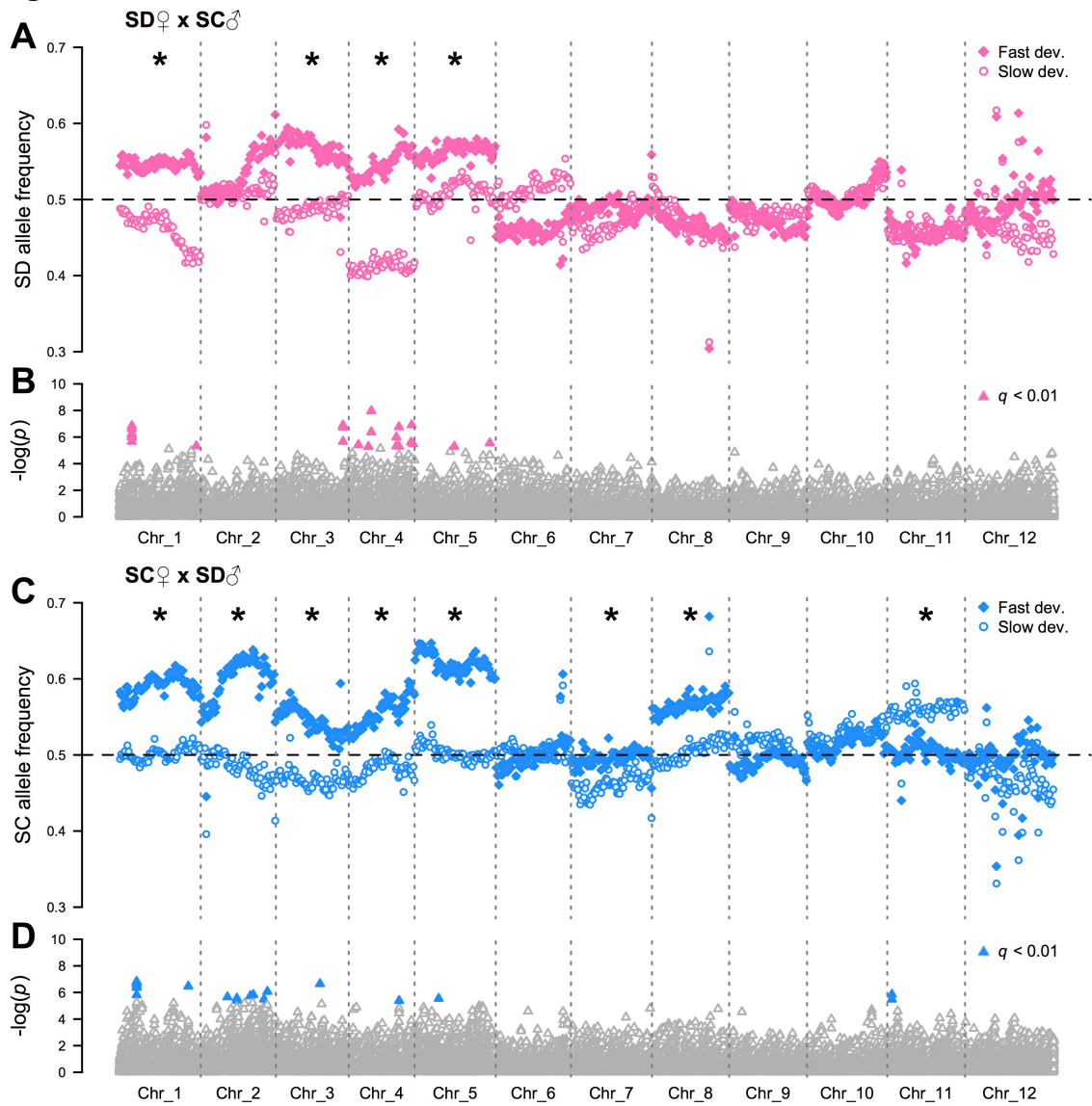
Fig. S1. Maternal allele frequencies for 759 chromosomal windows in SD♀xSC♂ (A; pink) and SC♀xSD♂ (C; light blue) fast (filled diamonds) and slow (open circles) developing F_2 hybrids. Asterisks indicate significant Kolmogorov-Smirnov tests for a chromosome between fast and slow developers. Panels B, D display statistical results for individual loci with ≥80X coverage (significant differences - filled pink or light blue triangles).

Fig. S2
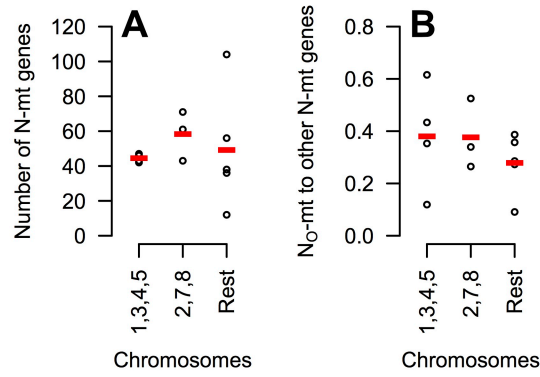
Fig. S2. Numbers of nuclear-encoded mitochondrial (N-mt) genes (A) and ratios of putative interacting (N_O-mt genes) to other N-mt genes (B) for all twelve *T. californicus* chromosomes. Chromosomes are grouped by those that were consistent with effects of mitonuclear matching on development rate detected in both the SD♀xSC♂ and SC♀xSD♂ crosses (1, 3, 4 and 5), in only the SC♀xSD♂ cross (2, 7 and 8), or in neither of the crosses (6, 9, 10, 11 and 12): individual chromosome values - black circles; mean values - red dashes.

Table S1. Sequencing and allele frequency summary for fast (8-12 dph) developers.

Chromosome	Number of 250kb windows	Number of SNPs	SNPs per window[1]	KS test P-value	Cross	Average SNP coverage	SC allele frequency for 250 kb windows		
							μ	$q_{0.1}$	$q_{0.9}$
One	66	182,789	2770 ± 424	4.1 x 10^{-5} *	SD♀xSC♂	66X	0.453	0.445	0.464
					SC♀xSD♂	69X	0.591	0.575	0.606
Two	61	150,451	2466 ± 660	1.6 x 10^{-4} *	SD♀xSC♂	67X	0.467	0.429	0.496
					SC♀xSD♂	70X	0.598	0.556	0.627
Three	59	157,563	2671 ± 740	1.6 x 10^{-4} *	SD♀xSC♂	66X	0.433	0.416	0.449
					SC♀xSD♂	69X	0.545	0.523	0.563
Four	54	145,651	2697 ± 743	5.8 x 10^{-4} *	SD♀xSC♂	66X	0.454	0.430	0.476
					SC♀xSD♂	69X	0.555	0.530	0.582
Five	66	177,977	2697 ± 485	4.1 x 10^{-5} *	SD♀xSC♂	66X	0.437	0.424	0.449
					SC♀xSD♂	69X	0.620	0.604	0.639
Six	61	151,542	2484 ± 803	1.6 x 10^{-4} *	SD♀xSC♂	66X	0.541	0.528	0.552
					SC♀xSD♂	70X	0.501	0.480	0.516
Seven	66	169,217	2564 ± 628	0.13	SD♀xSC♂	66X	0.511	0.500	0.526
					SC♀xSD♂	69X	0.496	0.485	0.508
Eight	63	164,275	2608 ± 583	2.5 x 10^{-3}	SD♀xSC♂	66X	0.536	0.521	0.549
					SC♀xSD♂	69X	0.567	0.553	0.581
Nine	63	161,669	2566 ± 558	1.6 x 10^{-4} *	SD♀xSC♂	66X	0.527	0.509	0.546
					SC♀xSD♂	69X	0.491	0.477	0.503
Ten	65	172,457	2653 ± 541	0.09	SD♀xSC♂	66X	0.494	0.465	0.512
					SC♀xSD♂	68X	0.516	0.500	0.529
Eleven	63	157,055	2493 ± 634	2.5 x 10^{-3}	SD♀xSC♂	66X	0.542	0.530	0.554
					SC♀xSD♂	70X	0.504	0.492	0.519
Twelve	72	119,344	1658 ± 914	0.73	SD♀xSC♂	68X	0.504	0.474	0.534
					SC♀xSD♂	73X	0.494	0.472	0.516

[1] μ ± σ; * significant after Bonferroni correction

Table S2. Sequencing and allele frequency summary for slow (>22 dph) developers.

Chromosome	Number of 250kb windows	Number of SNPs	SNPs per window[1]	KS test P-value	Cross	Average SNP coverage	SC allele frequency for 250 kb windows		
							μ	$q_{0.1}$	$q_{0.9}$
One	66	182,789	2770 ± 424	7.4×10^{-4} *	SD♀xSC♂	76X	0.539	0.517	0.574
					SC♀xSD♂	79X	0.502	0.491	0.513
Two	61	150,451	2466 ± 660	0.98	SD♀xSC♂	77X	0.488	0.480	0.499
					SC♀xSD♂	81X	0.482	0.460	0.503
Three	59	157,563	2671 ± 740	1.6×10^{-4} *	SD♀xSC♂	77X	0.514	0.500	0.525
					SC♀xSD♂	80X	0.469	0.455	0.480
Four	54	145,651	2697 ± 743	5.8×10^{-4} *	SD♀xSC♂	76X	0.587	0.576	0.597
					SC♀xSD♂	80X	0.481	0.463	0.496
Five	66	177,977	2697 ± 485	0.73	SD♀xSC♂	77X	0.492	0.475	0.509
					SC♀xSD♂	80X	0.503	0.493	0.516
Six	61	151,542	2484 ± 803	0.02	SD♀xSC♂	77X	0.488	0.467	0.508
					SC♀xSD♂	81X	0.504	0.492	0.515
Seven	66	169,217	2564 ± 628	7.4×10^{-4} *	SD♀xSC♂	77X	0.528	0.505	0.550
					SC♀xSD♂	80X	0.461	0.443	0.477
Eight	63	164,275	2608 ± 583	0.09	SD♀xSC♂	77X	0.530	0.500	0.552
					SC♀xSD♂	80X	0.506	0.488	0.523
Nine	63	161,669	2566 ± 558	0.66	SD♀xSC♂	77X	0.521	0.511	0.532
					SC♀xSD♂	80X	0.513	0.495	0.525
Ten	65	172,457	2653 ± 541	2.5×10^{-3}	SD♀xSC♂	76X	0.494	0.476	0.507
					SC♀xSD♂	79X	0.527	0.507	0.543
Eleven	63	157,055	2493 ± 634	0.02	SD♀xSC♂	77X	0.541	0.528	0.553
					SC♀xSD♂	81X	0.559	0.548	0.569
Twelve	72	119,344	1658 ± 914	6.3×10^{-3}	SD♀xSC♂	80X	0.533	0.498	0.563
					SC♀xSD♂	83X	0.462	0.436	0.488

[1] $\mu \pm \sigma$; * significant after Bonferroni correction