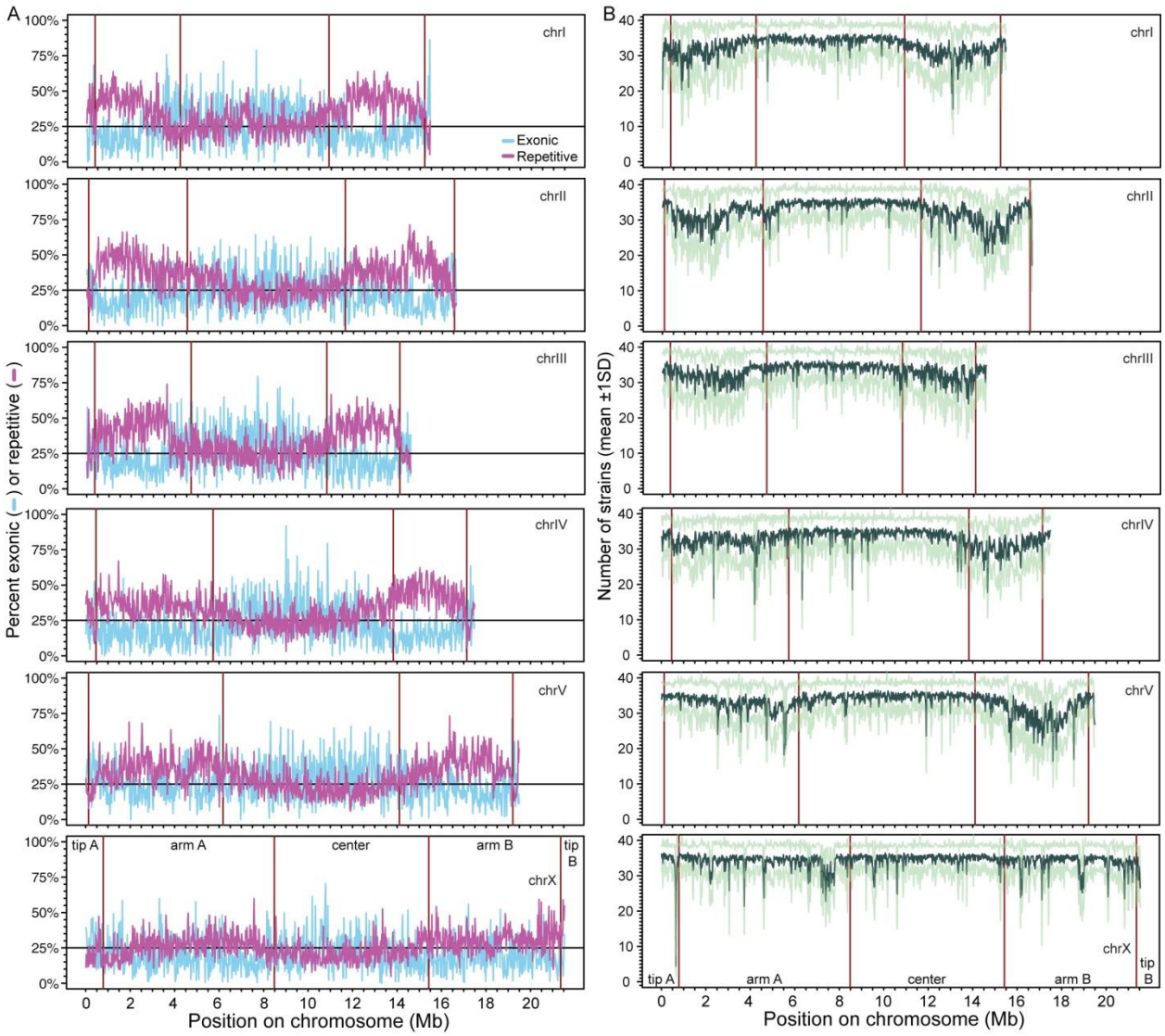
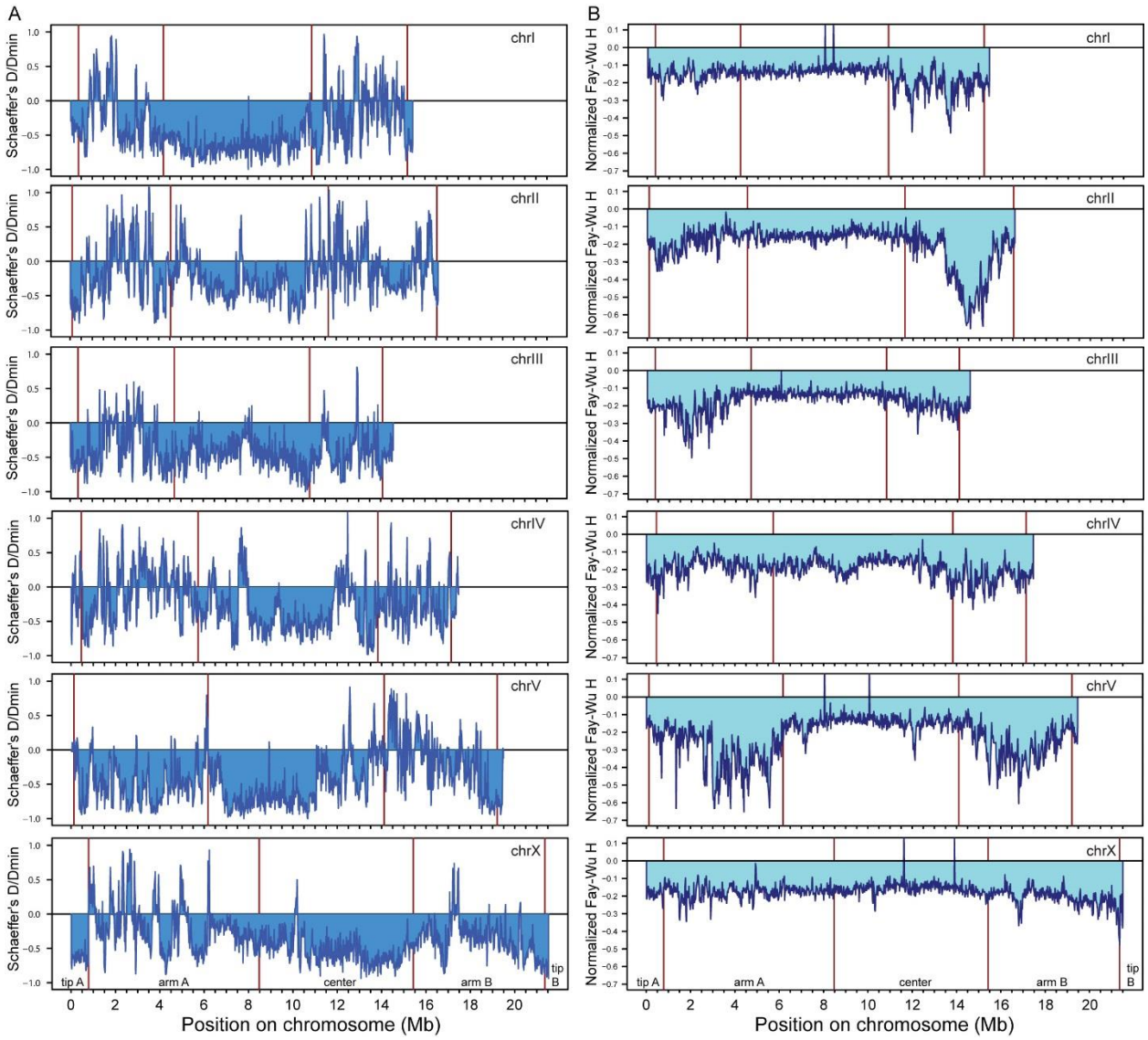
Supplementary Material for Thomas et al. 2015. *Genome Research*. "Full-genome evolutionary histories of selfing, splitting and selection in *Caenorhabditis*"

Supplementary Table 1.

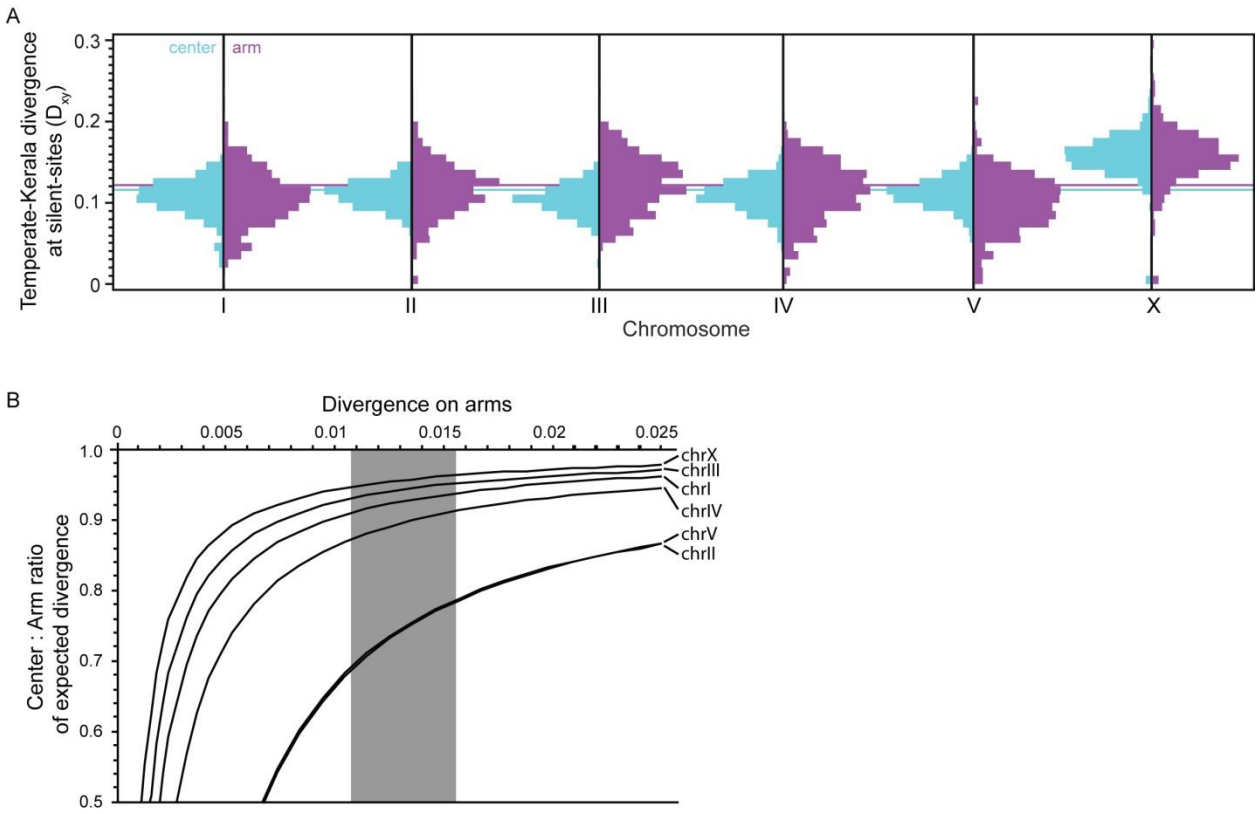
Strain	Phylogeographic clade	Mapped read coverage	Sequencing Platform	Library
ED3032	Tropical	33.75	GA	Paired End (100bp×2)
ED3034	Tropical	27.97	HiSeq2000	Paired End (100bp×2)
ED3036	Tropical	31.41	GA	Paired End (100bp×2)
GXW0023	Tropical	31.31	HiSeq2000	Paired End (100bp×2)
GXW0024	Tropical	268.27	MiSeq	Paired End (245bp+200bp)
JU1338	Tropical	102.21	MiSeq	Paired End (245bp+200bp)
JU1376	Tropical	28.77	HiSeq2000	Paired End (100bp×2)
JU1392	Tropical	25.91	HiSeq2000	Paired End (100bp×2)
JU1399	Tropical	70.5	HiSeq2000	Paired End (142bp+100bp)
JU1424	Tropical	23.34	HiSeq2000	Paired End (100bp×2)
JU1794	Tropical	45.37	GA + HiSeq2000	Paired End (100bp×2) + Single End (101bp)
JU1795	Tropical	19.78	HiSeq2000	Paired End (100bp×2)
JU1799	Tropical	31.06	HiSeq2000	Paired End (100bp×2)
JU1802	Tropical	32.08	HiSeq2000	Paired End (100bp×2)
JU1829	Tropical	79.49	HiSeq2000	Paired End (142bp+100bp)
JU1884	Tropical	16.72	GA	Paired End (100bp×2)
JU1907	Tropical	27.76	GA + HiSeq2000	Paired End (100bp×2) + Single End (67bp)
JU1908	Tropical	72.01	HiSeq2000	Paired End (142bp+100bp)
JU1914	Tropical	25.72	HiSeq2000	Paired End (100bpX×2)
JU726	Tropical	26.88	HiSeq2000	Paired End (100bp×2)
NIC109	Tropical	38.74	HiSeq2000	Paired End (100bp×2)
NIC115	Tropical	116.67	HiSeq2000	Paired End (142bp+100bp)
NIC65	Tropical	102.38	HiSeq2000	Paired End (142bp+100bp)
QX1410	Tropical	96.04	HiSeq2000	Paired End (142bp+100bp)
QX1805	Tropical	116.29	MiSeq	Paired End (245bp+200bp)
BW287	Temperate	24.28	GA	Paired End (100bp×2)
EG4181	Temperate	40.45	GA	Paired End (100bp×2)
JU516	Temperate	30.59	GA	Paired End (100bp×2)
QX1547	Temperate	28.59	HiSeq2000	Paired End (100bp×2)
JU1341	Kerala	57.51	GA + HiSeq2000	Paired End (100bp×2) + Single End (67bp)
JU1348	Kerala	34.01	HiSeq2000	Paired End (100bp×2)
ED3101	Nairobi	30.81	GA	Paired End (100bp×2)
VX0034	Hubei	60.17	GA	Paired End (100bp×2)
QR24	Quebec	15.42	GA	Paired End (100bp×2)
QR25	Quebec	21.92	GA	Paired End (100bp×2)
NIC19	Taiwan	31.95	HiSeq2000	Paired End (100bp×2)
NIC20	Taiwan	33.72	HiSeq2000	Paired End (100bp×2)



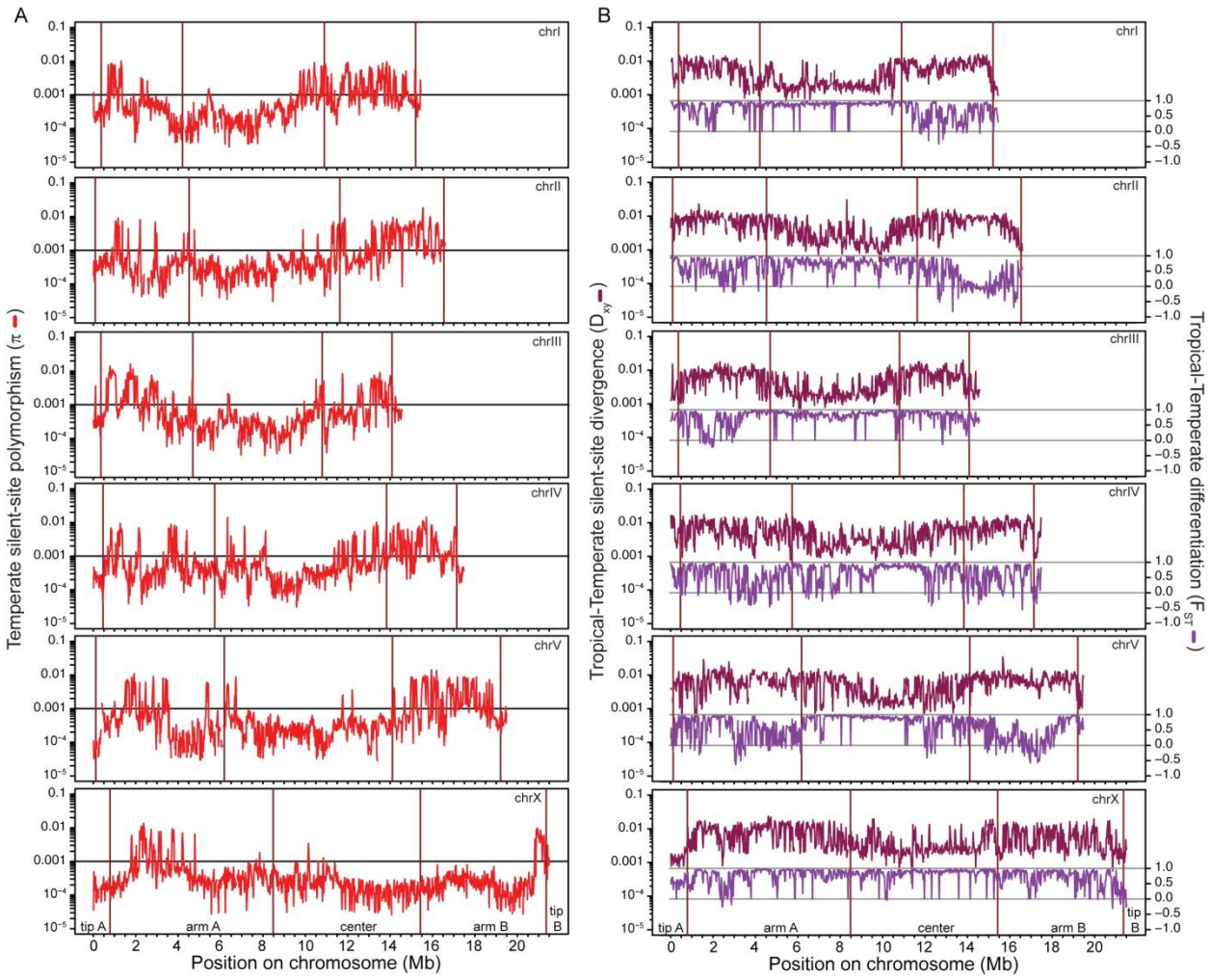
Supplementary Figure 1. Central chromosome domains are dense in coding sequence (blue) and sparse in repetitive DNA (pink), compared to arm domains of chromosomes in *C. briggsae* (A). Horizontal line at 25% is provided as a scale for comparing fluctuations along the length of a chromosome and among chromosomes. (B) The average number of strains with information passing filters in 20kb intervals, among the 37 wild isolates of *C. briggsae* sequenced mapped to reference genome version cb4 of strain AF16, varies modestly along chromosomes with repeat-rich arms having reduced strain coverage. Some regions of reduced strain coverage might represent copy number variation. Plots show values in non-overlapping 20kb intervals. Chromosome domain boundaries from (Ross et al. 2011).



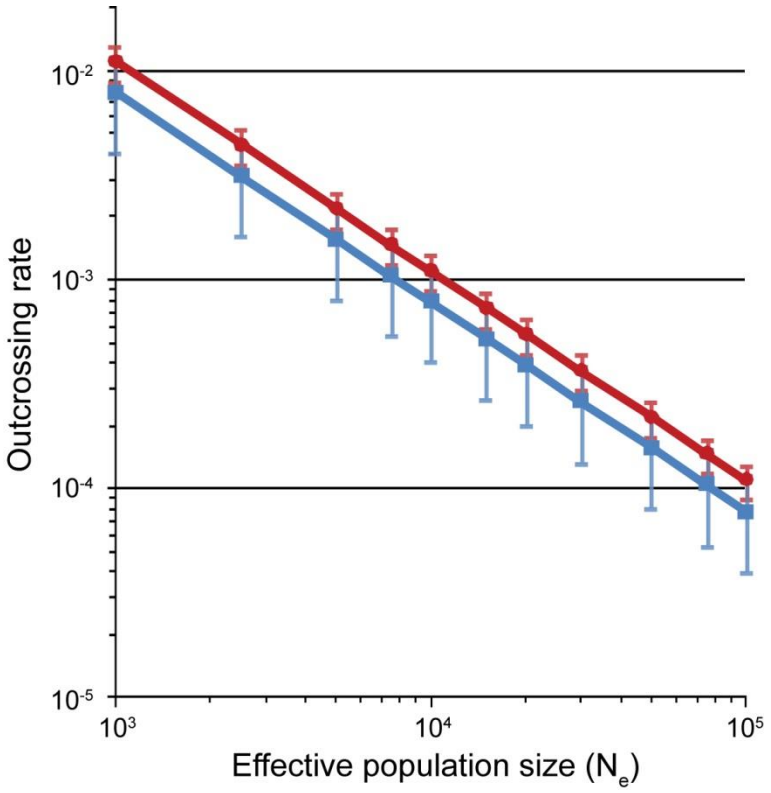
Supplementary Figure 2. The skew in the site frequency spectrum for silent sites in non-overlapping 20kb windows varies along chromosomes. (A) Shaeffer's D/D_{min} (Schaeffer 2002) for Tropical strains normalizes Tajima's D between [-1, +1] by scaling with the theoretical minimum value of D, given the sample of polymorphic sites. (B) Fay and Wu's H (Fay and Wu 2000), normalized by its expected variance (H_{norm}) (Zeng et al. 2006), describes the skew in high-frequency derived polymorphisms in the Tropical sample using Kerala strains as outgroup for ancestral state inference. The genome shows weakly negative values of H_{norm} with values that appear to deviate from zero more often on chromosome arms. However, polymorphism within the Tropical sample and divergence with Kerala are similar in corresponding regions of the arms (Figure 1, main text), suggesting that errors in ancestral state inference lead to spuriously low values of H_{norm} in these arm regions rather than reflecting recent selective sweeps.



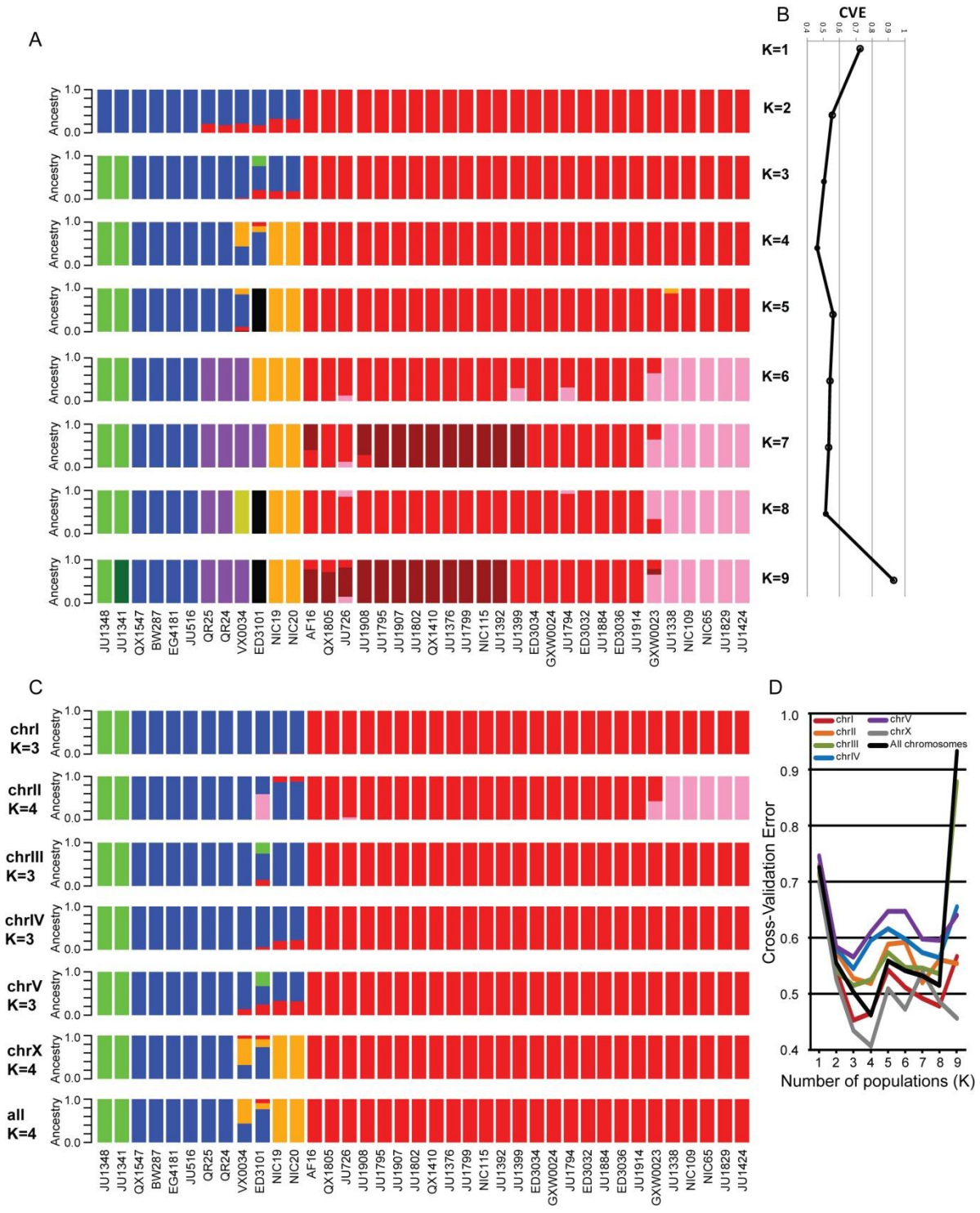
Supplementary Figure 3. (A) The distribution of divergence between Tropical and Kerala strains (D_{xy}) at silent sites differs only slightly for chromosome center and arm domains, being 7.6% lower in centers on average. Horizontal lines indicate the global median D_{xy} for center (blue) and arm (purple) domains. (B) Given the median observed π_{sil} for center and arm domains, the expected divergence between domains owing to ancestral polymorphism is ~10% lower for center regions. Gray shaded region indicates the range of median observed D_{xy} of arm domains among chromosomes. Chromosome-specific curves calculated according to the solution to f as a function of D_{xy_arm} for $D_{xy\_arm} - D_{xy\_center} = \pi_{arm} - \pi_{center}$, where $D_{xy\_center} = f \cdot D_{xy\_arm}$.



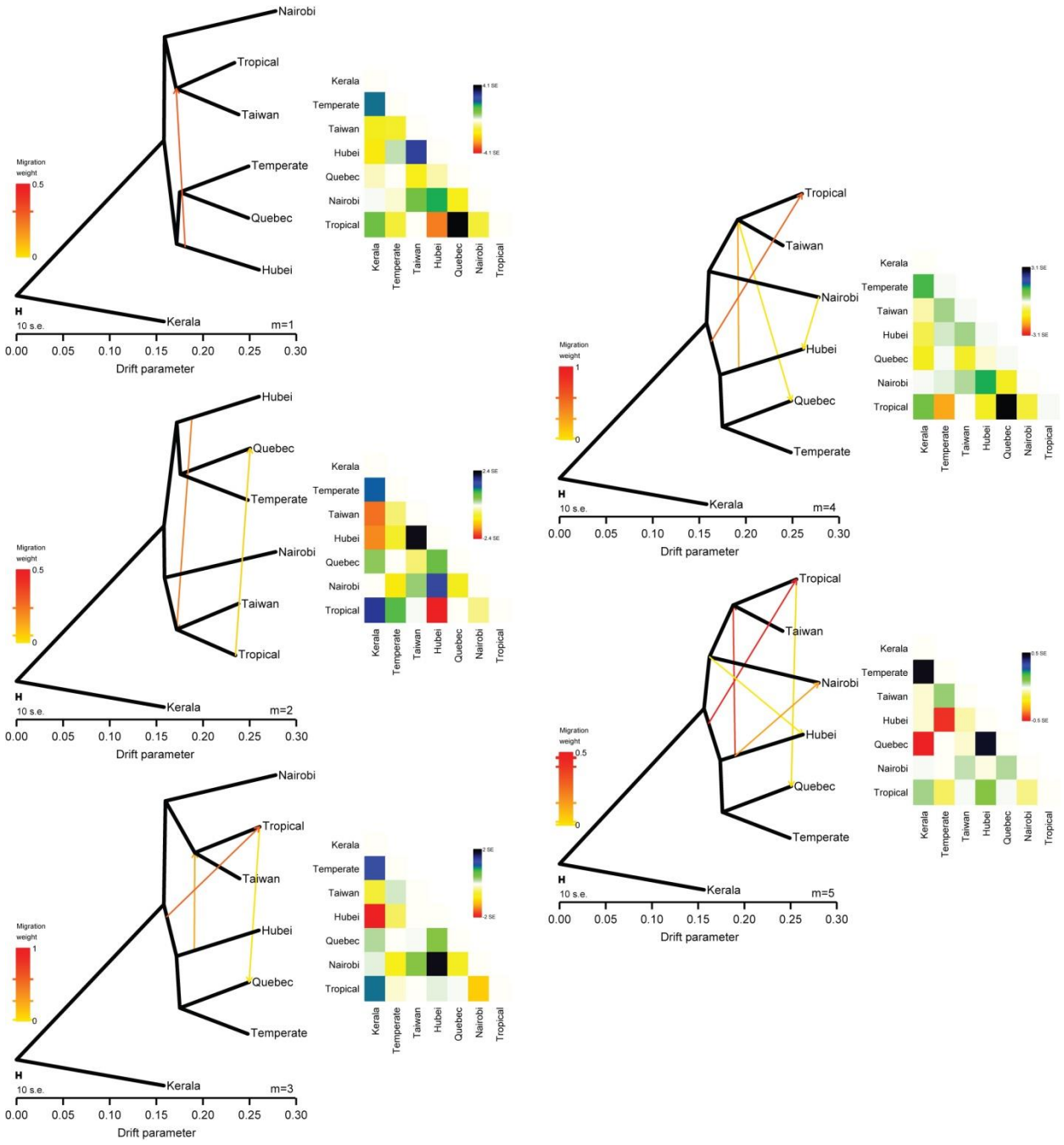
Supplementary Figure 4. (A) Silent-site nucleotide polymorphism among the 4 strains in the Temperate phylogeographic group varies along chromosomes, with center domains tending to have reduced variation than arm domains. (B) Genetic differentiation between Temperate and Tropical populations varies along chromosomes, and depends on whether absolute (D_{xy}) or relative (F_{ST}) measures of differentiation are applied. D_{xy} (dark purple, left axis) tends to be higher in chromosome arms than centers (20kb windows D_{xy} Wilcoxon test $P<0.0001$ for all chromosomes), reflecting deeper coalescence of arm loci owing to more ancestral polymorphism in those regions of the genome. By contrast, F_{ST} (light purple, right axis) tends to be higher in chromosome centers than arms (20kb windows F_{ST} Wilcoxon test $P<0.0001$ for all chromosomes), reflecting a greater proportion of fixed differences with the shallower coalescent depth of lower polymorphism in central regions. Both of these observations are consistent with stronger linked selection in the gene dense, low recombination center regions of chromosomes (Pease and Hahn 2013; Cruickshank and Hahn 2014). Plots show values in non-overlapping 20kb intervals.



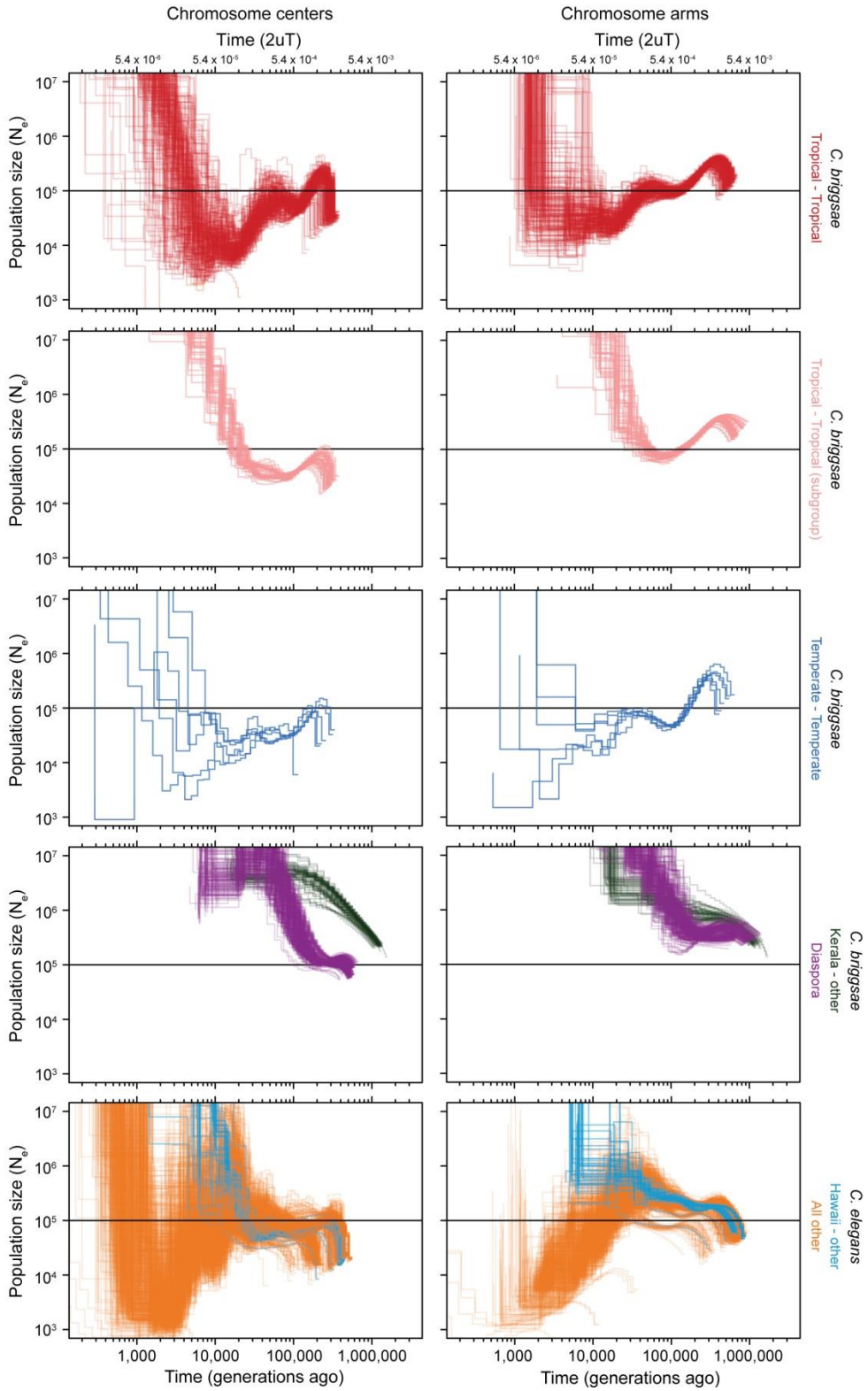
Supplementary Figure 5. Outcrossing rates estimated from inter-chromosomal linkage disequilibrium (r^2)for *C. briggsae* (red) and *C. elegans* (blue) as a function of the presumed effective population size (N_e). Error bars correspond to SEM around the mean of the 15 pairwise values of inter-chromosomal r^2 between the 6 chromosomes of each species, where a value for a given combination of chromosomes is the average r^2 between non-singleton polymorphic sites on the two chromosomes. Computation of outcrossing rate used the equation in (Cutter 2006), except that r^2 was substituted with $r^2 - r^2_{exp}$ where r^2_{exp} is the background value given the sample size (Weir and Hill 1980).



Supplementary Figure 6. (A) ADMIXTURE ancestry results for K=2 to 9 using data from all chromosomes together with (B) cross-validation error (CVE) indicated for each value of K. (C) ADMIXTURE ancestry proportions for each chromosome run separately, shown for the value of K that minimizes the CVE. (D) CVE values for each chromosome for K=2 to 9.

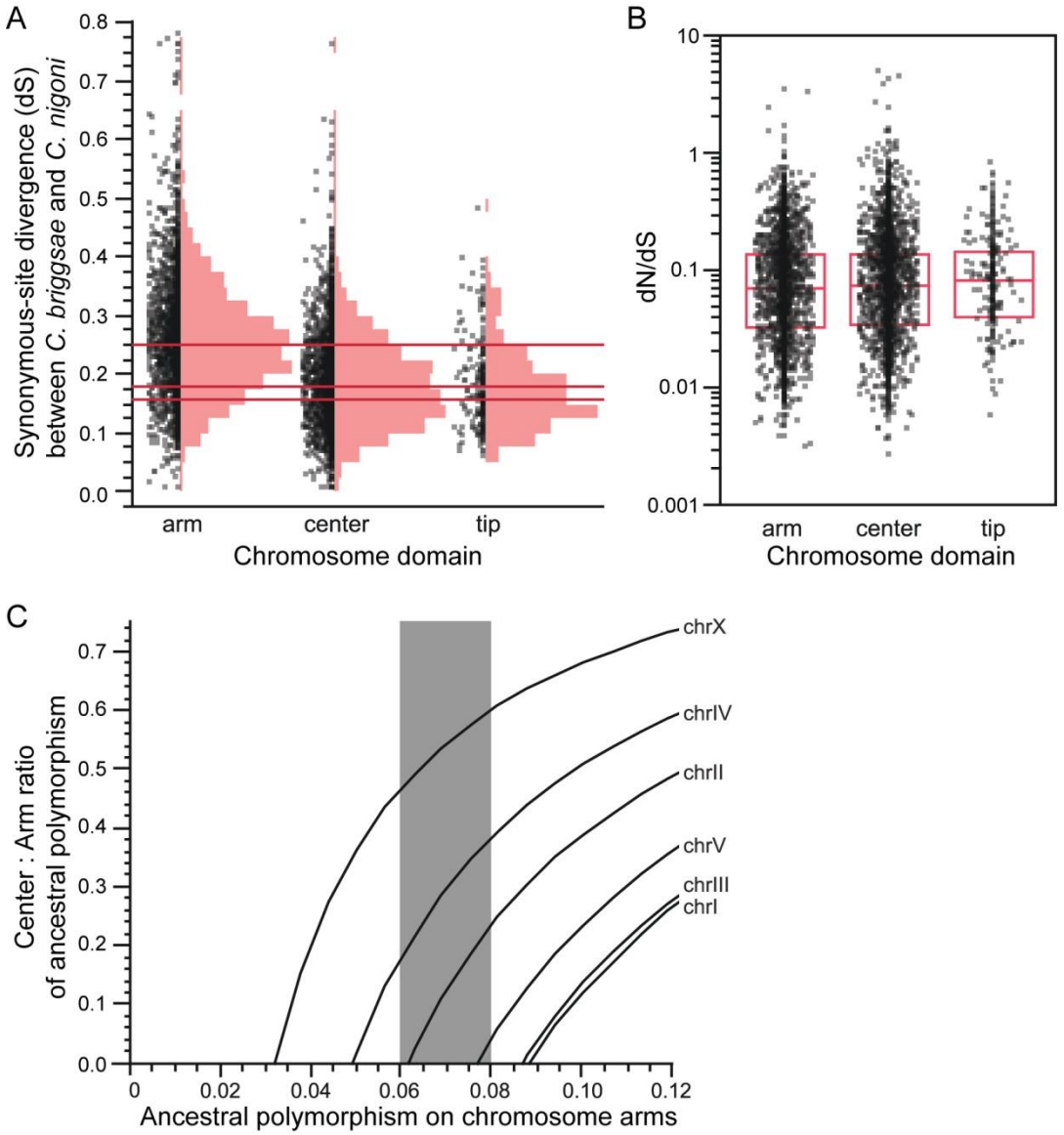


Supplementary Figure 7. Alternative Treemix analyses showing inferred migration events and their weights in the genealogy, using m=0 to 5 events. Inset matrix to the right of each genealogy indicates the residual variation unexplained by the topology.



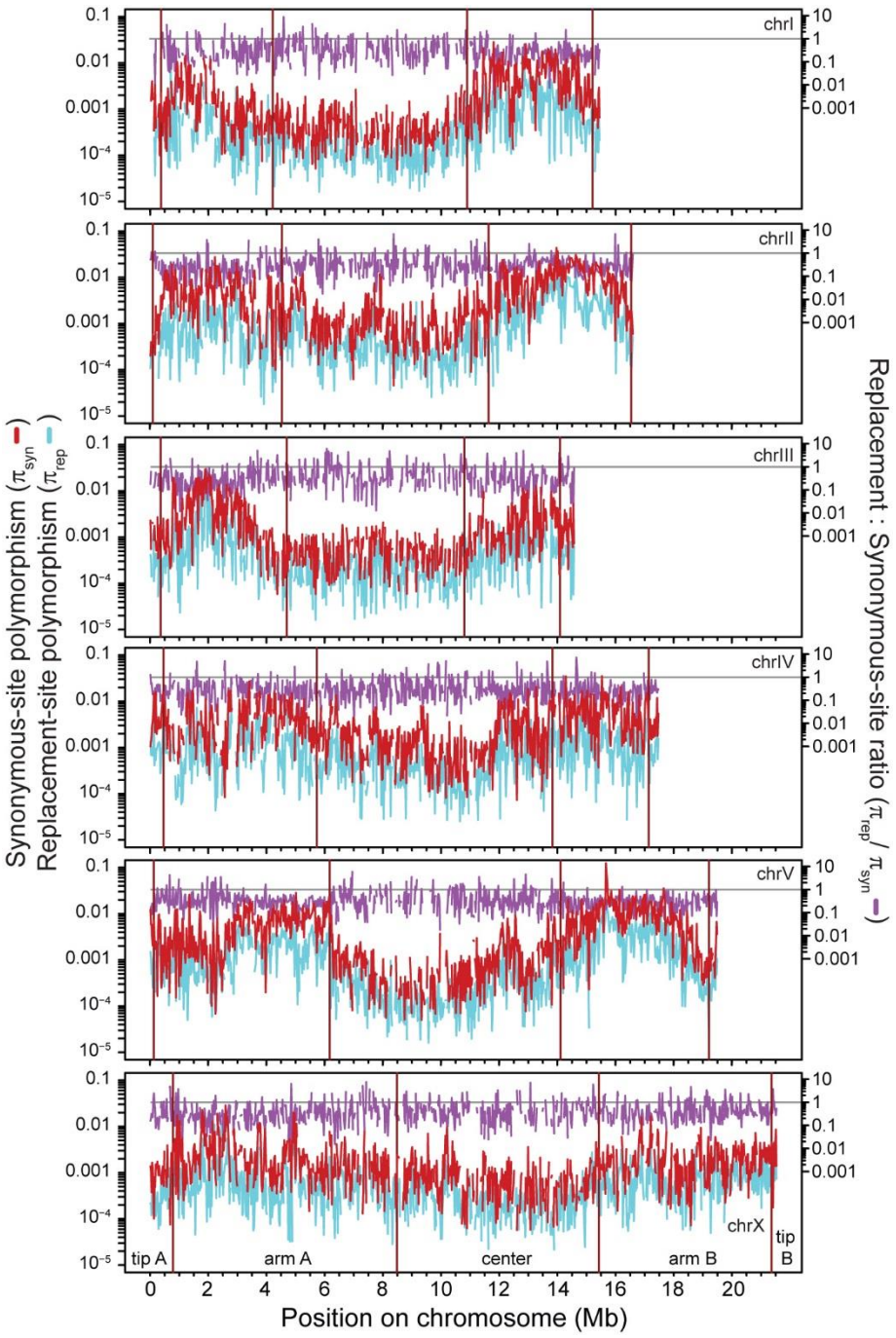
Supplementary Figure 8. PSMC profiles of effective population size (N_e) through time for different strain subsets of *C. briggsae* (top four rows) and *C. elegans*, performed separately for center (left column) and

arm domains (right column) of chromosomes. Top two rows show analysis of pseudo-diploid genome combinations of strains from the Tropical phylogeographic group, with the second row (pink) containing the subgroup of strains that we infer to have a recent population split from the main set of Tropical strains (red, top row). The middle row includes strain combinations from the Temperate phylogeographic group (dark blue). The fourth row includes pseudo-diploid genome combinations of strain from different genetic groups, with those including Kerala-derived genomes (green) separated from other combinations representing the diaspora splitting of populations in the past (purple). Bottom row shows *C. elegans* pseudo-diploid genome combinations, with those including Hawaiian strain CB4856 (light blue) separated from all others (orange). Each curve represents the PSMC output for a pseudo-diploid genome combination.



Supplementary Figure 9. (A) Divergence at synonymous sites between gene orthologs of *C. briggsae* and *C. nigoni*, separated by chromosomal domain. Red horizontal lines indicate median values for arm (top), center (middle), and tip (bottom) domains. Loci with $d_S > 0.8$ or ENC < 45 are excluded. (B) Rates of protein sequence divergence (d_N/d_S; substitution rate at nonsynonymous sites d_N, normalized by the substitution rate at synonymous sites, d_S) do not differ among chromosome domains. Chromosome domains defined according to gene positions in the *C. briggsae* genome according to (Ross et al. 2011). Plot excludes 348 loci with $d_N/d_S = 0$, although these loci are included in calculation of median and interquartile range (red boxes). (C) Ratio of ancestral polymorphism in chromosome centers relative to arms as a function of ancestral polymorphism on chromosome arms, for each chromosome. Given the median observed d_S for center and arm domains, the expected polymorphism is >60% lower for center regions than arms. Gray shaded region indicates the range of π_{arm} expected in the common ancestor of

C. briggsae and *C. nigoni*, based on unpublished synonymous-site polymorphism of several genes in *C. nigoni* and published data for related outcrossing species (Cutter et al. 2013). Chromosome-specific curves calculated according to the solution to f as a function of π_{arm} for $d_{S\text{-}arm} - d_{S\text{-}center} = \pi_{arm} - \pi_{center}$, where $\pi_{center} = f \cdot \pi_{arm}$.



Supplementary Figure 10. The density of polymorphisms at replacement sites (π_{rep}, blue) is lower than at synonymous sites (π_{syn}, red). The amount of polymorphism for both replacement and synonymous sites is reduced in central regions of chromosomes. However, the relative efficacy of selection in eliminating missense mutations is significantly weaker in chromosome center regions, indicated by the slightly elevated π_{rep} /π_{syn} ratio in center chromosome domains (π_{rep} / π_{syn}, purple; Figure 6, main text).

References for Supplementary Material

Cruickshank, T.E. and Hahn, M.W. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23:** 3133-3157.

Cutter, A.D. 2006. Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. *Genetics* **172:** 171-184.

Cutter, A.D., Jovelin, R., and Dey, A. 2013. Molecular hyperdiversity and evolution in very large populations. *Mol. Ecol.* **22:** 2074-2095.

Fay, J.C. and Wu, C.I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155:** 1405-1413.

Pease, J.B. and Hahn, M.W. 2013. More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution* **67:** 2376-2384.

Ross, J.A., Koboldt, D.C., Staisch, J.E., Chamberlin, H.M., Gupta, B.P., Miller, R.D., Baird, S.E., and Haag, E.S. 2011. *Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. *PLoS genetics* **7**.

Schaeffer, S.W. 2002. Molecular population genetics of sequence length diversity in the *Adh* region of *Drosophila pseudoobscura*. *Genetics Research* **80:** 163-175.

Weir, B.S. and Hill, W.G. 1980. Effect of mating structure on variation in linkage disequilibrium. *Genetics* **95:** 477-488.

Zeng, K., Fu, Y.-X., Shi, S., and Wu, C.-I. 2006. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174:** 1431-1439.