Population structure and introgression among recently differentiated *Drosophila melanogaster* populations

J.M. Coughlan[1,%,*], A.J. Dagilis[1]*, A. Serrato-Capuchina[2], H. Elias[1], D. Peede[3,4], K. Isbell[1], D.M. Castillo[5], B.S. Cooper[6] and D.R. Matute[1,%]

*Denotes equal contribution

[1] Biology Department, University of North Carolina

[2] Arnold Arboretum, Harvard University

[3]Department of Ecology, Evolution, and Organismal Biology, Brown University

[4] Center for Computational Molecular Biology, Brown University

[5] School of Biological Sciences, University of Utah

[6] Division of Biological Sciences, University of Montana

%Biology Department

University of North Carolina

250 Bell Tower Drive, Genome Sciences Building

Chapel Hill, NC 27510

Email: dmatute@email.unc.edu, jcoug@email.unc.edu

Running title: *Diversity and genetic structure in a model system*

Keywords: Female mate choice, Genetic differentiation, Genetic incompatibility, Hybridization, Polygenic adaptation, Reproductive isolation

Author contributions

Experimental design: JMC, AJD, DMC, and DRM designed all experiments and data analyses.

Data collection: JMC, ASC, HE, DP, KI, BSC, and DRM collected the data, including field work, mate choice experiments, and DNA extractions.

Data analyses: JMC and AJD performed all data analyses.

Manuscript writing: JMC, AJD, DMC, BSC, and DRM wrote the manuscript

ABSTRACT

Understanding the factors that produce and maintain genetic variation is a central goal of evolutionary biology. Despite a century of genetic analysis, the evolutionary history underlying patterns of exceptional genetic and phenotypic variation in the model organism *Drosophila melanogaster* remains poorly understood. In particular, how genetic and phenotypic variation is partitioned across global *D. melanogaster* populations, and specifically in its putative ancestral range in Subtropical Africa, remains unresolved. Here, we integrate genomic and behavioral analyses to assess patterns of population genetic structure, admixture, mate preference, and genetic incompatibility throughout the range of this model organism. Our analysis includes 174 new accessions from novel and under-sampled regions within Subtropical Africa. We find that while almost all Out of Africa genomes correspond to a single genetic ancestry, different geographic regions within Africa contain multiple distinct ancestries, including the presence of substantial cryptic diversity within Subtropical Africa. We find evidence for significant admixture—and variation in admixture rates—between geographic regions within Africa, as well as between African and non-African lineages. By combining behavioral analysis with population genomics, we demonstrate that female mate choice is highly polymorphic, behavioral types are not monophyletic, and that genomic differences between behavioral types correspond to many regions across the genome. These include regions associated with neurological development, behavior, olfactory perception, and learning. Finally, we discovered that many individual pairs of putative incompatibility loci likely evolved during or after the expansion of *D. melanogaster* out of Africa. This work contributes to our understanding of the evolutionary history of a key model system, and provides insight into the distribution of reproductive barriers that are polymorphic within species.

INTRODUCTION

Drosophila melanogaster has remained one of the most powerful genetic systems to understand the molecular underpinnings of animal biology since its development in the early 20th century [1–3]. The species is distributed globally and is commonly associated with human settlements [4,5]. Recent sampling efforts have strongly suggested that *D. melanogaster* originated in the African mopane forest and initially bred on marula fruits, with a transition to human commensalism within Africa approximately 10,000-13,000 years ago [4,5]. Yet, much remains unknown about the natural history, distribution, and evolutionary history of *D. melanogaster*, particularly in its ancestral range. Given the importance of model systems, like *D. melanogaster*, to our understanding of the genetic basis of morphological [6,7], physiological [8], and behavioral [9] traits, as well as our understanding of different evolutionary processes in both natural and experimental contexts [10–12], it remains critical to understand how genetic and phenotypic variation evolved and is maintained in the ancestral range.

Although significant population genetic structure between Africa and non-African populations has long been recognized in *D. melanogaster [13],* and the existence of population genetic structure within Africa has more recently been suggested [14–16], the partition of genetic diversity within the ancestral range is largely still unresolved. Early multilocus surveys found limited to modest structure within Africa [14,17], and supported distinct West and East African clades [15]. Most recent efforts suggest modern day remnants of ancestral *D. melanogaster* lineages exist as isolated, genetically unique populations within the putative ancestral range [5,16,18]. Human-aided migration following the transition to human commensalism is thought to have

87 contributed to both the within Africa expansion [4,5,19], and subsequent global
88 expansion, the latter of which likely resulted from a single out of African event [5,20–
89 22], with multiple bottlenecks [23,24]. However, despite housing the vast majority of
90 genetic diversity, the demographic processes shaping population genetic structure in
91 the ancestral range remain largely unknown (though see [4,5]).

92 Post expansion, multiple historical events have also created opportunities for
93 human-mediated admixture between genetically distinct lineages of *D. melanogaster*. It
94 has previously been suggested that the opening of western commercial routes and the
95 'Scramble for Africa' facilitated hybridization between local African populations of *D.*
96 *melanogaster* and invading non-African *D. melanogaster* individuals, particularly in more
97 urban areas [25,26]. Indeed, the extent of non-African ancestry in Africa is widely
98 variable between populations [18], with some evidence for more pronounced signatures
99 of admixture in urban populations, [25,26]. Second, human migration associated with
100 slavery roughly 400 years ago produced a secondary contact zone between African and
101 non-African populations of *D. melanogaster* in the southeastern United States and the
102 Caribbean [27,28]. However, it is unknown whether cryptic genetic lineages within Africa
103 [16] have differentially contributed African ancestry outside of Africa [29]. Furthermore,
104 the extent to which admixture between Out of Africa and African lineages contributes to
105 the modern-day genetic composition of *D. melanogaster* within Africa is relatively
106 unresolved (but see [18,25,26]). Lastly, patterns of admixture between African
107 populations are almost entirely unexplored, despite their relevance for unraveling the
108 evolutionary history of *D. melanogaster*.

109 One mechanism that could contribute to differentiation among *D. melanogaster*

110 populations is behavioral variation between populations which results in assortative
111 mating. Behavioral surveys of female mate choice within *D. melanogaster* from
112 Subtropical Africa revealed a potential case of incipient speciation within *D.*
113 *melanogaster* ([13,30–32]. At least two more cases of region-specific female mate
114 preference have also been described [26,33]). Cosmopolitan (denoted 'M') flies are
115 globally distributed, including within Subtropical Africa, while a second lineage is largely
116 restricted to Zambia and Zimbabwe (denoted 'Z'). While Z females show significant
117 preference for Z males, M females show no preference, and males of both lineages
118 court both types of females at similar rates [30,32,34]. Early attempts to map the genetic
119 basis of female preference found the trait to be highly polygenic, involving loci on all
120 major chromosome arms, and significant non-additive effects [30,31]. Despite the
121 contributions of *D. melanogaster* to the field of speciation [35,36], it remains unknown
122 whether the frequencies of these behavioral types covary with patterns of *D.*
123 *melanogaster* population structure within Subtropical Africa. Understanding if/how
124 behavioral isolation is associated with genetic structuring can allow us to determine if
125 divergence in complex, quantitative traits—such as mate choice—can contribute to
126 stable genetic divergence within a species, a hypothesis which is still widely debated
127 [37–39].
128 In addition to variation in female mate choice, negative epistatic interactions
129 between polymorphic loci, similar to hybrid incompatibilities, can cause substantial
130 variation in fitness, including the production of low fitness individuals within a species
131 [40,41]. While it has been hypothesized that these incompatibilities are geographically
132 structured across the *D. melanogaster* range [40], very little is known about the

133 geographic origins or distributions of these alleles. In particular, it remains unknown if
134 these alleles correspond to previously described behavioral and/or genetic lineages (i.e.
135 [13,16,32]). Determining how these forms of reproductive isolation correspond with
136 population structure and patterns of gene flow can contribute to our understanding of
137 the early stages of population differentiation and speciation.

138 Here we address the extent of genetic differentiation and admixture in *D.*
139 *melanogaster*, relate patterns of genetic structure to behavioral differences in the
140 putative ancestral range, and describe the geographic distribution of putative
141 incompatibility alleles and their contributions to admixture and population structure. This
142 work leverages genome-wide information from 420 individuals, including 174 samples
143 from novel and previously undersampled geographic regions as well as extensive
144 behavioral assays of flies from Subtropical Africa. Specifically, we ask: does *D.*
145 *melanogaster* show population genetic structure in Africa, particularly within the
146 ancestral range? What is the extent of gene flow between unique genetic lineages? And
147 how do two forms of polymorphic RI (mate choice and genetic incompatibilities)
148 influence population structure in this system? Our results help to clarify the
149 demographic history of *D. melanogaster* and provide some insights into the persistence
150 of genetically unique clades within *D. melanogaster*.

151

152 **METHODS**

153 *Sampling*

154 We collected *D. melanogaster* from seven novel locations in Zambia, Namibia, and
155 Zimbabwe using a similar approach to previously described efforts ([5]; see Table S1 for

156 sampling locations). Our approach differed from other samplings in that we used
157 multiple potential substrates. Each trap consisted of buckets adjacent to each other and
158 separated by about 50cms. The buckets were filled with mashed bananas (purchased
159 locally), Marula (*Sclerocarya birrea*) or muzinzila fruits (*Berchemia discolor).* In all
160 cases, we removed the husk of the fruit, added yeast (Red Star Active Dry Yeast - 16
161 oz. #201265), and allowed them to ferment for about 24 hours. These traps were put
162 underneath trees. We collected all flies in the bucket using a sweeping net (BioQuip;
163 Rancho Domingo, CA) after 24, 48, and 72 hours. We then netted and aspirated flies
164 with a pooter (1135A Aspirator–BioQuip; Rancho Domingo, CA)) and anesthetized
165 within 20 minutes of collection using FlyNap (triethylamine, Carolina Biological Supply
166 Co.). Females and males were separated. Males and individuals from other species
167 were placed in ethanol; *D. melanogaster* females were placed in 30mL plastic vials with
168 cornmeal food and allowed to oviposit. Of 339 collected females, we were able to
169 establish 244 isofemale lines (i.e., shelf-stable lines derived from a single matriarchal
170 lineage).

171 *DNA extraction and sequencing*

172 We extracted DNA from 20 individuals from each of the 174 unique isofemale lines
173 using a Gentra Puregene Tissue Kit (Qiagen, Valencia, CA, USA) following the
174 recommended tissue protocol with volumes of reagents as suggested for processing 5 -
175 10 mg of tissue (see Table S1 for collection details). To prepare the genomic DNA
176 libraries we used KAPA HyperPrep kits (Roche Sequencing, Pleasanton, CA) with a
177 target fragment size of 300-500 bp at the University of North Carolina (UNC) School of
178 Medicine's high-throughput sequencing facility. Next, we pooled individually barcoded

179 libraries into groups of ~10 individual libraries and each pool was sequenced on either a
180 single lane of an Illumina HiSeq 4000 or a Noveseq6000S4XP platform, in both cases
181 generating paired-end 150 bp reads. This sequencing strategy yielded between 2.6-
182 23.1 billion bp of raw sequence data for each individual (See Table S1 for coverage
183 information). Additionally, we sequenced 32 lines that were advanced generation
184 isofemale lines collected in Malawi, Zimbabwe and Botswana (outlined in Table S1).
185

186 *Public data*

187 We obtained whole genome sequences for an additional 247 isolines via NCBI SRA. Of
188 these publicly available genomes, 72 were of flies sampled outside of Africa and 175
189 were of flies collected from within Africa, 30 of which resided in the preported ancestral
190 range in Subtropical Africa [16,18] (see Table S1 for details).

191 *Variant calling*

192 We aligned 420 genomes of *D. melanogaster* to the *D. melanogaster* v6.32 reference
193 genome [42] using *bwa mem* function [43]. We then used *Picard Tools*
194 (http://broadinstitute.github.io/picard/) to clean, sort and dedupe individual files before
195 individually genotyping them in *GATK4* [44] with the *HaplotypeCaller* function. All
196 samples were then jointly genotyped in *GATK4* using the *GenotypeGVCfs* function,
197 following *GATK* best practices [44]. The resultant VCF was filtered so that indels were
198 removed, and only biallelic sites with a minimum quality score of 30, minimum coverage
199 of 5X, minimum genotype quality of 30, a maximum of 50% missing data were kept. We
200 additionally removed 10 individuals with poor quality genomes (e.g. less than 5X

201 average coverage). For analyses requiring an outgroup (such as our phylogenetic
202 reconstruction, outlined below), we also included twelve *D. simulans* and one *D. yakuba*
203 genomes (see Table S1 for SRA accession numbers).These sequences were
204 processed as above, and VCF files were merged using *bcftools merge* function [45]. We
205 used this VCF file to perform the population genomic analyses, outlined below.

206

207 *Lineage relationships: population structure, PCA, and phylogenetic reconstruction*

208 To better understand the relationships among a global sampling of *D. melanogaster*, we
209 constructed maximum likelihood (ML) phylogenies for the autosomes and *X*
210 chromosome separately using *iqtree* version 1.6.12 [46–48]. We generated ML trees for
211 non-overlapping regions of 100KB using the model-finder and ultra-fast bootstrap
212 approach with 1,000 bootstraps. We then used the resulting ML trees as input for
213 *ASTRAL v5.1.1 [49]* in generating a consensus tree for the autosomes and *X*
214 chromosome independently (Figure S2).

215 To characterize fine scale population genetic structure we did a *K*-means
216 clustering analysis and PCA using *PCAngsd* [50]. PCAngsd uses genotype likelihoods
217 to first perform a genome-wide PCA, then assess population structure wherein the
218 number of ancestry types (*K*) is defined as the number of significant PCs + 1. For these
219 analyses, we used the deduped .bam files generated above. We only included sites with
220 <10% missing data, a minimum mapping quality of 30, and a minimum base quality of
221 20. This generated a dataset of 461,147 sites for 420 individuals.

222 Given our phylogenetic, PC, and *K*-means clustering analysis, we identified six
223 geo-genetic lineages which correspond to five geographic regions (all Out of Africa lines
224 (OOA); all East African lines excluding lines from Ethiopia (East), all West African lines,
225 all Ethiopian lines, nine lines from Subtropical Africa which form a unique ancestry
226 group that we refer to as Hahare-Distinct (HD), and all other Subtropical African lines
227 (Central)). We thus calculated all pairwise measures of divergence and differentiation
228 between all lineages identified from our phylogenetic, PC, and *K*-means clustering
229 analysis (e.g. F_{ST} and D_{xy}; defined in Figure 1) and nucleotide diversity within each
230 lineage (e.g. π) in 1KB windows across the genome using *pixy* [51] with default filtering
231 expressions (i.e. DP>=10,GQ>=20,RGQ>=20). For this analysis we used an all-sites
232 VCF to include invariant sites, and performed the same filtering above, but retained
233 sites with a maximum of 2 alleles, so as to include invariant and biallelic sites. We also
234 calculated Tajima's D using *VCFTools* [52], and estimated the Site-Frequency Spectra
235 (SFS) using *SweeD* [53] for each geo-genetic lineage.

236

237 *Introgression Analyses*

238 To estimate broad patterns of gene flow between genetic lineages of *D.*
239 *melanogaster,* we first calculated Patterson's *D* and f_G (a Patterson's *D* derivative which
240 more accurately estimates the proportion of the genome experiencing introgression
241 [54]) using *Dsuite* [55] with *D. simulans* and *D. yakuba* as outgroups for all possible
242 trios, given the following phylogeny: (((OOA, ((East, West),Ethiopia)),HD), Central).
243 Significance of Patterson's *D* was determined using a standard block jackknife

procedure [55]. In specific cases (outlined below) we also quantified differences in the extent of introgression between genetic lineages that show significant Patterson's *D*, using f_{dM} in non-overlapping 20 SNP windows for focal trios (outlined below). f_{dM} is a Patterson's *D* derivative that is more appropriate for windowed analysis and provides a more accurate estimate of the proportion of the genome that has experienced introgression than Patterson's *D [55]*. Finally, we bolstered our introgression analyses by assessing heterogeneity in the genome in the relationships between potentially introgressing groups by calculating tree topology weights using *twisst [56]*. For each focal trio (outlined below), we calculated the topology weight at each non-overlapping 100KB window for trees among four groups: *D. simulans* individuals as an outgroup, the two potentially introgressing groups (outlined below) and the remaining African flies as a single group (AFR). While f_{dM} calculates the proportion of shared derived variants between non-sister lineages, *twisst* assesses the proportion of phylogenetic trees that fit particular taxonomic relationships. These analyses thus provide complimentary, but uniquely informative quantifications of introgression.

Our goals were threefold: (1) assess if female mate choice in Subtropical Africa is associated with reduced gene flow with other African lineages, (2) quantify if urbanization in Subtropical Africa is associated with increased levels of gene flow with OOA, and (3) identify the source(s) of African ancestry in the SE United States. To address each of the specific questions listed above, we used specific trios.

<u>Gene flow between Central Africa and other African lineages</u>: we first assessed whether Patterson's *D* was significantly elevated from 0 between all trios involving HD or Central Africa as P3, and Ethiopia, West, and/or East Africa as P1 and P2. For trios that had

267 significant Patterson's *D*, we then calculated f_{dM} in 20 SNP windows across the genome
268 to evaluate the extent of introgression. To determine whether the extent of introgression
269 varied between trios, we used an ANOVA with windowed f_{dM} as the response variable,
270 and P3 (HD or Central Africa), P2 (East or West), and their interaction as independent
271 variables with a Type III Sum of Squares using the *car* package in R [57]. In all cases,
272 Ethiopia was used as P1, as there were no scenarios in which Ethiopia demonstrated
273 significant introgression with either HD or Central Africa (Table S3; Figure 3).

274 Urbanization and introgression with OOA: we calculated Patterson's *D* for all trios in
275 which Central Africa (composed largely of rural individuals) or HD (composed largely of
276 urban individuals) was P3, and OOA and either Ethiopia, East, or West Africa was P1 or
277 P2. Then, to determine if either HD or Central Africa has experienced more
278 introgression with OOA, we calculated f_{dM} in 20 SNP windows across the genome for
279 both comparisons with Ethiopia as P1, OOA as P2, and either Central Africa or HD as
280 P3. We assessed differences in windowed f_{dM} using an ANOVA with Type III SS using
281 the *car* package in R [57], with f_{dM} as the independent variable and P3 (HD or Central
282 Africa) as the independent variable.

283 Source(s) of African ancestry in the SE United States: we performed two sets of
284 analyses. Again, we first calculated Patterson's *D* with either lines from the SE United
285 States or all other OOA lines as P3, and Ethiopia, East and/or West Africa as P1 and
286 P2. Secondly, we calculated Patterson's *D* with HD or Central Africa as P3 and either
287 lines from the SE United States or all other OOA as P2, and each of Ethiopia, East and
288 West Africa as P1. As flies from the SE United States are hypothesized to represent a
289 contact zone between OOA and an unidentified African lineage, we expect that flies

290 from the SE United States should have increased levels of introgression with the source
291 African lineage(s), relative to the rest of OOA flies. Therefore, we again calculated f_{dM} in
292 20 SNP windows across the genome for trios with significant Patterson's *D*, and asked
293 which (if any) African lineages had higher f_{dM} with the SE-United States than with OOA
294 using ANOVAs with Type III SS using the *car* package in R [57] with windowed f_{dM} as
295 the response variable and each OOA lineage (SE-United States or all other OOA lines)
296 as the independent variable.

297 Last, to assess heterogeneity in introgression across the genome, we leveraged
298 the windowed f_{dM} analyses described above and first asked do chromosomes vary in the
299 extent of introgression using an ANOVA with type III SS in the *car* package in R [57].
300 For each focal trio we performed these analyses with f_{dM} as the response variable and
301 chromosome arm as the independent variable. We then used pairwise *t*-tests to assess
302 significant differences between chromosome arms. To determine what genomic regions
303 were specifically more likely to introgress in each of our focal comparisons, we also
304 defined introgression outliers as windows with the top 1% of f_{dM} values.

305

306 *Assessing mate preference in Subtropical African samples*

307 We next tested if patterns of diversity, differentiation, and population structure
308 correspond to Z and M mating types within *D. melanogaster* (as described in [30,32]).
309 While Z mating behavior seems largely restricted to Subtropical Africa, previous work
310 has identified that female mate choice is likely variable in the ancestral range of *D.*
311 *melanogaster [34]*—however, the frequency of Z behavior within the ancestral range, as

312 well as it's phylogenetic distribution among Subtropical African flies, is largely unknown.
313 To assess the prevalence and strength of female mate choice in Subtropical Africa, we
314 performed a series of replicated choice experiments for 47 focal isofemale lines. These
315 lines originate from four sites in Zambia (a total of 17 lines), three sites in Zimbabwe (a
316 total of 21 lines), one population from Malawi (two lines), one population from Botswana
317 (one line), and one population from Namibia (six lines). Each focal isofemale line was
318 tested in a choice experiment where 7-10 day old virgin females were presented with a
319 standard Z and M male in a vial containing cornmeal/Karo/agar medium. We used the
320 lines ZS2 and RAL371 as representative Z and M lines, respectively. We note that while
321 RAL lines have mixed African-European ancestry, RAL371 is 81.2% European
322 ancestry, slightly above the average 80.2% for the population as a whole [40]. Vials
323 were watched continually for up to three hours, and once mating began the unmated
324 male was removed from the vial by aspiration.

325 As Z and M males are morphologically indistinguishable, males were transferred
326 to vials containing cornmeal/Karo/agar medium that had been dyed with either red or
327 blue food safe dye approximately one hour prior to the mate choice experiment, and
328 were easily distinguished based on abdominal coloration. Across replicates, we
329 switched which males were fed what color to control for any bias abdominal coloration
330 might have on female preference, and we found no effect of food color on female
331 preference (based on an ANOVA with type III SS, with proportion of Z males chosen as
332 the response variable, and isoline and color of Z male as fixed effects: effect of color of
333 Z male: F=0.088, df=1, p=0.77; effect of isoline: F=1.57, df=47, p=0.039). In total, we
334 set up an average of 53 trials (range: 12-100) per focal isofemale line, and, on average,

335 37 trials per inbred line (i.e. ‘isoline’; range: 8-71) resulted in a successful mating (e.g.
336 1,724/2,507 total trials were successful). We tested for female mate preference using a
337 Fisher’s exact test to determine if the observed ratio of Z:M successful matings
338 significantly deviated from the expected under random mating (e.g. 1:1) for each
339 isofemale line (we note that the results are qualitatively the same when mate preference
340 is assessed using a binomial test). To assess population level differences, we also
341 performed two additional models: first, to assess if populations varied in the proportion
342 of females with strong preference, we did a logistic regression with the response
343 variable being whether lines had mating preference and population as the independent
344 variable. Second, to assess if the strength of preference differed between populations,
345 we performed a generalized linear mixed effect regression (glmer) with a poisson
346 distribution, with the count of Z males chosen as the dependent variable, and the
347 isoline, replicate trial date, and total matings as random effects. For both of these
348 analyses we used the *lme4* package [58] and assessed significance of fixed effects
349 using an Type III Wald’s X^2 test using the *Anova* function in the *car* package in R [57].

350 To test for genetic differences associated with female mate preference, we first
351 calculated the population branch excess (*PBE*) statistic [59] in 1KB windows across the
352 genome with Subtropical African lines with significant female mate choice as the focal
353 population, and Subtropical African lines that do not have strong female mate choice
354 and RAL lines as the non-focal populations. *PBE* is a derivative of the population branch
355 statistic (i.e. *PBS*; [60]), but quantifies branch-specific evolution for a focal population
356 relative to two non-focal populations [59]. We then defined behavioral *PBE* outliers as
357 the top 1% of *PBE* values (368 windows). Using this list of outliers, we did a Gene

358 Ontology (GO) Enrichment Analyses with a Fisher's Exact test with an FDR threshold of
359 0.05 using *PANTHER* 16.0 [61] to assess if behavioral *PBE* outliers were enriched for
360 biological GO terms relating to behavior, sensory perception, memory, learning or
361 neurological development, each of which may have specific function in female
362 choosiness or male attractiveness [62].

363 Next, we studied whether female mate choice in Subtropical Africa is associated
364 with global patterns of differentiation or introgression in *D. melanogaster,* as a test of the
365 hypothesis that female mate choice may serve as a significant reproductive barrier in
366 nature and contribute to global patterns of population genetic structure. To do this, we
367 calculated whether behavioral *PBE* outliers had elevated differentiation relative to the
368 rest of the genome for a broader sample of lines from Subtropical Africa. We used the
369 aforementioned estimates of F_{ST} calculated in 1KB windows between all Central Africa
370 (including all lines, regardless of phenotype) and each OOA geographic region (Asia,
371 Europe, North Africa, North America, and Tazmania), as well as F_{ST} between Central
372 Africa and each within-Africa geo-genetic lineage (West, East, Ethiopia, South, and
373 HD). This analysis enabled us to determine whether Central Africa broadly shows
374 increased differentiation for regions associated with female choosiness, in line with a
375 polygenic genetic basis of female mate preference [30,31]. We then performed an
376 ANOVA with Type III SS to assess whether F_{ST} differed between locus type (genome-
377 wide versus behavioral *PBE* outlier), between comparison (each geographic region),
378 and their interaction using the *car* package in R [57]. We used the *emmeans* package to
379 estimate significance of specific contrasts [63]. To determine whether female mate
380 choice was associated with differential patterns of introgression at the level of individual

381 loci, we also ascertained whether windowed f_{dM} between Central Africa and each of
382 West Africa and Europe differed between behavioral *PBE* outlier loci and the rest of the
383 genome, using the f_{dM} datasets described above. To do this, we performed a *t*-test with
384 windowed f_{dM} as the dependent variable, and locus type (behavioral *PBE* outlier vs
385 genome-wide) as the independent variable.

386 Finally, we asked whether isolines that had significant female mate choice also
387 showed an overall lower history of introgression than isolines with no strong female
388 mate choice. To do this, we recalculated Patterson's *D* and f_G for trios involving OOA,
389 West Africa, East Africa, and/or Ethiopia as P1 and P2, and isolines of each behavioral
390 type as P3. We restricted these analyses to include only Central Africa flies for which
391 we had known behavioral phenotypes, and excluded phenotyped flies from the HD
392 lineage, as these lines have substantial population structure from Central Africa and HD
393 and Central Africa are not monophyletic (Figure 1; Figure S5). We then performed a *t*-
394 test with f_G as the dependent variable and the identity of P3 (i.e. lines that showed
395 behavioral preference versus those that did not) as the dependent variable.

396

397 *Determining the global distribution of previously identified incompatibilities*

398 We characterized patterns of differentiation for loci that have previously been implicated
399 in genetic incompatibilities [40,41]. We note that while many these loci likely represent
400 genetic incompatibilities in the classical sense (i.e. Dobzhansky-Muller Incompatibilities,
401 [64,65]), some of these loci—particularly those from natural admixture zones—may be
402 involved in assortative mating or extrinsic reproductive isolation [66]. For simplicity, we

403 refer to these loci as putatively incompatible, and we focus on loci previously described
404 in two studies. First, [41] used a global panel of *D. melanogaster* inbred lines to create
405 synthetically admixed populations from a series of round-robin matings followed by
406 continual inbreeding. This design enabled the identification of pairs of alleles that
407 appear less frequently than expected under random mating and Mendelian segregation
408 in their final recombinant inbred line population (i.e. Genotype Ratio Distortion). Using, a
409 similar premise, but in a naturally admixed population, Pool [40] used patterns of linkage
410 disequilibrium in the SE United States to assess pairs of alleles that occur together less
411 frequently than expected based on their allele frequencies (i.e. Ancestry Disequilibrium).
412 [40] also determined that many of these loci were highly differentiated between Africa
413 and Europe, using populations from West Africa and France, respectively.

414 Elevated differentiation of these putative incompatibility alleles between West
415 Africa and France may stem from multiple evolutionary scenarios, and differentiating
416 these scenarios can help elucidate the geographic distribution and potential origins of
417 putative incompatibilities within *D. melanogaster*. Here, we aim to differentiate two
418 potential scenarios: First, putative incompatibilities between Europe and Africa may
419 have arisen with or after the Out of Africa expansion and thus may represent more
420 recently derived incompatibilities (i.e. within the last 10-23kya; [5,20,22]). Under this
421 scenario, we predict that differentiation at putative incompatibility loci should be low
422 between geo-genetic lineages in Africa, but high between Europe and all African
423 populations. Second, it is also plausible that putative incompatibilities between Europe
424 and West Africa are very old and also exist between West Africa and another geo-
425 genetic lineage within Africa. Shared ancestry or subsequent introgression could explain

426 the presence of these alleles in Europe. Under this scenario, we expect that
427 differentiation should be high between West Africa and both Europe and other African
428 geo-genetic lineages, but relatively low between Europe and other African geo-genetic
429 lineages.

430 To differentiate these scenarios, we used the F_{ST} windows from above for all
431 pairwise comparisons of Central Africa, West Africa, and Europe. We first ask if putative
432 incompatibility loci have elevated divergence relative to the whole genome for any
433 specific geographic pair using an ANOVA with Type III SS using the *car* library in *R* [57].
434 Specifically, F_{ST} was the dependent variable, and locus type (e.g. genome-wide, loci
435 identified by [41], or loci identified by [40]), population pair (Central Africa-West Africa,
436 Europe-Central Africa, or West Africa-Europe), and their interaction were the
437 independent variables. We then used the *emmeans* package in R [63] to determine
438 significance for specific contrasts. Second, we aim to assess the history and distribution
439 of individual putative incompatibility pairs and differentiate the two evolutionary
440 scenarios outlined above. For these analyses, we identified putative incompatibility pairs
441 in which both interacting loci have the predicted patterns of differentiation. We define
442 highly differentiated loci as those with F_{ST} values within the top 2.5% of F_{ST} for that
443 population pair.

444

445 **RESULTS**

446 *Diversity, divergence, and evolutionary relationships among populations of* D.
447 melanogaster

448 To understand the global distribution of diversity and population structure of *D.*
449 *melanogaster,* we combined whole genome resequence data from 420 lines of *D.*
450 *melanogaster* from around the world, including 174 newly sequenced genomes from un-
451 and undersampled, rural locales within the proposed ancestral range in Subtropical
452 Africa. PCA revealed that genetic variation within *D. melanogaster* is mainly structured
453 between flies from Subtropical Africa and Out of Africa (OOA), as reflected by PC1
454 (which explains 46.7% of the variation). While somewhat intermediate along PC1,
455 Ethiopia, as well as East and West Africa, are much more distinct from OOA and
456 Subtropical Africa along PC2 (which explains 18.6% of the total variation). We further
457 identified eight unique genetic ancestries using *K*-means clustering in *PCAngsd* [50].
458 Four of these ancestries predominantly occur in Subtropical Africa, three largely
459 correspond to one each of Ethiopia, West Africa, and East Africa, and a final ancestry
460 type is most common in all OOA accessions (including *D. melanogaster* from North
461 Africa; Figure 1).

462 *K*-means clustering and PC analyses largely agree with our consensus
463 phylogeny with *D. simulans* and *D. yakuba* as outgroups. Of the eight ancestry types
464 identified, four correspond to largely monophyletic clades; OOA, Ethiopia, West Africa,
465 and a unique ancestry type from Subtropical Africa that comprises nine individuals
466 predominantly from Harare, Zimbabwe (denoted by gold in Figure 1). We refer to this
467 lineage as Harare-Distinct (HD). Of the remaining four ancestries, individuals from East
468 Africa cluster together in a PCA, but are split across our phylogeny, with some
469 individuals sister to West Africa, and others sister to all OOA individuals. The remaining
470 three ancestries are largely found in individuals from Subtropical Africa, but do not

471 correspond to monophyletic clades or unique sampling locales. They also do not exist in
472 pure form; the vast majority of Subtropical African individuals comprise two, and
473 sometimes three of these ancestries (denoted as yellow, red, and orange in Figure 1).
474 While we find no strict correspondence between ancestry type and monophyly in these
475 Subtropical African individuals, we note that the largest and most distantly related clade
476 of Subtropical African flies has the largest proportion of one ancestry type (indicated by
477 orange in Figure 1), and largely contain samples from more remote sampling locales.
478 Combining these results, we define six geo-genetic lineages that we use for subsequent
479 analyses: OOA, Ethiopia, East Africa, West Africa, HD, and all other Subtropical African
480 flies (which we refer to as Central Africa). In some analyses, we also compare
481 individuals from South Africa, Northern Africa, and the SE United States, but we note
482 that these geographic groupings are mere subsets of individuals from our Central
483 African (for South Africa) or OOA (for North Africa and the SE United States) geo-
484 genetic lineages.

485 Despite substantial population genetic structure, global F_{ST} estimates were
486 relatively low in almost every comparison, with the lowest global F_{ST} found between
487 Central and South Africa (F_{ST}=0.022), and the highest between HD and OOA
488 (F_{ST}=0.246; Table S2). Variation in pairwise global F_{ST} is likely due to differences in
489 nucleotide diversity, as all pairwise comparisons of D_{xy} were very similar (Table S2), and
490 the populations with the highest global F_{ST} were those in which both populations also
491 showed the lowest pairwise nucleotide diversity (π; Table S2; Figure 2).

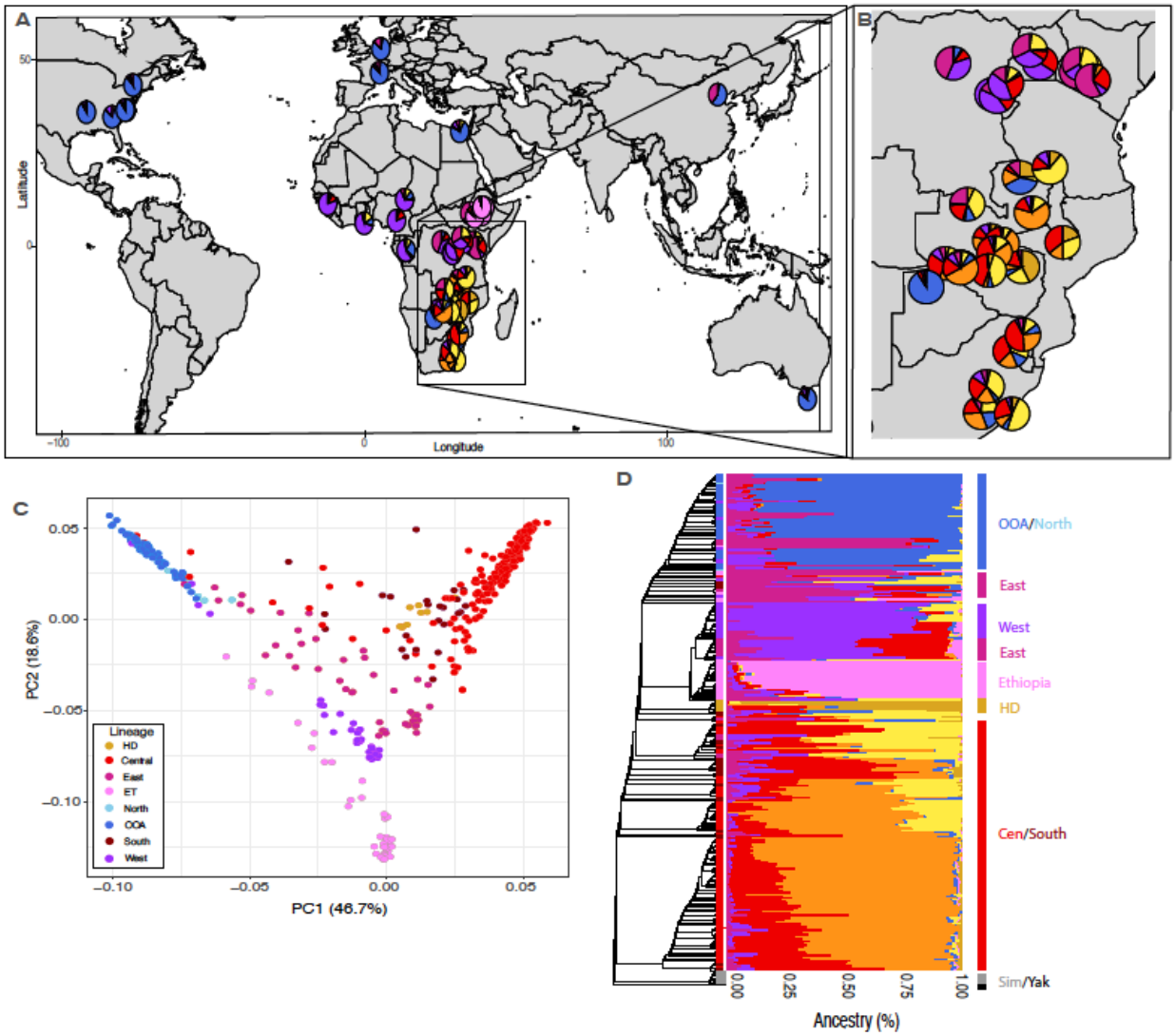


FIGURE1. Broad sampling of *Drosophila melanogaster* from Subtropical Africa reveals substantial cryptic genetic structure. *(**A**) Geographic sampling of 420 genomes from a global distribution of D. melanogaster; pie charts represent the average ancestry determined by PCAngsd from that sampling local. (**B**) Zoom-in panel from (A), focussing on Subtropical African accessions. (**C**) PCA of all genomes, colors indicate the genetic clade the samples reside in based on the phylogeny. (**D**) Phylogeny of all accessions with an ancestry plot denoting their average genomic composition of each sample. Labels to the right of the ancestry plot indicate how samples cluster into the 6 genetic lineages that we identify using PCA, K-means clustering, and phylogenetic reconstruction.*

In addition to substantial cryptic genetic structure, the Central African lineage contains substantially higher levels of genetic diversity than any other geo-genetic

505 lineage (Figure 2). This diversity is likely caused by an excess of rare alleles, as the
506 Central African lineage also shows the most negative values of Tajima's D, and a left-
507 skewed Site-Frequency Spectrum (SFS), in line with this lineage having a much larger
508 effective population size and a recent history of population expansion (as suggested in
509 [22]). In contrast, two lineages show signals of recent population contraction and lower
510 levels of diversity: OOA (including North Africa) and HD (Figure 2). South Africa shows
511 both higher values of Tajima's D, and a slightly less left-skewed SFS but does not show
512 a substantial reduction of genetic diversity. Tajima's D and the SFS also differ between
513 chromosomes, although the direction of these differences depended on the geo-genetic
514 lineage (based on an ANOVA with Type III SS: chromosome arm × geo-genetic lineage
515 effect: F=163.12, df=28, p<0.0001). For five lineages, the *X* chromosome had lower
516 values of Tajima's D and a slightly less left-skewed SFS (Central Africa, East Africa,
517 West Africa, HD, and South Africa), while Ethiopia, North Africa, and OOA show the
518 opposite pattern (Figure 2; Figure S3). Overall, these results suggest that both different
519 lineages, and different chromosome arms, of *D. melanogaster* have experienced
520 different demographic histories and even within Subtropical Africa, the HD lineage has a
521 significantly different demographic history than the Central African lineage.

522

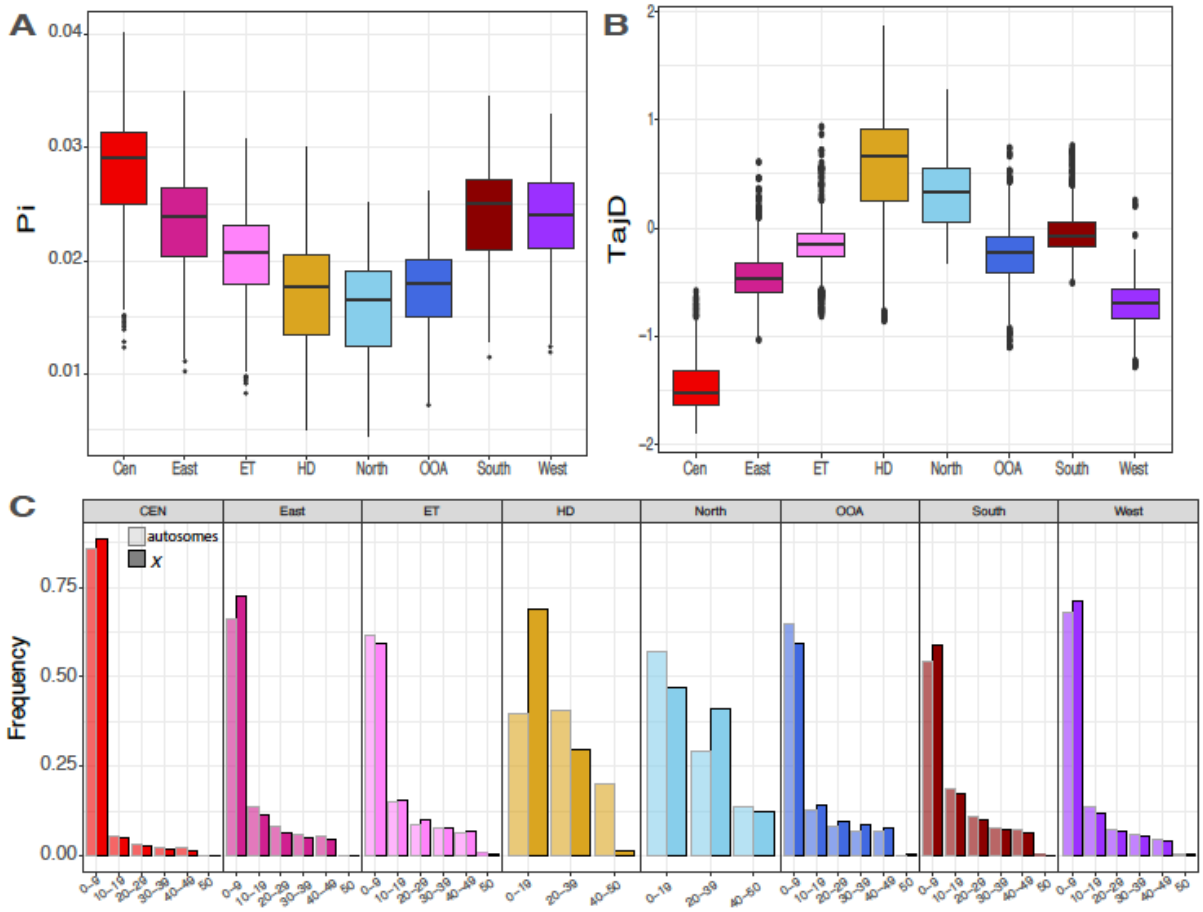


FIGURE 2. Genome-wide statistics for each geo-genetic lineage. *(A,B) Average pairwise diversity (π) and Tajima's D. (C) Folded site frequency spectrum for each genetic clade/ geographic lineage for autosomes (lighter, outlined in grey) and the X chromosome (darker, outlined in black). Data for the autosomes is averaged across 2L, 2R, 3L, 3R, and chromosome 4.*

<u>*Patterns of gene flow throughout the range of* D. melanogaster</u>

We next evaluated the extent of gene flow among distinct lineages within a global sampling of *D. melanogaster*. Specifically, we focus on three potential cases to better understand the sources and dynamics of gene flow across the range of *D. melanogaster*: (1) between Subtropical Africa and other African lineages, (2) the extent of gene flow between OOA and both HD and Central Africa, and (3) the source(s) of

536 African ancestry in the SE United States (as proposed by [27,28]). We describe the
537 results for each of these cases as follows.

538 First, we evaluated the extent of admixture among geo-genetic lineages within
539 Africa. We find evidence for extensive gene exchange between Central Africa and both
540 East and West Africa, but not between Central Africa and Ethiopia, nor between Central
541 Africa and HD (Table S3, Figure 3). Although gene flow between Central and both East
542 and West Africa is pervasive, the magnitude of gene flow is likely quite low, as only 1-
543 3% of the genome is inferred to be admixed between Central Africa and each of East
544 and West Africa, respectively (depending on whether f_G or f_{dM} is used; Table S3, Figure
545 3). Similarly, we find significant gene flow between HD and both East and West Africa
546 (Table S3, Figure 3). Despite closer geographic proximity between most East African
547 samples and Central African samples, introgression between West Africa and both
548 Central Africa and HD is significantly elevated compared to introgression between East
549 Africa and either Central Africa or HD (Figure 3; f_G between West Africa and either HD
550 or Central Africa ranged from 0.02-0.033, while f_G between East Africa and either HD or
551 Central Africa ranged from 0.017-0.022; we find no significant difference between
552 Central Africa and HD in the extent of gene flow with either West or East Africa).
553 Overall, this suggests that the magnitude of admixture is not equivalent between
554 Subtropical African and other African *D. melanogaster* lineages and may not simply
555 correspond to geographic distance.

556 Second, we evaluated the extent of introgression between OOA and both Central
557 Africa and HD. Specifically, we aimed to test whether back migration from OOA may be
558 responsible for the genetically unique HD lineage (denoted in gold in Figures 1 and 2)

559 as a test of whether urbanization has facilitated introgression in Africa. Other genomic
560 evidence suggests that this may be the case- the HD lineage shows substantially
561 reduced diversity and shows genomic patterns of a bottleneck relative to Central Africa,
562 despite no geographic separation (Figure 1, Figure 2), and mimics patterns of diversity
563 and Tajima's D seen in OOA populations (Figure 2). In line with this hypothesis, we find
564 evidence of admixture between HD and OOA; f_{dM} is elevated between OOA and HD
565 when compared to OOA and Central Africa (Figure 3), wherein 18-20% of OOA and HD
566 genomes have introgressed, while only 15% of genomes between OOA and Central
567 Africa have introgressed (depending on whether f_G or f_{dM} is used; Table S3, Figure 3).
568 However, tree topology weights generated by *twisst* further reveal that the unique HD
569 lineage does not appear to be simply a hybrid of flies from Central Africa and OOA
570 (Figure 3C). While large proportions of the genome show near complete support for
571 topologies in which HD is sister to OOA (purple, in Figure 3C), the second most
572 common topology does not place HD sister to samples from Africa, but rather places all
573 other African samples sister to OOA (teal, Figure 3C). This suggests that rather than
574 being a patchwork of African and OOA ancestry (in which we expect yellow and purple
575 topologies to be common in Figure 3C), HD carries distinct genetic variation, at least in
576 genomic regions where the topology supports Africa sister to OOA (such as the center
577 of 2L). Extensive introgression between other African lineages and OOA may also
578 contribute to these patterns. In total, this suggests that while HD has likely experienced
579 substantial introgression with OOA, it has also likely experienced a unique evolutionary
580 history from either OOA or the rest of Central Africa.

581 Last, we studied the origin of the African alleles harbored in the SE United

582 States. We find significantly elevated Patterson's *D* between SE United States and each
583 of Central Africa, West Africa, and HD which suggests these populations have
584 contributed to *D. melanogaster* from the SE United States. We do not observe a similar
585 pattern for Ethiopia or East Africa (Table S3, Figure 3). Patterns of introgression are not
586 unique to the SE United States, as we find elevated Patterson's *D* between all other
587 OOA lines and each of Central Africa, West Africa, and HD (Table S3, Figure 3),
588 suggesting that the contribution of these African populations might precede the split of
589 different OOA populations. However, if *D. melanogaster* from the SE United States
590 experienced a second pulse of introgression unique from other OOA populations, then
591 we predict that SE United States flies should show elevated signals of introgression with
592 at least one African source relative to other OOA flies. In line with this prediction, both f_G
593 and f_{dM} between West Africa and the SE United States are elevated relative to either f_G
594 and f_{dM} between West Africa and other OOA populations (Table S3, Figure 3), while
595 Central Africa and HD show no difference in either f_G and f_{dM} with the SE United States
596 or other other OOA lines (Table S3, Figure 3). However, it is challenging to detect
597 regions of the genome that show unique signals of introgression between West Africa
598 and the SE United States relative to West Africa and OOA, as the landscape of f_{dM}
599 across the genome is highly correlated between these two comparisons (r^2=0.59,
600 p<0.001; as it is between Central Africa and OOA and Central Africa and SE United
601 States: r^2=0.509, p<0.001;, and between HD and OOA and HD and SE United States:
602 r^2=0.607, p<0.0001). Moreover, the weighted topologies with West Africa as sister to all
603 OOA versus only accessions from the SE United States are highly similar (Figure S7).
604 In total, this suggests that all OOA populations have experienced some level of gene

605 flow with Central Africa, West Africa, and HD, but flies from the SE United States may
606 have experienced a second, and independent pulse of West African ancestry (as has
607 been hypothesized by [27,28]). The lack of distinction may also be caused by rapid
608 purging of introgressed alleles, reducing any signals of introgression in a short time-
609 span [67].

610 We next examined how patterns of introgression vary across the genome within
611 Africa and between African and non-African lineages. For these analyses, we focus on
612 the three scenarios outlined above: (1) Introgression within Africa (as measured
613 between West Africa and Central Africa), (2) introgression between Out of Africa and
614 the unique HD lineage, and (3) introgression between West Africa and the SE United
615 States. In all instances, Ethiopia was used as P1 as we find little to no evidence of
616 admixture between Ethiopia and any P3 used herein (Table S3). We find that
617 chromosomes significantly differ in the extent of introgression for all comparisons (as
618 measured by f_{dM}; West-SE-US: F=47.11, df=5, p<0.001; West-Cen: F=18.36, df=5,
619 p<0.001; OOA-HD: F=16.01, df=5, p<0.001). For all three comparisons, the *X*
620 chromosome showed significantly higher f_{dM} values, than almost every autosome (Table
621 S4). However, definitive evidence of increased introgression on the *X*-chromosome is
622 much less apparent when using weighted topologies (Figure 3, Figure S7).

623 Many genes which may be involved in the transition to human commensalism fall
624 within the top 1% of f_{dM} windows (see Table S5 for full list). For example, these windows
625 include several genes involved in insecticide resistance, (including *Cyp6a18* (between
626 OOA-HD), *Cyp313a* and *ACE* (West-SE-US), and *Cyp12a4* and *LRR* (West-Central)),
627 several genes related to metabolism, feeding behavior, and perception of and response

628 to food sources (including *happyhour*, *pbx, for,* and *Gfat1* (West-Cen), *NPFR, lovit,* and
629 *Ald1* (West-SE-US), and *Gr5a* (OOA-HD)), as well as genes involved in immune
630 function (*Dcr-2*, *Rab4*, *DptA*, *DptB*, and *Tl* (OOA-HD), *Npc2h*, *Npc2g*, and *ben* (West-
631 Cen), and *Charon* and *Ance-2* (West-Se-US)). Although causative connections are still
632 needed, it is perhaps unsurprising that alleles putatively related to human
633 commensalism may be overrepresented in three examples of potentially long-range,
634 and potentially human-mediated admixture.

635 Lastly, we find several genes associated with mating behavior fall within the top
636 1% of introgressed regions including *Desat2* (West-SENA; although we note that this
637 gene may also be involved in desiccation resistance, as is the case with other cuticular
638 hydrocarbons [68,69]), *fru* (West-CEN), and *5-HT1A*, *clt*, and *Oamb* (OOA-HD). While
639 these regions may not be involved in human commensalism *per se*, introgression at
640 these loci may have implications for the distribution of female mate choice, and the
641 distribution of female mate choice may be influenced by human commensalism; a
642 possibility that we explore below.

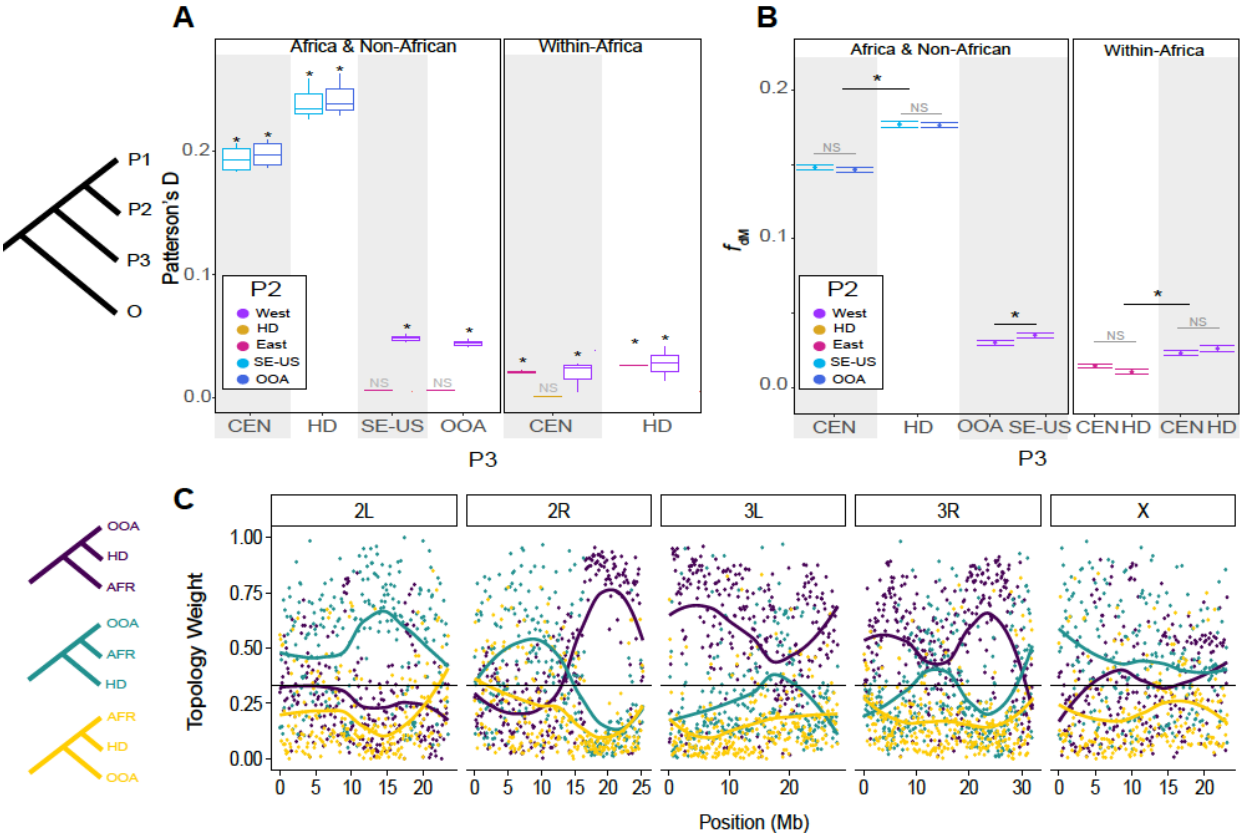


FIGURE 3: Patterns of gene flow between different African and non-African lineages of D. melanogaster. *(**A**) Boxplots of Patterson's D for all trios given the phylogeny (((OOA, ((East, West),Ethiopia),HD), Central Africa). P3 is given on the X-axis, with the identity of the P2 population denoted by the color of the boxplot. Values of D represent the range of values over multiple P1 populations. Comparisons that yielded a significant Patterson's D based on a standard block jackknife procedure [55] are denoted with an asterisk, otherwise they were deemed Non-Significant (NS). (**B**) Mean and standard errors of 20-SNP windows of f_{dM} from across the genome for each African geo-genetic lineage showing significant introgression with either other African or non-Africa lineages. The OOA lineage is split into those that heil from the SE United States (SE-US) and all others (OOA). Significance was determined by* ANOVA*s with Type III SS: NS= Not Significant, * = 0.01<p<0.05. (**C**) Weighted topologies for three configurations of OOA, HD, and Africa (denoted by colored phylogenies to the left).*

<u>*Patterns of female mate choice in Subtropical Africa*</u>

660 To understand if and how strong female mate preference originally described in
661 Subtropical Africa corresponds to any of the Subtropical African ancestries that we
662 identify herein, we surveyed a subset of lines from Subtropical Africa for the existence
663 of female mate choice using replicated choice experiments. Briefly, 47 focal isofemales
664 from ten sampling locales in Subtropical Africa were presented with both a Subtropical
665 Africa Z (ZS2) or OOA M (RAL371) isoline male. We quantified female mate preference
666 as a significant deviation from random choice, wherein ZS2 males were more likely to
667 copulate than RAL371 males. We find that female mate preference was common in our
668 experiment, but far from fixed across Subtropical Africa (e.g. 19/47 lines had a
669 preference for ZS2 males; Table S6, Figure S4). The frequency of lines showing female
670 mate preference varied among sampling locales, with the average proportion of females
671 exhibiting preference per population ranging from 0-100% (mean across all
672 populations= 44%; X^2=19.04, *df*=9, *p*=0.025; Figure 4; Table S6). Populations also
673 varied in the strength of preference. Some populations had nearly complete preference
674 for Z males, and others showed a mild aversion to Z males (X^2=18.63, *df*=9, *p*=0.029;
675 Table S6; Figure S4; average proportion Z males chosen per population ranged from
676 0.398-0.969). Lastly, we find that flies that with female mate choice are not
677 monophyletic, nor do any major PCs ascribing genetic variation among phenotyped flies
678 separate isolines with female preference from those that do not (Figure S5). Our results
679 suggest that the dynamics of mate preference are complex with Subtropical Africa; both
680 the presence and strength of female mate choice vary within and among populations,
681 and this trait is not monophyletically distributed.

682 We leveraged the natural diversity in female mate choice among these 47 lines

683 from Subtropical Africa to determine what regions of the genome are associated with
684 female mate choice. Given that isolines with strong preference do not show significant
685 population structure relative to isolines with no strong preference (Figure S5), we
686 instead quantified branch-specific evolution for lines with significant female mate
687 preference. We calculated the population branch excess (*PBE*) statistic [59], with
688 Subtropical African lines with strong female mate preference as the focal population and
689 Subtropical African lines with no strong female mate choice and RAL as the non-focal
690 populations. We defined the top 1% (368 windows) as outliers. These regions represent
691 all 5 chromosomal arms, although there is a significant excess of outliers on 3L and on
692 the *X* chromosome (X^2=183.01, df=4, p<0.0001). Of these 368 windows, 283 contain at
693 least one gene, and 84 of the remaining 85 windows are proximal to at least one gene
694 (of the windows in which genes are proximal, genes occur within 16.2KB on average,
695 range: 2.5-100KB). A total of 407 unique genes are included either within these
696 windows, or in close proximity (i.e. within 16.2KB on average). Gene Ontology
697 overrepresentation analyses indicate that of these 407 genes, there is an excess of
698 genes associated with learning, memory, cognition, chemotaxis, sensory perception,
699 detection of stimulus, behavior, and several ontology categories associated with
700 neurological development (as well as other biological functions; see Table S8 for
701 details).

702 Furthermore, several individual genes are of potential interest for future study
703 (Table S7). We find thirty genes that are known to influence behavior in the top 1% of
704 *PBE* outliers, including five genes specifically involved in courtship behavior (*Shep, Rdl*,
705 *Tbh, lig*, and *eag*). Thirty-eight genes that are involved in sensory perception and

706 sensory learning are also within the top 1% of *PBE* values, particularly those involved in
707 sensory perception of smell or olfactory learning, including three odorant receptor genes
708 (*Or22b*, *Or67c*, and *Or85e*), four defective proboscis extension response genes (*dpr8*,
709 *dpr13*, *dpr14*, *dpr18*), seven ionotropic receptor genes (*Ir11a*, *Ir67b*, *Ir67c*, *Ir75a*, *Ir75b*,
710 *Ir94f*, *Ir94h*), *Dop2R* and *vn* (both involved in olfactory learning), as well as *Ank2* and *sif,*
711 which are involved in sensory perception of sound and visual perception, respectively.
712 Five genes which are known to influence male aggression fall in the top 1% of
713 behavioral *PBE* values (*Dop2R, Rpb6, CkIIalpha, Tbh,* and *Rdl*). Z males tend to be
714 substantially more aggressive during courtship than their M male counterparts,
715 particularly to Z females (Figure S8). Lastly, ten genes that are involved in mushroom
716 body development are also included in the top 1% of *PBE* values, including *Frl, DAAM,*
717 *PsGEF*, *rg*, and *rad*. Mushroom bodies play a central role in learning and memory,
718 particularly of olfactory perception, and therefore may be important in differentiating
719 male mates. Furthermore, mutants with aberrant mushroom bodies and inhibition of
720 mushroom bodies can cause virgin females to reject matings [70,71]. In total, we find
721 that *D. melanogaster* from Subtropical Africa with significant mate preference differ in
722 many regions across the genome, indicative of a polygenic basis of female mate choice.

723 We next sought to assess whether loci putatively involved in female mate choice
724 were also more highly differentiated among a broader geographic sampling of *D.*
725 *melanogaster* as a test of polygenic adaptation of this trait. Specifically, we asked
726 whether behavioral *PBE* outliers were among the most highly differentiated loci between
727 all Central Africa lines (regardless of phenotype) and other geographic regions. We find
728 that F_{ST} among behavior outliers is elevated between Central Africa and all other

729 genetic lineages relative to genome-wide F_{ST} (locus type (e.g. genome-wide versus
730 *PBE* outlier): F=7.57, df=1, p=0.0059; geo-genetic lineage effect: F=29085, df=6,
731 p<0.0001, locus type × lineage interaction: F=1.29, df=6, p=0.26, Figure 4, Figure S6).
732 Further analysis revealed that differences in F_{ST} between behavioral *PBE* outlier loci
733 and genome-wide estimates are greater between Central Africa and OOA and between
734 Central Africa and HD than between Central Africa and any other African lineage
735 (Figure S6). However, despite this modest evidence for selection, we do not find that
736 *PBE* outlier loci are less likely to introgress than the rest of the genome, as would be
737 expected if female mate choice served as a significant barrier to introgression (f_{dM} for
738 behavioral *PBE* outliers is not significantly reduced relative to the rest of the genome
739 between Central Africa and either West Africa: t=-0.064, df=406.27, p=0.95 or Central
740 Africa and Europe: t=-1.52, df=428.69, p=0.13). Furthermore, individual isolines that
741 differ in the strength of female mate preference show no difference in the history of
742 introgression experienced (mean f_G for lines with female mate choice = 0.097, mean f_G
743 for lines with no strong female mate choice= 0.112; t=-0.27, df=9.7, p=0.79). In total,
744 this work suggests that while female mate choice may show modest signals of
745 polygenic adaptation, it is likely not a very effective barrier to gene flow in natural *D.*
746 *melanogaster* populations.

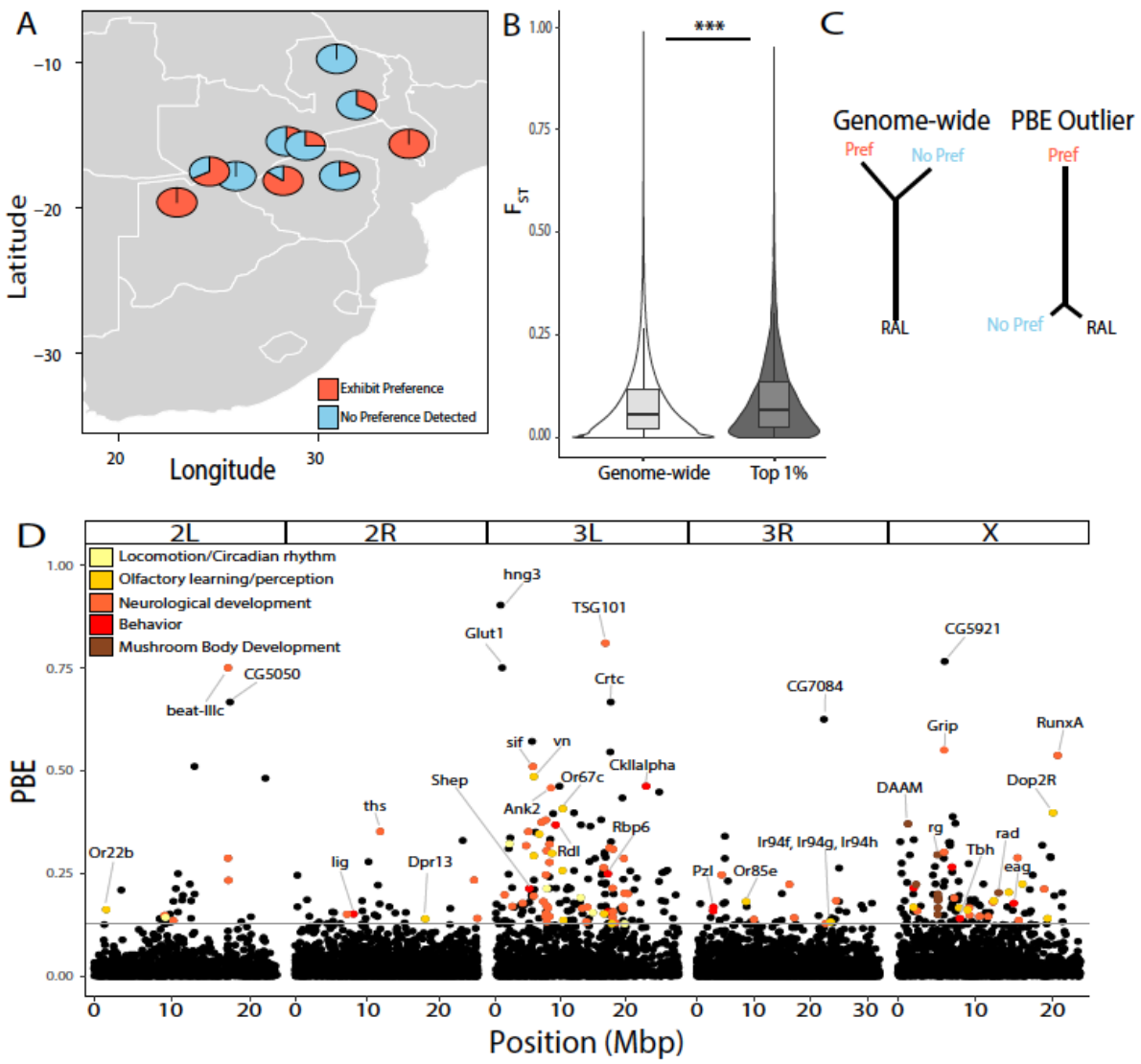


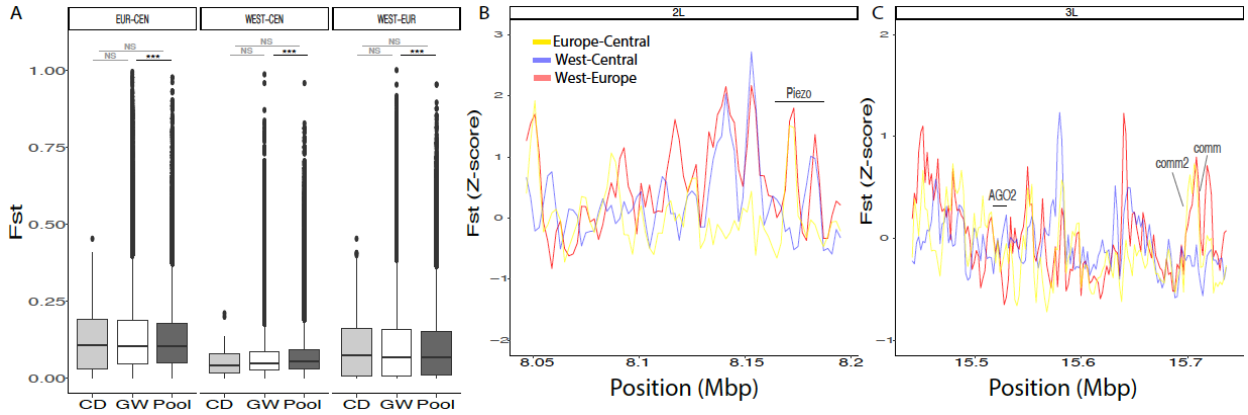
FIGURE 4: Geographic and genetic dissection of female preference in Subtropical Africa. *(**A**) Frequency of isofemale lines that show significant female mate choice in two-male choice experiments per sampling locale. (**B**) Average F_{ST} is elevated in the top 1% PBE regions between the Central Africa lineage (regardless of mate preference phenotype) and all other geographic regions. (**C**) Cartoon illustrating that the PBE statistic can infer branch-specific evolution of the focal population (in this case, flies with significant female mate preference (red, "pref"), relative to flies with no preference (blue, "no pref") and a third population of RAL)). (**D**) PBE of D. melanogaster lines with strong mate preference, measured in 1KB windows across the five major chromosome arms of D. melanogaster. Grey horizontal line indicates the top 1% threshold. Colored dots indicate a sample of genes with functions relevant to female mate preference.*

760 *Distribution of putative incompatibilities throughout D. melanogaster*

761 We last assessed the distribution of alleles potentially involved in negative
762 epistatic fitness effects in admixed populations (herein 'incompatibility loci') throughout
763 the range of *D. melanogaster.* Although we refer to these loci as incompatibilities, they
764 include any loci that are found in repulsion of one another, and thus may include
765 traditional incompatibilities (i.e., involved in intrinsic postzygotic isolation), alleles
766 involved in ecological hybrid breakdown, as well as loci involved in assortative mating
767 [66]. We find that loci identified by [40], show slightly increased overall levels of
768 differentiation than the genome-wide average between Central and West Africa, but
769 significantly lower differentiation than the genome-wide average between Europe and
770 each of Central and West Africa (Figure 5). In contrast, we find no significant difference
771 in average F_{ST} between loci identified by [41] and the genome-wide average for any
772 population pair. Thus, while on average, putatively incompatible loci are within the most
773 differentiated loci between West and Central Africa, they are not broadly differentiated
774 on a global scale.

775 We next assessed whether specific pairs of incompatibility loci showed elevated
776 differentiation, and used patterns of differentiation to characterize their potential
777 geographic origins. Out of 445 putative incompatibility loci identified by [40] and 45
778 putative incompatibility loci identified by [41], we identified only eight pairs of interacting
779 incompatibility alleles with high differentiation within Africa, and low differentiation
780 between Central Africa and Europe; indicative of potentially older incompatibilities that
781 divide genetic lineages within Africa. Within these eight loci, eight unique genes are
782 included in the F_{ST} outlier windows, two of which were also found in our *PBE* behavioral

783 analysis (*Abl,* and *eag;* Table S9). In contrast, we identified 71 pairs of loci with a
784 signature of more recent origin (i.e. low differentiation in Africa, high differentiation
785 between Europe and both West and Central Africa). We find that the F_{ST} peaks within
786 these 71 pairs of loci contain 133 unique genes (Table S9), 13 of which are also
787 behavioral *PBE* outliers. These include *DAAM, Rbp6, beat-IIIC, Doa, Lasp, luna, Eip75,*
788 *Mp, olf413, Pex7, pk, Svil,* and *eag* (we note this is a separate 1KB window than the
789 window which shows a signal of an putatively older incompatibility, described
790 immediately above). Many of these genes are involved in neurological development
791 and/or behavior (*DAAM, Rbp6, eag, beat-IIIC, Eip75, pk,* and *Mb*), and reproduction
792 (*Doa, Lasp*). Thus, many putative incompatibilities show a signature of more recent
793 evolution, and a significant proportion of these loci may also be associated with
794 behavioral differences within Subtropical Africa.



795
796 **FIGURE 5: Global differentiation of putative incompatibility loci.** *(**A**) Distribution of*
797 *F_{ST} for incompatibility loci from CD [41], Pool [40] and genome-wide for three*
798 *comparisons: Central Africa vs Europe, West Africa vs Central Africa, and West Africa*
799 *vs Europe. (**B,C**) Two zoomed in windows representing one incompatibility pair, as*
800 *identified by Pool [40]. In panel (**B**) divergence is elevated between Europe-Central*
801 *Africa and West-Europe (but not West-Central Africa) for the length of the gene Piezo,*
802 *indicating this allele may be more recently derived in OOA populations. In (**C**) we show*
803 *the corresponding locus: again, divergence is elevated between Europe-Central Africa*
804 *and West-Europe (but not West-Central Africa) in a window containing no genes, but*
805 *slight up and down stream of comm and comm2, respectively. We also highlight the*

806 *approximate position of AGO2, the hypothesized interacting partner (based on [40]),*
807 *and show there is no increased divergence in this region. F_{ST} is plotted as a*
808 *chromosome-level Z score, so as to more easily compare the three measures of*
809 *divergence.*

810

811 **DISCUSSION**

812 The evolutionary history of genetic model systems has been the target of extensive
813 research, including *D. melanogaster* [4,5,16,18]. Nonetheless, gaps in the sampling of
814 *D. melanogaster* across regions of its range have left crucial aspects of its history
815 unexplored. We collect samples and explore the partitioning of genetic diversity and
816 patterns of gene flow across the global range of *D. melanogaster,* and evaluate the
817 distribution of two sources of polymorphic reproductive isolation that have been
818 previously described (mate choice and putative genetic incompatibilities). We find that
819 flies from Subtropical Africa harbor previously unknown genetic diversity and population
820 genetic structure. We also find that *D. melanogaster* has experienced a complex history
821 of gene flow, particularly within Africa. Finally, we find that behavioral isolation and
822 hybrid incompatibilities segregating within the species have a multilayered history, but
823 overall, do not explain patterns of population genetic structure in Subtropical Africa.
824 These results contribute not only to our understanding of the natural and evolutionary
825 history of a powerful model system, but also in our understanding of how reproductive
826 isolation influences global patterns of population genetic structure. We discuss each of
827 these implications in the following paragraphs.

828

829 *Population genetic structure in the ancestral range of* Drosophila melanogaster

830 *Drosophila melanogaster* individuals from its purported ancestral range of Subtropical
831 Africa show the highest genetic diversity and the lowest values of Tajima's D, in line
832 with a demographic history of population expansion and/or strong purifying selection (in
833 agreement with [13,15,16,20,22]). While these flies, originating largely from Zimbabwe,
834 Zambia, Malawi and Namibia cluster together in a genome-wide PCA, our phylogenetic
835 analyses indicate that Subtropical African individuals comprise a non-monophyletic
836 clade that is most distantly related to all other *D. melanogaster,* as would be expected if
837 the mopane forest in countries like Zambia and Zimbabwe are the likely origin of *D.*
838 *melanogaster* variation [4,5]. However, we also describe a strong signal of structure
839 among individuals from the ancestral range. We identify at least four unique ancestries
840 using *K*-means clustering within Subtropical Africa, only one of which is monophyletic in
841 our phylogeny. This monophyletic ancestry group—which we refer to as Harare Distinct
842 (HD) —comprise a distinct cluster in a whole-genome PCA (Figure 1). HD has
843 substantially reduced nucleotide diversity and elevated Tajima's D, indicative of a
844 population bottleneck or recent introgression (or less likely, balancing selection; [72]).
845 These individuals also have evidence of elevated introgression with OOA relative to
846 Central Africa. Given that seven of nine individuals in this lineage are derived from the
847 urban center of Harare, HD may be a product of human-assisted migration of OOA
848 individuals into urban centers in Subtropical Africa, in line with earlier work which used
849 microsatellites [25], and similar to patterns shown in West Africa [26]. While our results
850 are partially in line with this hypothesis, we also show that HD is likely not simply an
851 admixed population between OOA and the Central African lineage. Although much

852 remains unknown about the HD clade, the presence of previously undescribed diversity
853 within the ancestral range of arguably one of the most well studied model organisms
854 highlights the need for further sampling and genetic interrogating of flies from
855 Subtropical Africa.

856 *Drosophila melanogaster* from other regions within Africa also show strong
857 population genetic structure, with West Africa, East Africa, and Ethiopia largely being
858 identified as three distinct ancestry types, each clustering uniquely in a PCA, and West
859 Africa and Ethiopia being largely monophyletic in our phylogenetic analyses. Individuals
860 from these areas also show intermediate levels of diversity and Tajima's D, in line with
861 the hypothesis that, while these flies likely do not represent ancestral populations, they
862 were also likely some of the first steps of range expansion from Subtropical Africa (as
863 suggested by [5]). We find evidence of gene flow between both East and West Africa
864 and Subtropical Africa. In contrast to East and West Africa, Ethiopia does not show any
865 introgression with Subtropical Africa and represents a genetically unique lineage from
866 other East Africa flies (in agreement with [16,18]). Many of the Ethiopian samples herein
867 originate from the high elevation populations from the Ethiopian Highlands; an
868 environment that might not be permissive to migration from the lowlands [73–75].
869 Substantial geographic and/or ecological barriers to migration may thus isolate
870 individuals from the Ethiopian Highlands from the rest of Africa.

871 In contrast, individuals from outside of Africa show substantially reduced
872 complexity; all OOA flies cluster together in a genome-wide PCA, and largely ascribe to
873 a single ancestry. As a whole, these samples have reduced diversity and elevated
874 Tajima's D, which is suggestive of a bottleneck, likely accompanying an out of Africa

875 expansion (as has been found previously [13,20,22]). We find that all OOA individuals
876 are sister to a subset of East African individuals, specifically those from Kenya, but as a
877 whole, OOA lines do not show evidence of introgression with East Africa. It is therefore
878 possible that OOA lineages are derived from East Africa, and specifically from the
879 ancestors of modern Kenyan flies. Additionally, we find that African lineages have
880 differentially contributed to genetic variation within OOA via post-expansion
881 introgression. We find that all OOA flies have similar levels of introgression with the
882 Central Africa and HD lineages, while individuals from the SE United States show
883 slightly elevated levels of gene flow with flies from West Africa, relative to other OOA
884 individuals. It has previously been suggested that flies from the SE United States
885 represent a secondary contact zone between individuals of European and African
886 ancestry [16,18,27,40], here we show that different African lineages have contributed
887 differentially to *D. melanogaster* populations outside of Africa.

888 Taken together, these broad-scale analyses suggest that *D. melanogaster* has
889 substantial population genetic structure within the ancestral range in Subtropical Africa,
890 including the presence of a unique, and potential admixed lineage found in Harare,
891 Zimbabwe. Patterns of gene flow within Africa are complex, as there is at least some
892 degree of gene flow among most geo-genetic lineages; highland Ethiopian lines
893 remaining relatively distinct and unconnected. Gene flow between African and non-
894 African lineages is also variable, depending on the African and non-African population in
895 question, with relatively even patterns of introgression between Subtropical Africa and
896 OOA populations, but an increase in West African ancestry in SE United States
897 populations, relative to other OOA populations.

898 *Gene flow is elevated in genes related to human-commensalism*

899 In addition to variation in broad-scale patterns of gene flow between distinct geo-genetic
900 lineages of *D. melanogaster,* we also find a dynamic landscape of introgression across
901 the genome. First, using f_{dM}, we find that for all comparisons introgression is higher on
902 the *X* chromosome than autosomes (Table S4). This is unusual as the sex
903 chromosomes are often depleted for introgression [76–81], and indeed even within *D.*
904 *melanogaster,* OOA populations often lack African ancestry on the *X* (for example see
905 [40]), although several examples of systems with high levels of gene flow on the *X* also
906 exist [82–85]. As the *X* chromosome also contains an excess of alleles associated with
907 Z/M female mate choice, the finding that the *X* chromosome exhibits the highest levels
908 of introgression has implications for the efficacy of female mate choice as a barrier to
909 introgression, a possibility that we explore below.

910 Genomic windows with the highest levels of introgression contain an excess of
911 genes associated with human commensalism, specifically genes involved in insecticide
912 resistance, metabolism and immune system function. Humans present organisms with
913 novel, often adverse conditions. Human habitat change is a primary driver of
914 hybridization in many systems [86]. While the loss of local adaptation due to
915 introgression of domesticated traits is often considered to be one of the threats of
916 hybridization [87], wild populations without certain domesticated traits may simply be
917 unable to survive in human proximity. The genes we identify as outliers in introgression
918 analyses are often associated with the unique pressures of human commensalism (see
919 [88] for an overview)— insecticide resistance [5,89,90] and diet [91,92] are the most
920 salient cases. Increased population densities associated with human habitation may

921 also favor changes to the immune system [93], and genes involved with metabolism
922 have been identified as potentially important to domestication in several cases [94–96].
923 Moreover, previous work has shown that *D. melanogaster* from non-human associated
924 collection sites do show elevated, branch specific evolution for many immunity related
925 genes [5], suggesting that the transition to human commensalism in *D. melanogaster*
926 specifically may involve allelic changes in immunity.

927 We also find that several genes involved in mate choice and courtship behavior
928 fall within the top 1% of introgressed regions. One potential explanation is that strong
929 female mate preference comes at a fitness cost, potentially by limiting mating potential
930 or delaying reproduction. Thus, it is possible that the cost:benefit ratio of female choice
931 may differ under scenarios of human commensalism, ultimately favoring the loss of
932 female choice in these environments. Under this hypothesis, subsequent introgression
933 of less-choosy alleles back into Subtropical Africa may be favored. This may be
934 particularly true in more urban areas, which may suffer from both continual introgression
935 from non-choosy donor populations, and potentially a higher cost of female choosiness.
936 Cost in urban areas may be mediated by selection for faster reproduction via increased
937 resource competition in regions with higher *D. melanogaster* density. Although much
938 more work is needed to further test this hypothesis, previous work has shown the loss of
939 female preference in more urban populations of *D. melanogaster* from Brazzaville,
940 Congo [26]. Similarly, we find that the urban HD clade (which is potentially admixed)
941 encompasses fewer lines with strong female mate choice than the rest of our sample
942 from Subtropical Africa, despite close geographic proximity (Figure 4, Figure S5).

943

944 *Female mate choice is polymorphic and likely highly polygenic*

945 Female mate preference in Subtropical African *D. melanogaster* is a classic example of
946 incipient speciation since its original discovery in the mid 1990s [30–32,34,97–99]. Here
947 we revisit this classic system to survey flies from Subtropical Africa for the presence and
948 strength of female mate choice, particularly in previously unsampled, remote locales in
949 Zambia, Zimbabwe, and Namibia. We find that female mate choice is relatively
950 common, but highly polymorphic both within and among sampling locales, and not
951 monophyletically distributed on our phylogenetic tree. These results are in line with
952 previous work which has shown substantial variation in a female mate preference for a
953 smaller number of isolines [34], as well as recent work highlighting the variation in male
954 courtship behavior within Subtropical Africa [30–32,34,97–99]. Three evolutionary
955 scenarios may explain this polymorphism: (1) female mate choice is an ancestral
956 polymorphism which is maintained in Subtropical Africa, (2) strong female mate choice
957 is the ancestral condition, and introgression from outside of this region has eroded its
958 prevalence within Subtropical Africa, or (3) female mate choice is a derived character
959 within Subtropical Africa and has either been eroded by introgression or never fully fixed
960 throughout this region. While we cannot fully disentangle these scenarios with our
961 current dataset, which of these scenarios has occurred in this system has implications
962 for the stability of reproductive isolation in natural populations, as well as the potential
963 evolutionary costs and benefits of female choosiness.

964 One potential caveat to the above female mate choice work is that we used a
965 standard Z and M line for all assays as a way to standardize the female preference
966 tests. However, one possibility is that female mate choice is more common than we

967 estimate, but preference is specific to local males (e.g. not all Z males are equivalent;
968 [97–99]). Several lines of evidence suggest that this may be true for some populations;
969 for example, the population with one of the highest levels of female preference was the
970 population in which the focal male was derived from (e.g. ZS). Additionally, work that
971 has used a standard M line and a male of the same isoline as the focal female generally
972 showed greater values of female mate choice [34]. Lastly, African males vary in both
973 cuticular hydrocarbon profiles and mating displays [99,100]. Therefore, while we present
974 the most widespread survey of female mate choice from Subtropical Africa to date,
975 more research will be needed to more precisely characterize the dynamics of female
976 choice in this system.

977 We leveraged natural variation in the presence of female mate choice within
978 Subtropical Africa to assess genomic differences associated with female pickiness and
979 male attractiveness. Although genomic outliers exist on all major chromosome arms,
980 chromosomes 3L and the *X* are highly enriched for outliers. These results are largely
981 congruent with previous analyses [30,31], which found that the third chromosome had
982 the largest effect on both female preference and male attractiveness, including a large
983 contribution from 3L. In contrast, [30], found a much smaller contribution of the *X*
984 chromosome in single chromosome replacement experiments, but substantial epistasis
985 between the *X* chromosome and both major autosomes.

986 Behavioral *PBE* outlier regions are significantly enriched for genes that are
987 involved in male courtship behavior, olfactory perception, learning, memory, and
988 neurological development. Several individual genes present interesting candidates for
989 future work, including genes involved in mushroom body development and sensory

990 perception. Mushroom body ablation in virgin females results in higher rejection rates
991 [70,71], and plays a role in male memory and courtship effort [101]. Additionally, while
992 the precise male traits controlling attractiveness are unknown, cuticular hydrocarbons
993 likely play an important role [97,99,102], and therefore, evolution of sensory perception-
994 particularly sensory perception of smell and taste may also help regulate female mate
995 preference (see [103] for an example of sensory perception of taste in male mate
996 choice). Lastly, we find that two genes— *shep* and *eag*— which have been shown
997 directly to affect mating behavior are highly differentiated in Subtropical African lines
998 with significant female preference. *Shep* is responsible for neuron remodeling during
999 metamorphosis, and whose loss-of-function results in increased rejection of male mates
1000 in virgin females [104]. *Eag* regulates potassium permeability, and mutants of this gene
1001 have been shown to influence the amount of time males spend courting [62,105]. Thus,
1002 while we present some intriguing candidate genes for future functional analyses, overall
1003 our work suggests that the genetic basis of mate preference is likely highly polygenic.

1004 Lastly, while we find that *PBE* behavioral outliers show elevated differentiated
1005 between a broader sampling of the Central African lineage and non-Central African
1006 populations (particularly those from OOA), we also find that these behavioral *PBE*
1007 outliers do not show decreased levels of introgression relative to the rest of the genome,
1008 nor do lines with strong female preference show a reduced history of introgression.
1009 Taken together, these results could arise from a scenario in which female mate
1010 preference has experienced polygenic local adaptation, but this asymmetric mate
1011 preference is insufficient to noticeably dampen gene flow between mating types in
1012 Subtropical Africa. In particular, female mate choice may still be maintained if the

1013 genetic basis of mate choice involves genetic redundancy (e.g. multiple alleles are
1014 sufficient to cause female mate choice, and therefore the loss of any given allele by
1015 introgression does not dampen female mate choice). Earlier mapping efforts [30,31]
1016 suggest the existence of genetic redundancy in female mate preference and male
1017 attractiveness, though more detailed genetic analyses are needed.

1018

1019 *Putative incompatibilities are highly differentiated, and many are likely recently derived*

1020 We lastly studied the distribution of previously identified putative incompatibility
1021 loci across our global sampling of *D. melanogaster.* We find that while, on average,
1022 these loci show slightly elevated differentiation between Central and West Africa.
1023 However, while some pairs of loci show a signature of differentiation within Africa, there
1024 are substantially more pairs of loci that are highly differentiated between Europe and
1025 both West and Central Africa. Moreover, we find that of the putative incompatibility loci
1026 that are highly differentiated (either between Europe and both African lineages or within
1027 Africa), 10-25% of genes within these windows overlap with our behavioral *PBE* outliers.
1028 There are two explanations for the high overlap between these two datasets. First,
1029 some of the loci identified by [40] may represent loci involved in assortative mating, as
1030 many of these putative incompatibilities were ascertained by detecting pairs of alleles
1031 with ancestry disequilibrium in naturally admixed populations [66]. Genes identified by
1032 these two approaches may constitute alleles involved in female mate choice, including
1033 *eag,* which has been shown to directly influence courtship behavior, *DAAM*, which is
1034 involved in mushroom body development, and *rbp6* which influences male aggressive

1035 behavior. Second, while *PBE* analyses are an effective approach to detect branch
1036 specific evolution for a focal population [5,59], by definition, these statistics cannot
1037 differentiate branch specific evolution which is specific to a trait of interest versus subtle
1038 population structure and/or branch specific evolution of non-focal traits. Therefore, if
1039 some incompatibility alleles exist at higher frequencies in isolines with strong female
1040 choice relative to isolines with no strong female mate choice, they too may be included
1041 in our *PBE* outlier analysis (even though these alleles may not directly influence female
1042 mate choice). Regardless, our results suggest that putative incompatibility alleles do
1043 show some genetic structure across the global distribution of *D. melanogaster*, and
1044 better understanding the function of these alleles can better our understanding of both
1045 reproductive isolation and barriers to gene flow within this model system.

1046

1047 **CONCLUSION**

1048 Our work contributes to our understanding of the evolutionary history of *D.*
1049 *melanogaster*, including the distribution of genetic diversity and putative reproductive
1050 isolation within this model organism. We combined population genomics to assess how
1051 genetic and phenotypic diversity is partitioned across the globe in *D. melanogaster*,
1052 especially within its putative ancestral range in Subtropical Africa. While we find
1053 significant population genetic structuring throughout the range, we also find that
1054 polymorphic reproductive barriers are largely decoupled from broader patterns of
1055 population genetic structure in this system. Furthermore, this work provides insights into
1056 the complexity of genetic changes underlying female mate choice, and highlights many

genes of interest for future study. In total, our work contributes to our understanding of the factors influencing how genetic diversity is structured across the range in a model system.

REFERENCES

1. Morgan TH. 1911. Random segregation versus coupling in Mendelian inheritance. *Science* **34**, 384–384.

2. Morgan TH. 1910. Sex limited inheritance in *Drosophila*. *Science* **32**, 120–122.

3. Green MM. 2010 2010: A Century of Drosophila Genetics Through the Prism of the white Gene. *Genetics* **184**, 3–7.

4. Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehermann G, Becher PG, Pool JE, Stensmyr MC. 2018 Wild African *Drosophila melanogaster* are seasonal specialists on Marula fruit. *Curr. Biol.* **28**, 3960–3968.e3.

5. Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC, Pool JE, Larracuente A. 2020 Recurrent collection of *Drosophila melanogaster* from wild African environments and genomic insights into species history. *Mol. Biol. Evol.* **37**, 627–638.

6. Kopp A, Duncan I, Carroll SB. 2000 Genetic control and evolution of sexually dimorphic characters in Drosophila. *Nature* **408**, 553–559.

7. McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, Payre F, Stern DL. 2007 Morphological evolution through multiple cis -regulatory mutations at a single gene. *Nature* **448**, 587–590.

8. Montooth KL, Marden JH, Clark AG. 2003 Mapping determinants of variation in energy metabolism, respiration and flight in *Drosophila*. *Genetics* **165**, 623–635.

9. Ding Y, Berrocal A, Morita T, Longden KD, Stern DL. 2016 Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature* **536**, 329–332.

10. Markow TA. 2015 The secret lives of *Drosophila* flies. *Elife* **4**, e06793.

11. Harshman LG, Hoffmann AA. 2000 Laboratory selection experiments using *Drosophila*: what do they really tell us? *Trends Ecol. Evol.* **15**, 32–36.

1087 12. White NJ, Snook RR, Eyres I. 2020 The Past and future of experimental speciation.
1088 *Trends Ecol. Evol.* **35**, 10–21.

1089 13. Begun DJ, Aquadro CF. 1993 African and North American populations of
1090 *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550.

1091 14. Dieringer D, Nolte V, Schlötterer C. 2005 Population structure in African *Drosophila*
1092 *melanogaster* revealed by microsatellite analysis. *Mol. Ecol.* **14**, 563–573.

1093 15. Pool JE, Aquadro CF. 2006 History and structure of sub-Saharan populations of
1094 *Drosophila melanogaster*. *Genetics* **174**, 915–929.

1095 16. Pool JE *et al.* 2012 Population Genomics of Sub-Saharan *Drosophila*
1096 *melanogaster*: African Diversity and Non-African Admixture. *PLoS Genet.* **8**.
1097 (doi:10.1371/journal.pgen.1003080)

1098 17. Schlötterer C, Neumeier H, Sousa C, Nolte V. 2006 Highly Structured Asian
1099 *Drosophila melanogaster* Populations: A New Tool for Hitchhiking Mapping?
1100 *Genetics* **172**, 287–292.

1101 18. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA,
1102 Langley CH, Pool JE. 2015 The *Drosophila* genome nexus: a population genomic
1103 resource of 623 *Drosophila melanogaster* genomes, including 197 from a single
1104 ancestral range population. *Genetics* **199**, 1229–1241.

1105 19. Adrion JR, Hahn MW, Cooper BS. 2015 Revisiting classic clines in *Drosophila*
1106 *melanogaster* in the age of genomics. *Trends Genet.* **31**, 434–444.

1107 20. Grenier JK, Roman Arguello J, Moreira MC, Gottipati S, Mohammed J, Hackett SR,
1108 Boughton R, Greenberg AJ, Clark AG. 2015 Global diversity lines-a five-continent
1109 reference panel of sequenced *Drosophila melanogaster* strains. *G3: Genes,*
1110 *Genomes, Genetics* **5**, 593–603.

1111 21. Baudry E, Viginier B, Veuille M. 2004 Non-African populations of *Drosophila*
1112 *melanogaster* have a unique origin. *Mol. Biol. Evol.* **21**, 1482–1491.

1113 22. Arguello JR, Laurent S, Clark AG, Gaut B. 2019 Demographic History of the
1114 Human Commensal *Drosophila melanogaster*. *Genome Biol. Evol.* **11**, 844–854.

1115 23. Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005 Multilocus patterns
1116 of nucleotide variability and the demographic and selection history of *Drosophila*
1117 *melanogaster* populations. *Genome Res.* **15**, 790–799.

1118 24. Thornton K, Andolfatto P. 2006 Approximate Bayesian inference reveals evidence
1119 for a recent, severe bottleneck in a Netherlands population of *Drosophila*
1120 *melanogaster*. *Genetics* **172**, 1607–1619.

1121 25. Kauer M, Dieringer D, Schlötterer C. 2003 Nonneutral admixture of immigrant

1122 genotypes in African *Drosophila melanogaster* populations from Zimbabwe. *Mol.*
1123 *Biol. Evol.* **20**, 1329–1337.

1124 26. Capy P, Veuille M, Paillette M, Jallon JM, Vouidibio J, David JR. 2000 Sexual
1125 isolation of genetically differentiated sympatric populations of *Drosophila*
1126 *melanogaster* in Brazzaville, Congo: The first step towards speciation? *Heredity*
1127 **84**, 468–475.

1128 27. Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D. 2015 Population genomic
1129 analysis uncovers African and European admixture in *Drosophila melanogaster*
1130 populations from the south-eastern United States and Caribbean Islands. *Mol. Ecol.*
1131 **24**, 1499–1509.

1132 28. Yukilevich R, True JR. 2008 Incipient sexual isolation among cosmopolitan
1133 *Drosophila melanogaster* populations. *Evolution* **62**, 2112–2121.

1134 29. Caracristi G, Schlötterer C. 2003 Genetic differentiation between American and
1135 European *Drosophila melanogaster* populations could be attributed to admixture of
1136 African alleles. *Mol. Biol. Evol.* **20**, 792–799.

1137 30. Hollocher H, Ting CT, Wu ML, Wu CI. 1997 Incipient speciation by sexual isolation
1138 in *Drosophila melanogaster*: Extensive genetic divergence without reinforcement.
1139 *Genetics* **147**, 1191–1201.

1140 31. Ting C-T, Takahashi A, Wu C-I. 2001 Incipient speciation by sexual isolation in
1141 *Drosophila*: Concurrent evolution at multiple loci. *Proc. Natl. Acad. Sci. U. S. A.* **98**,
1142 6709–6713.

1143 32. Wu CI, Hollocher H, Begun DJ, Aquadro CF, Xu Y, Wu ML. 1995 Sexual isolation
1144 in *Drosophila melanogaster*: a possible case of incipient speciation. *Proc. Natl.*
1145 *Acad. Sci. U. S. A.* **92**, 2519–2523.

1146 33. Korol A, Rashkovetsky E, Iliadi K, Michalak P, Ronin Y, Nevo E. 2000 Nonrandom
1147 mating in *Drosophila melanogaster* laboratory populations derived from closely
1148 adjacent ecologically contrasting slopes at 'Evolution Canyon'. *Proc. Natl. Acad.*
1149 *Sci. U. S. A.* **97**, 12637–12642.

1150 34. Hollocher H, Ting C-T, Pollack F, Wu C-I. 1997 Incipient Speciation by Sexual
1151 Isolation in *Drosophila melanogaster*: Variation in Mating Preference and
1152 Correlation Between Sexes. *Evolution* **51**, 1175.

1153 35. Sawamura K, Tomaru M. 2002 Biology of reproductive isolation in *Drosophila*:
1154 toward a better understanding of speciation. *Popul. Ecol.* **44**, 0209–0219.

1155 36. Barbash DA. 2010 Ninety Years of *Drosophila melanogaster* Hybrids. *Genetics*
1156 **186**, 1–8.

1157 37. Jiggins CD, Mallet J. 2000 Bimodal hybrid zones and speciation. *Trends Ecol. Evol.*

1158 **15**, 250–255.

1159 38. Irwin DE. 2020 Assortative Mating in Hybrid Zones Is Remarkably Ineffective in
1160 Promoting Speciation. *Am. Nat.* **195**, E150–E167.

1161 39. Coughlan JM, Matute DR. 2020 The importance of intrinsic postzygotic barriers
1162 throughout the speciation process. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**,
1163 20190533.

1164 40. Pool JE. 2015 The mosaic ancestry of the drosophila genetic reference panel and
1165 the *D. melanogaster* reference genome reveals a network of epistatic fitness
1166 interactions. *Mol. Biol. Evol.* **32**, 3236–3251.

1167 41. Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF. 2013 Genetic
1168 incompatibilities are widespread within species. *Nature* **504**, 135–137.

1169 42. dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J,
1170 Emmert DB, Gelbart WM, FlyBase Consortium. 2015 FlyBase: introduction of the
1171 *Drosophila melanogaster* Release 6 reference genome assembly and large-scale
1172 migration of genome annotations. *Nucleic Acids Res.* **43**, D690–7.

1173 43. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows–
1174 Wheeler transform. *Bioinformatics* **25**, 1754–1760.

1175 44. McKenna A *et al.* 2010 The Genome Analysis Toolkit: a MapReduce framework for
1176 analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.

1177 45. Li H *et al.* 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
1178 **25**, 2078–2079.

1179 46. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018 UFBoot2:
1180 Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522.

1181 47. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017
1182 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat.*
1183 *Methods* **14**, 587–589.

1184 48. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and
1185 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol.*
1186 *Biol. Evol.* **32**, 268–274.

1187 49. Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018 ASTRAL-III: polynomial time
1188 species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*
1189 **19**, 153.

1190 50. Meisner J, Albrechtsen A. 2018 Inferring population structure and admixture
1191 proportions in low-depth NGS data. *Genetics* **210**, 719–731.

1192 51. Korunes KL, Samuk K. 2021 pixy: Unbiased estimation of nucleotide diversity and
1193 divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368.

1194 52. Danecek P *et al.* 2011 The variant call format and VCFtools. *Bioinformatics* **27**,
1195 2156–2158.

1196 53. Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N. 2013 SweeD: likelihood-based
1197 detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–
1198 2234.

1199 54. Martin SH, Davey JW, Jiggins CD. 2015 Evaluating the use of ABBA–BABA
1200 statistics to locate introgressed loci. *Mol. Biol.*

1201 55. Malinsky M, Matschiner M, Svardal H. 2021 Dsuite - Fast D-statistics and related
1202 admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595.

1203 56. Martin SH, Van Belleghem SM. 2017 Exploring Evolutionary Relationships Across
1204 the Genome Using Topology Weighting. *Genetics* **206**, 429–438.

1205 57. Fox J, Weisberg S. 2018 *An R Companion to Applied Regression*. SAGE
1206 Publications.

1207 58. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B,
1208 Scheipl F. 2012 Package 'lme4'. *CRAN. R Foundation for Statistical Computing,*
1209 *Vienna, Austria*

1210 59. Yassin A, Debat V, Bastide H, Gidaszewski N, David JR, Pool JE. 2016 Recurrent
1211 specialization on a toxic fruit in an island *Drosophila* population. *Proc. Natl. Acad.*
1212 *Sci. U. S. A.* **113**, 4771–4776.

1213 60. Yi X *et al.* 2010 Sequencing of 50 human exomes reveals adaptation to high
1214 altitude. *Science* **329**, 75–78.

1215 61. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K,
1216 Muruganujan A, Narechania A. 2003 PANTHER: A Library of Protein Families and
1217 Subfamilies Indexed by Function. *Genome Res.* **13**, 2129–2141.

1218 62. Greenspan RJ, Ferveur J-F. 2000 COURTSHIP IN DROSOPHILA. *Annu. Rev.*
1219 *Genet.* **34**, 205–232.

1220 63. Russell L. 2018 emmeans: Estimated marginal means, aka Least-Squares Means.
1221 R package version 1.3. 0.

1222 64. Müller, H. 1942 Isolating mechanisms, evolution, and temperature. *Biol. Symp.* **6**,
1223 71–125.

1224 65. Dobzhansky, T. (1937). Genetics and the Origin of Species. New York: Columbia
1225 University Press.

1226 66. Schumer M, Brandvain Y. 2016 Determining epistatic selection in admixed
1227 populations. *Mol. Ecol.* **25**, 2577–2591.

1228 67. Veller C, Edelman NB, Muralidhar P, Nowak MA. 2019 Recombination, variance in
1229 genetic relatedness, and selection against introgressed DNA. *BioRxiv*, 846147.
1230 (doi:10.1101/846147)

1231 68. Rouault J-D, Marican C, Wicker-Thomas C, Jallon J-M. 2004 Relations between
1232 cuticular hydrocarbon (HC) polymorphism, resistance against desiccation and
1233 breeding temperature; a model for HC evolution in *D. melanogaster* and *D.*
1234 *simulans*. *Genetica* **120**, 195-212.

1235 69. Foley BR, Telonis-Scott M. 2011 Quantitative genetic analysis suggests causal
1236 association between cuticular hydrocarbon composition and desiccation survival in
1237 *Drosophila melanogaster*. *Heredity* **106**, 68–77.

1238 70. Carhan A, Allen F, Armstrong JD, Hortsch M, Goodwin SF, O'Dell KMC. 2005
1239 Female receptivity phenotype of icebox mutants caused by a mutation in the L1-
1240 type cell adhesion molecule neuroglian. *Genes Brain Behav.* **4**, 449–465.

1241 71. Charles BS. 2019 Neural Regions Affecting Female Aggression and Receptivity in
1242 *Drosophila melanogaster*. Western University Thesis and Dissertation Repository.

1243 72. Tajima F. 1989 Statistical method for testing the neutral mutation hypothesis by
1244 DNA polymorphism. *Genetics* **123**, 585–595.

1245 73. Lack JB, Yassin A, Sprengelmeyer QD, Johanning EJ, David JR, Pool JE. 2016
1246 Life history evolution and cellular mechanisms associated with increased size in
1247 high- altitude *Drosophila*. *Ecol. Evol.* **6**, 5893–5906.

1248 74. Bastide H, Lange JD, Lack JB, Yassin A, Pool JE. 2016 A Variable Genetic
1249 Architecture of Melanic Evolution in *Drosophila melanogaster*. *Genetics* **204**, 1307–
1250 1319.

1251 75. Lack JB, Monette MJ, Johanning EJ, Sprengelmeyer QD, Pool JE. 2016
1252 Decanalization of wing development accompanied the evolution of large wings in
1253 high-altitude *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1014–1019.

1254 76. Payseur BA, Krenz JG, Nachman MW. 2004 Differential Patterns of Introgression
1255 Across the X Chromosome in a Hybrid Zone Between Two Species of House Mice.
1256 *Evolution* **58**, 2064.

1257 77. Geraldes A, Carneiro M, Delibes-Mateos M, Villafuerte R, Nachman MW, Ferrand
1258 N. 2008 Reduced introgression of the Y chromosome between subspecies of the
1259 European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Mol. Ecol.* **17**,
1260 4489–4499.

1261 78. Cortés-Ortiz L, Nidiffer MD, Hermida-Lagunes J, García-Orduña F, Rangel-Negrín

1262 A, Kitchen DM, Bergman TJ, Dias PAD, Canales-Espinosa D. 2019 Reduced
1263 Introgression of Sex Chromosome Markers in the Mexican Howler Monkey
1264 (*Alouatta palliata* × *A. pigra*) Hybrid Zone. *Int. J. Primatol.* **40**, 114–131.

1265 79. Turissini DA, Matute DR. 2017 Fine scale mapping of genomic introgressions within
1266 the *Drosophila yakuba* clade. *PLoS Genet.* **13**, e1006971.

1267 80. Garrigan D, Kingan SB, Geneva AJ, Andolfatto P, Clark AG, Thornton KR,
1268 Presgraves DC. 2012 Genome sequencing reveals complex speciation in the
1269 *Drosophila simulans* clade. *Genome Res.* **22**, 1499–1511.

1270 81. Fraïsse C, Sachdeva H. 2020 The rates of introgression and barriers to genetic
1271 exchange between hybridizing species: sex chromosomes vs autosomes. *Genetics*
1272 **217**. (doi:10.1093/genetics/iyaa025)

1273 82. Dixon G, Kitano J, Kirkpatrick M. 2018 The Origin of a New Sex Chromosome by
1274 Introgression between Two Stickleback Fishes. *Mol. Biol. Evol.* **36**, 28–38.

1275 83. Dufresnes C, Majtyka T, Baird SJE, Gerchen JF, Borzée A, Savary R, Ogielska M,
1276 Perrin N, Stöck M. 2016 Empirical evidence for large X-effects in animals with
1277 undifferentiated sex chromosomes. *Sci. Rep.* **6**, 1–7.

1278 84. Macholán M, Baird SJE, Munclinger P, Dufková P, Bímová B, Piálek J. 2008
1279 Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone?
1280 *BMC Evol. Biol.* **8**, 271.

1281 85. Hartmann FE, Rodríguez de la Vega RC, Gladieux P, Ma W-J, Hood ME, Giraud T.
1282 2020 Higher Gene Flow in Sex-Related Chromosomes than in Autosomes during
1283 Fungal Divergence. *Mol. Biol. Evol.* **37**, 668–682.

1284 86. Ottenburghs J. 2021 The genic view of hybridization in the Anthropocene. *Evol.*
1285 *Appl.* (doi:10.1111/eva.13223)

1286 87. Allendorf FW, Leary RF, Spruell P, Wenburg JK. 2001 The problems with hybrids:
1287 setting conservation guidelines. *Trends Ecol. Evol.* **16**, 613–622.

1288 88. Hulme-Beaman A, Dobney K, Cucchi T, Searle JB. 2016 An Ecological and
1289 Evolutionary Framework for Commensalism in Anthropogenic Environments.
1290 *Trends Ecol. Evol.* **31**, 633–645.

1291 89. Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlotterer C.
1292 2004 World-wide survey of an Accord insertion and its association with DDT
1293 resistance in Drosophila melanogaster. *Mol. Ecol.* **13**, 2491–2504.

1294 90. Daborn PJ *et al.* 2002 A single p450 allele associated with insecticide resistance in
1295 Drosophila. *Science* **297**, 2253–2256.

1296 91. Ravinet M, Elgvin TO, Trier C, Aliabadian M, Gavrilov A, Sætre G-P. 2018

1297 Signatures of human-commensalism in the house sparrow genome. *Proc. Biol. Sci.*
1298 **285**, 20181246.

1299 92. Guiry E, Buckley M. 2018 Urban rats have less variable, higher protein diets. *Proc.*
1300 *Biol. Sci.* **285**. (doi:10.1098/rspb.2018.1441)

1301 93. Harris SE, O'Neill RJ, Munshi-South J. 2015 Transcriptome resources for the white-
1302 footed mouse (*Peromyscus leucopus*): new genomic tools for investigating
1303 ecologically divergent urban and rural populations. *Mol. Ecol. Resour.* **15**, 382–394.

1304 94. Axelsson E *et al.* 2013 The genomic signature of dog domestication reveals
1305 adaptation to a starch-rich diet. *Nature* **495**, 360–364.

1306 95. Harris SE, Munshi-South J, Obergfell C, O'Neill R. 2013 Signatures of rapid
1307 evolution in urban and rural transcriptomes of white-footed mice (*Peromyscus*
1308 *leucopus*) in the New York metropolitan area. *PLoS One* **8**, e74938.

1309 96. Harpak A, Garud N, Rosenberg NA, Petrov DA, Combs M, Pennings PS, Munshi-
1310 South J. 2020 Genetic Adaptation in New York City Rats. *Genome Biol. Evol.* **13**.
1311 (doi:10.1093/gbe/evaa247)

1312 97. Grillet M, Everaerts C, Houot B, Ritchie MG, Cobb M, Ferveur J-F. 2012 Incipient
1313 speciation in *Drosophila melanogaster* involves chemical signals. *Sci. Rep.* **2**, 1–11.

1314 98. Grillet M, Ferveur JF, Everaerts C. 2018 Behavioural elements and sensory cues
1315 involved in sexual isolation between *Drosophila melanogaster* strains. *Royal*
1316 *Society Open Science* **5**. (doi:10.1098/rsos.172060)

1317 99. Jin B, Barbash DA, Castillo DM. 2020 Divergent selection on behavioral and
1318 chemical traits contributes to isolation between populations of *Drosophila*
1319 *melanogaster*. *BioRxiv*, 2020.09.28.316703. (doi:10.1101/2020.09.28.316703)

1320 100. Dembeck LM, Böröczky K, Huang W, Schal C, Anholt RRH, Mackay TFC. 2015
1321 Genetic architecture of natural variation in cuticular hydrocarbon composition in
1322 *Drosophila melanogaster*. *Elife* **4**, e09861.

1323 101. McBride SM, Giuliani G, Choi C, Krause P, Correale D, Watson K, Baker G, Siwicki
1324 KK. 1999 Mushroom body ablation impairs short-term memory and long-term
1325 memory of courtship conditioning in *Drosophila melanogaster*. *Neuron* **24**, 967–
1326 977.

1327 102. Coyne JA, Wicker-Thomas C, Jallon JM. 1999 A gene responsible for a cuticular
1328 hydrocarbon polymorphism in *Drosophila melanogaster*. *Genet. Res.* **73**, 189–203.

1329 103. Combs PA, Krupp JJ, Khosla NM, Bua D, Petrov DA, Levine JD, Fraser HB. 2018
1330 Tissue-Specific cis-Regulatory Divergence Implicates eloF in Inhibiting Interspecies
1331 Mating in *Drosophila*. *Curr. Biol.* **28**, 3969–3975.e3.

104. Chen D, Qu C, Bjorum SM, Beckingham KM, Hewes RS. 2014 Neuronal Remodeling During Metamorphosis Is Regulated by the alan shepard (shep) Gene in *Drosophila melanogaster*. *Genetics* **197**, 1267–1283.

105. Griffith LC, Wang J, Zhong Y, Wu CF, Greenspan RJ. 1994 Calcium/calmodulin-dependent protein kinase II and potassium channel subunit eag similarly affect plasticity in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10044–10048.

1352 **SUPPLEMENTAL MATERIALS**

1353 **Table S1: Sample information for all individuals in this study.** For each accession,
1354 the unique ID (sample), species (each are species names of the genus *Drosophila*),
1355 NCBI SRA code (new samples will be given a unique code upon data upload), the
1356 sequencing platform from which the data came, the geo-genetic lineage which the
1357 sample belongs to, collection location information, and the average depth of coverage
1358 are given.

1359 **Table S2:** F_{ST} (blue; lower triangle), D_{xy} (green; upper triangle), and π (diagonal) for all
1360 geo-genetic lineages of *D. melanogaster* and *D. simulans*.

1361 **Table S3:** *D* statistics of focal comparisons. Population codes: ET= Ethiopia, East= East
1362 Africa, West= West Africa, SE-US= SE United States, CEN= Central Africa, HD= Harare
1363 Distinct, OOA= Out of Africa (excluding SE North America).

1364 **Table S4:** Mean f_{dM} values for each chromosome arm by each focal comparison.
1365 Significant differences between chromosome arms for each comparison are indicated
1366 by superscript letters.

1367 **Table S5: Introgression Outliers.** Top 1% of f_{dM} outliers for each of three
1368 comparisons: West Africa and Central Africa (WestCen), HD and OOA (OOAHD), and
1369 West Africa and the SE United States (WestSena). Genes within the outlier window and
1370 those proximal (as well as their proximity in KB) are given.

1371 **Table S6: Summary of behavioral survey.** Average proportion of Z males chosen
1372 across trials for each isoline phenotyped in our behavioral survey, as well as the total
1373 trials set up, the total matings recorded, and the proportion of successful trials (mating
1374 proportion). We then defined each line as a behavioral type (no preference= NP;
1375 preference= P) based on both a Fisher's Exact and binomial test (these analyses
1376 agreed completely on type definition with a significance cutoff of $p<0.05$).

1377 **Table S7: Behavioral *PBE* outliers.** Top 1% of *PBE* outliers with Central African lines
1378 that show significant mate preference as the focal population, and Central African lines
1379 with no strong mate preference and RAL lines as the non-focal populations. Names of
1380 genes directly within the outlier window, as well as names of genes in close proximity
1381 are listed for each window, along with their proximity in KB up/downstream.

1382 **Table S8: Behavioral *PBE* GO analyses.** GO analyses of the genes in or proximal to
1383 *PBE* outliers. Performed using *PANTHER* v.16.0.

1384

1385 **Table S9: Putative incompatibilities with F_{st} outliers.** A list of putative
1386 incompatibilities identified by [40] and [41] in which both interacting loci show signatures
1387 of older or more recent derivation (denoted in the Popgen signature). Window positions
1388 were translated between V.5 and V.632 of the *D. melanogaster* reference genome, as
1389 different reference genomes were used between [40] and the current study. As some
1390 windows identified by pool contain multiple F_{ST} outliers, the number of outliers contained
1391 for each locus window, plus the genes that fall within that outlier are listed.

1392

1393

1394

1395

1396

1397

1398

1399

1400

1401

1402

1403

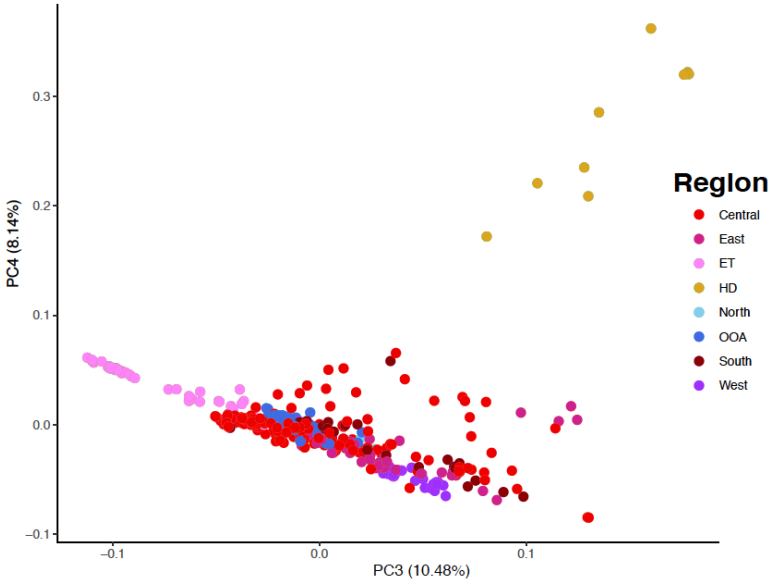


Figure S1: PCs 3 and 4 of a genome-wide PCA. *Points are colored based on the geo-genetic lineage. Percent of variance explained by each PC is indicated in brackets.*

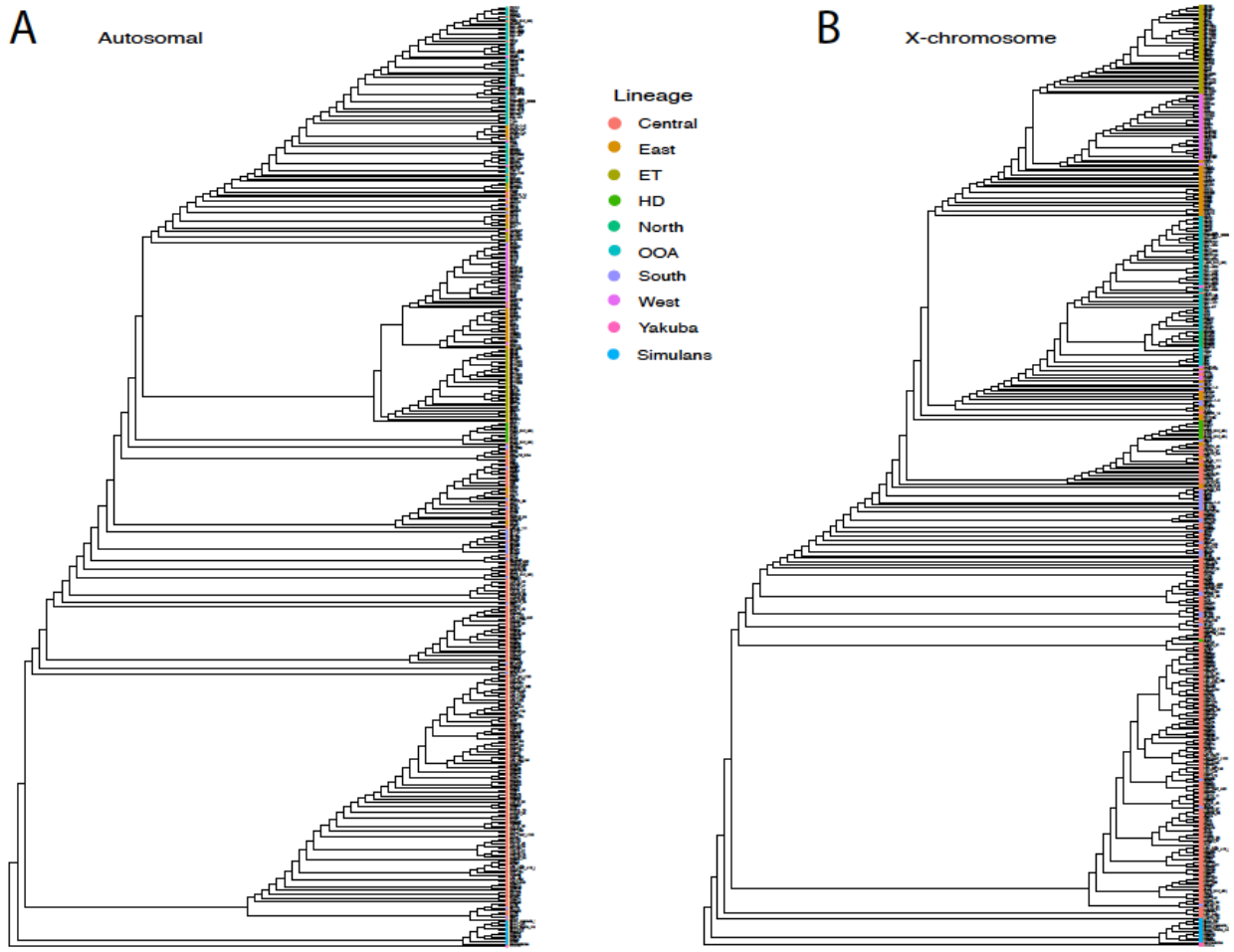


Figure S2: Full ML phylogeny of all samples based on (A) autosomes and (B) the *X* chromosome only. Tip colors correspond with the geo-genetic lineages defined herein, plus *D. simulans* and *D. yakuba* as outgroups.

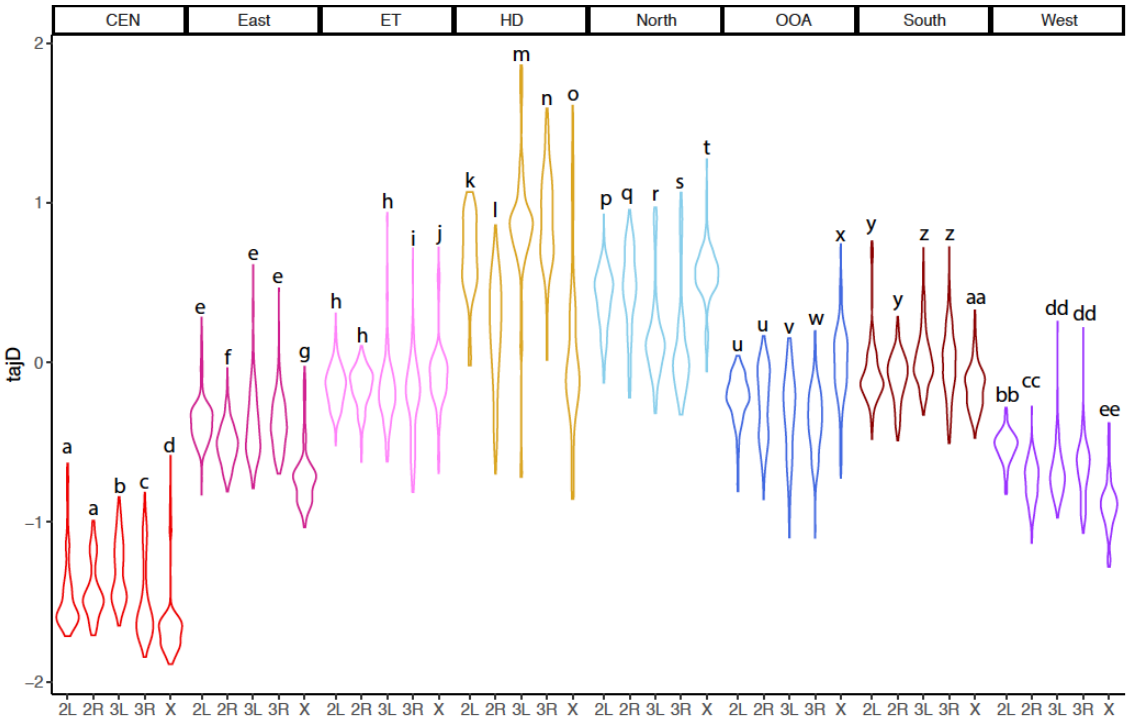


Figure S3: Tajima's D by chromosome for each geo-genetic lineage. Letters denote significantly different groups.

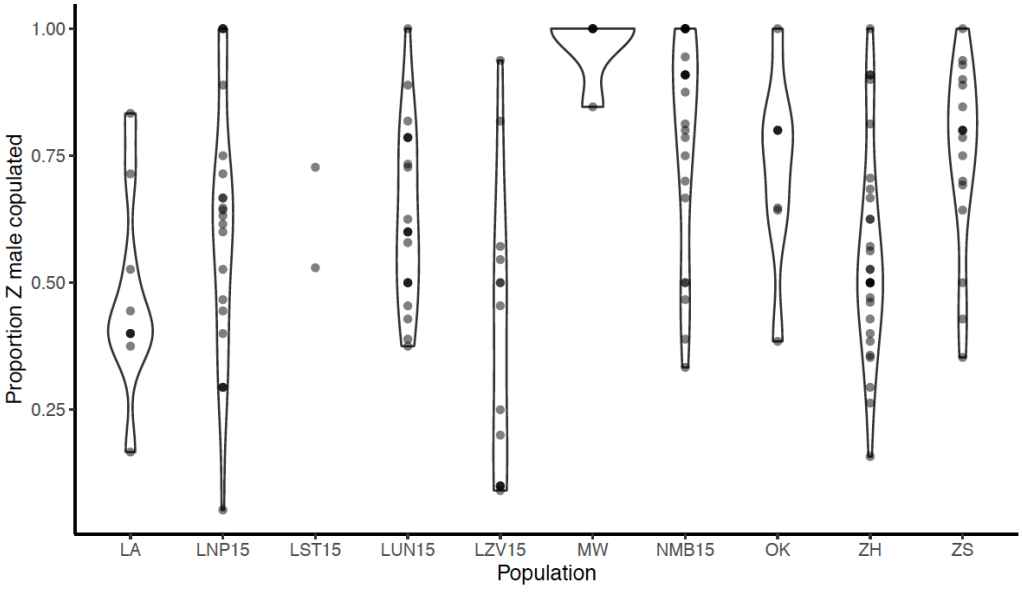


Figure S4: Strength of female mate preference differs by population. Proportion of Z males copulated in a two-way choice trial between Z and M males and female flies from 10 sampling locales in Subtropical Africa. Points represent individual trials for individual isolines.

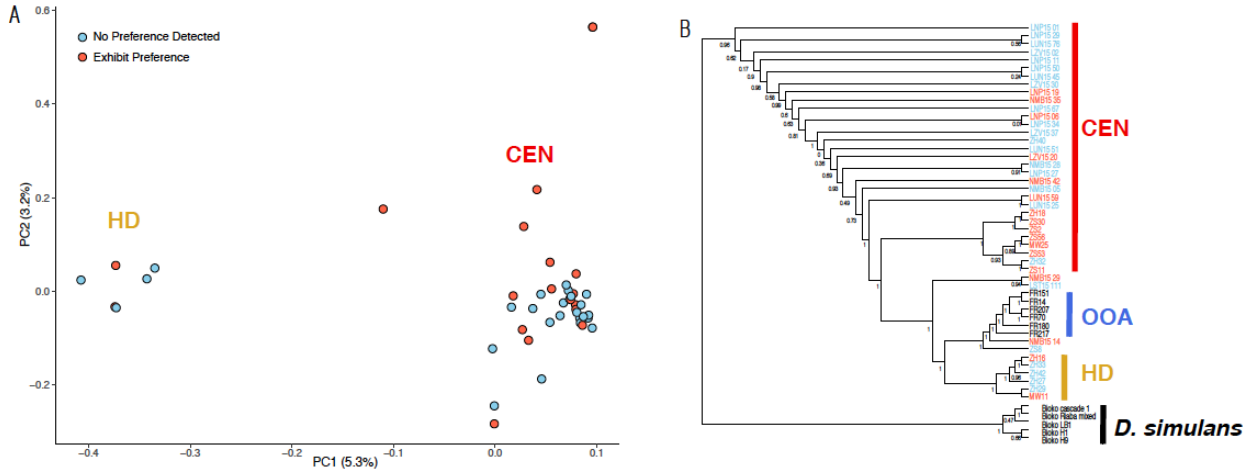
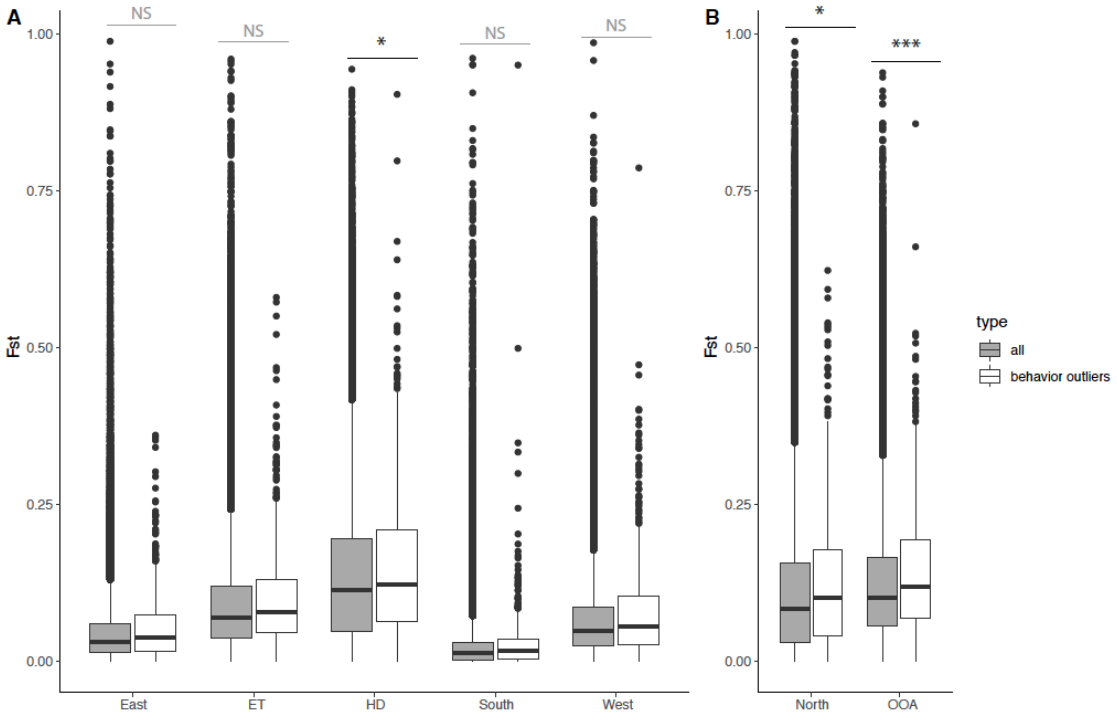


Figure S5: No genetic structure between lines with or without strong mate preference; *denoted in red and blue respectively. (A) PCA of all Central African lines that were phenotype for mate preference. Substantial genetic structure between HD and Central clades, but no structuring between lines that do/do not show significant female mate preference. Moreover, there is no genetic structure between behavioral types across the first 8 PCs. Numbers in brackets denote the percent of variance explained by PC1 and PC2, respectively. (B) ML consensus phylogeny built from non-overlapping regions of 100KB using the rapid-hill climbing algorithm in RAxML, with 50 bootstraps per tree. Consensus phylogeny was built using these trees as input to ASTRAL. Each genetic clade is outlined with a thick vertical bar, with colors replicating the represented clades in Figure 1. We also include six lines from France (denoted by the blue OOA section) for context. We note that only 17 of 19 individuals with strong preference and have sequenced genomes are included in the phylogeny as 2 genomes did not pass quality thresholds for our phylogenetic analyses (LA66 and LA69).*



1446

1447 **Figure S6: Behavioral outliers are highly differentiated between Central Africa and**
1448 **other geo-genetic lineages.** *PBE outliers for behavior show increased F_{st} relative to*
1449 *the rest of the genome for all comparisons (genome-wide versus outlier: F=7.5732,*
1450 *df=1, p=0.0059). However, when contrasting F_{st} between Central Africa and specific*
1451 *geo-genetic lineages, differences between F_{st} for genome-wide versus PBE behavior*
1452 *outliers is only significant for Central Africa versus OOA, North Africa, and HD.*

1453

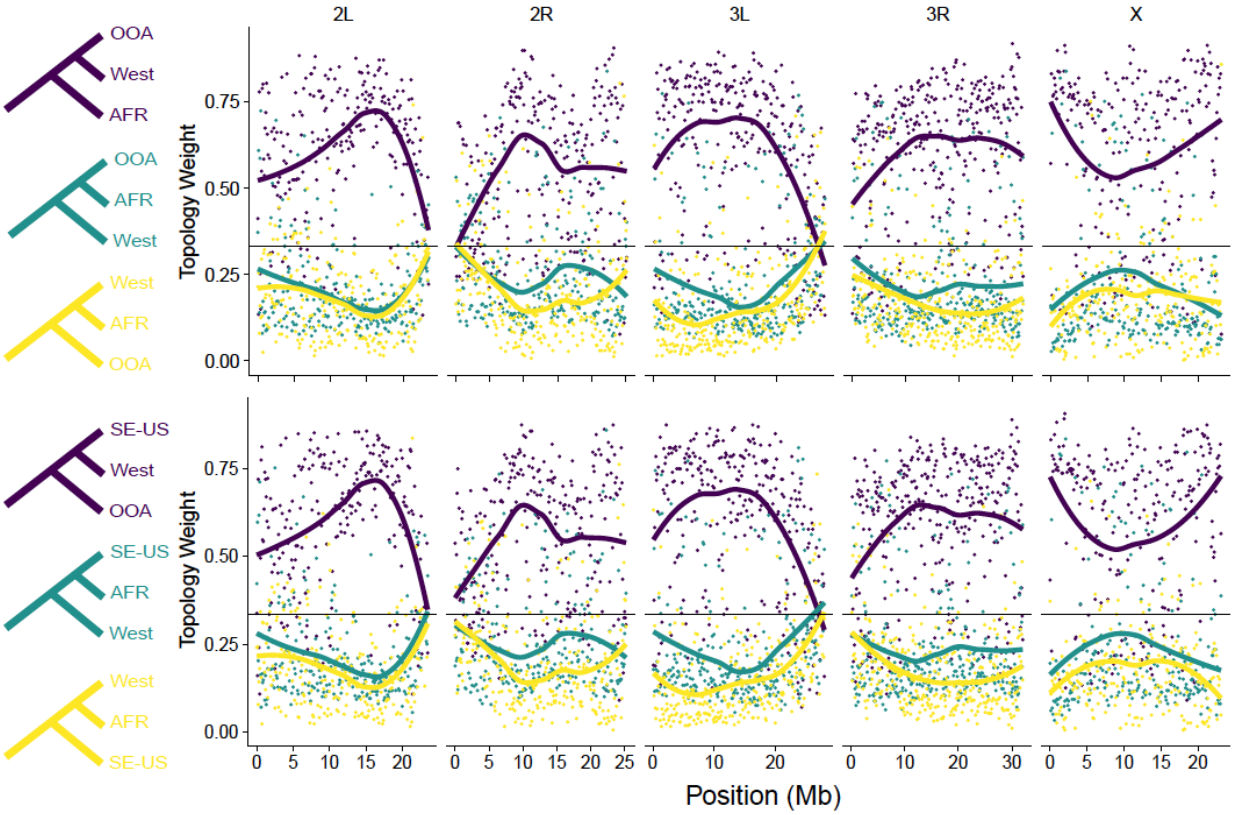


Figure S7: Weighted topologies between West Africa and OOA. *Highly similar landscapes of weighted topologies between West Africa and each of all OOA lines and just those from the SE-United States (SE-US) suggest a largely shared landscape of introgression.*

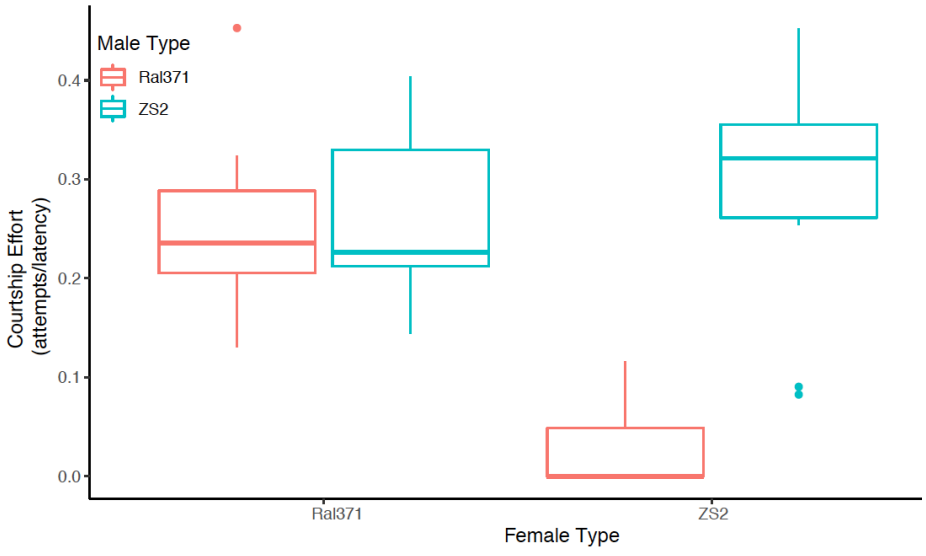
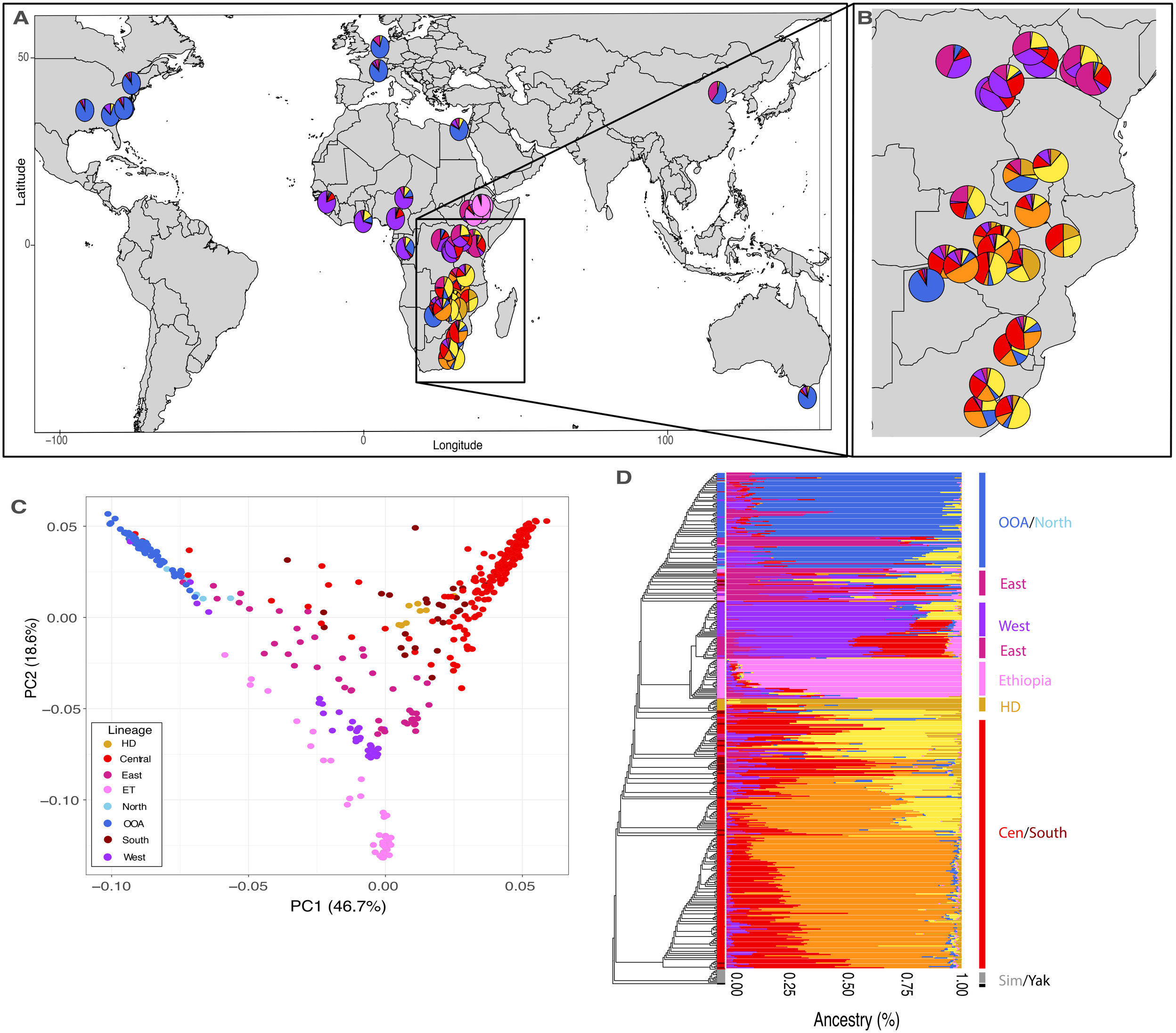
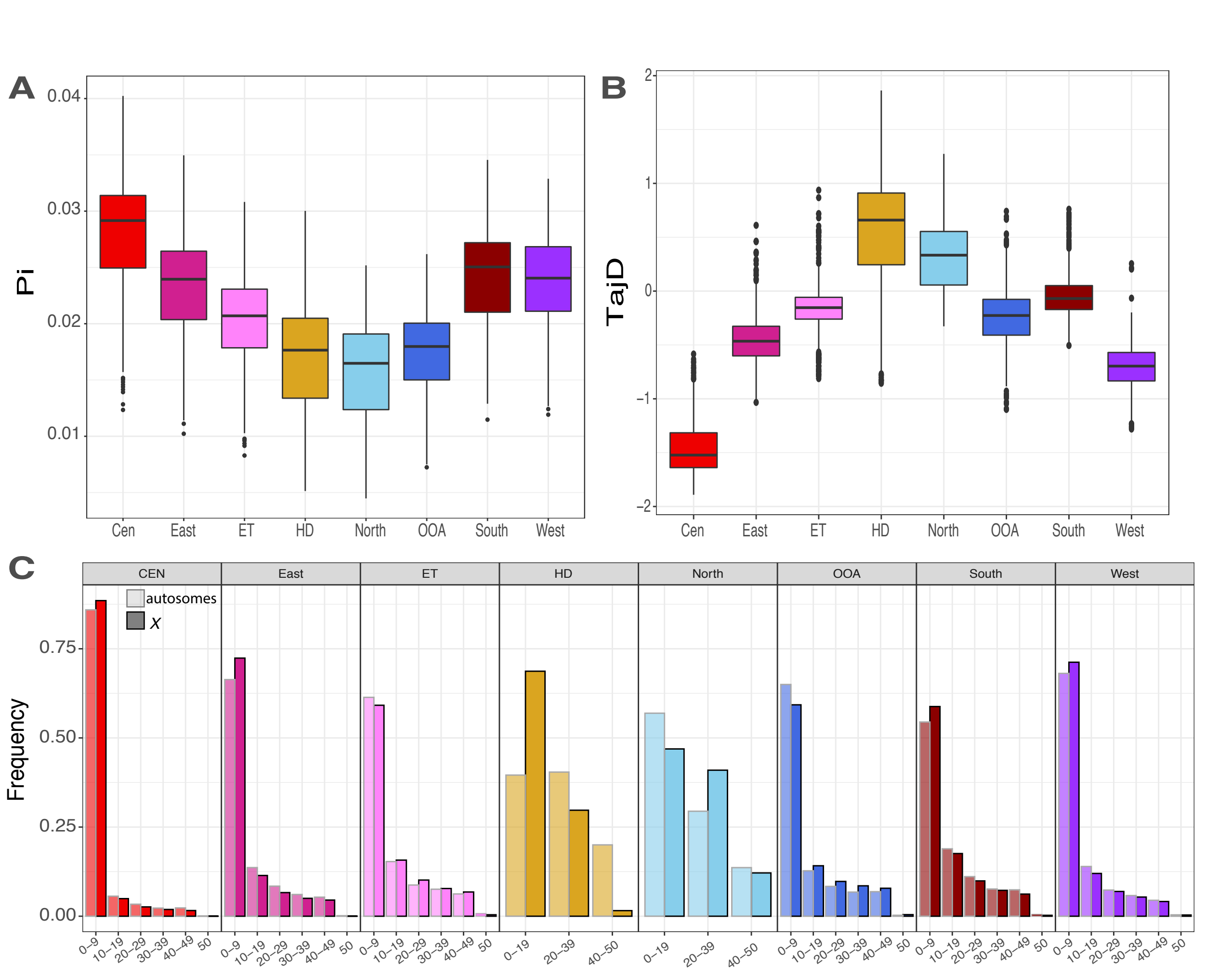
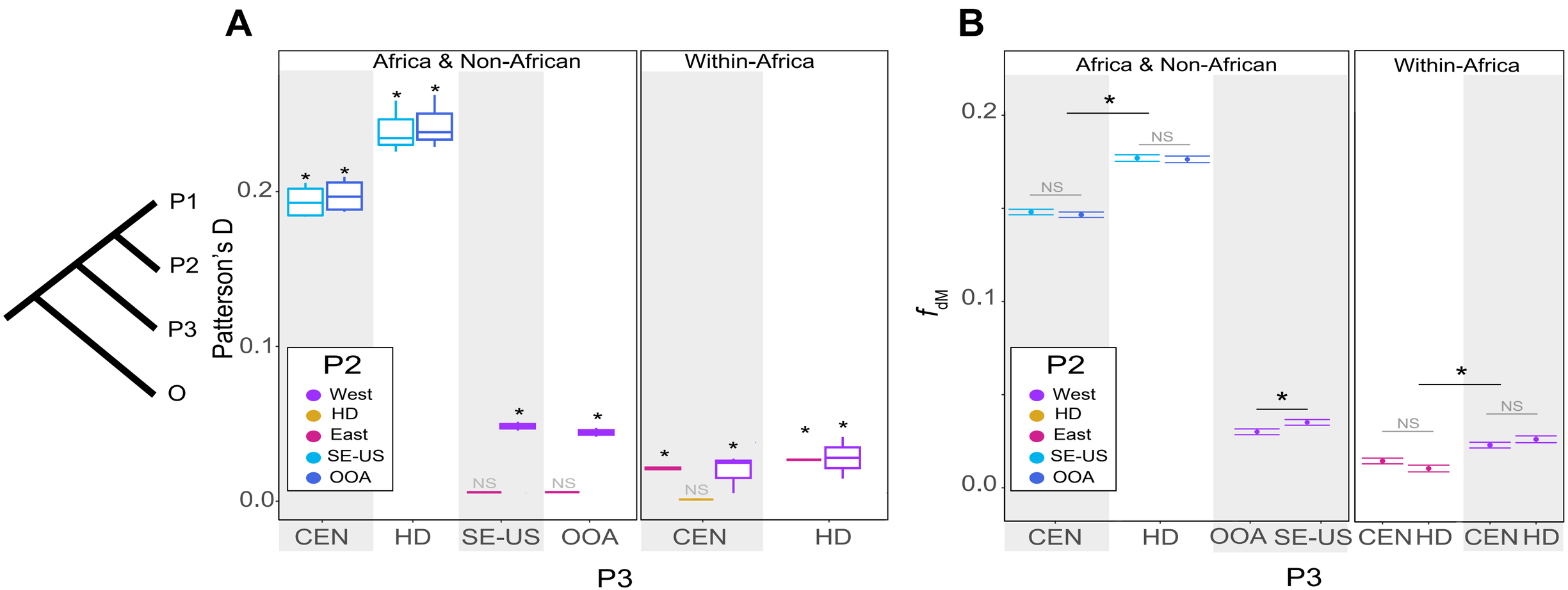
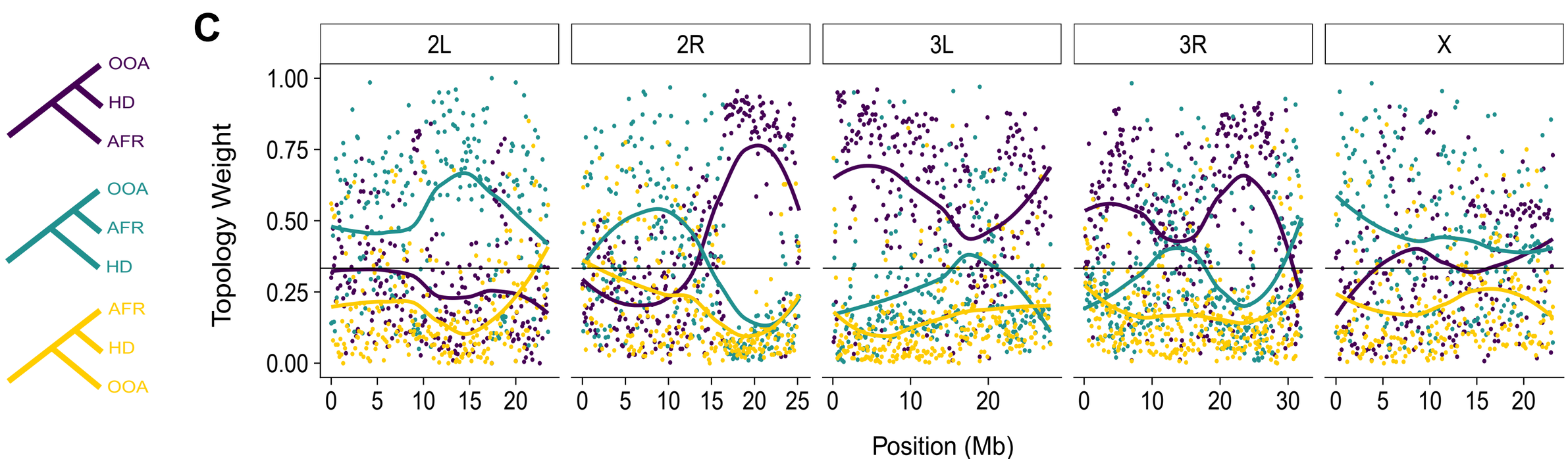


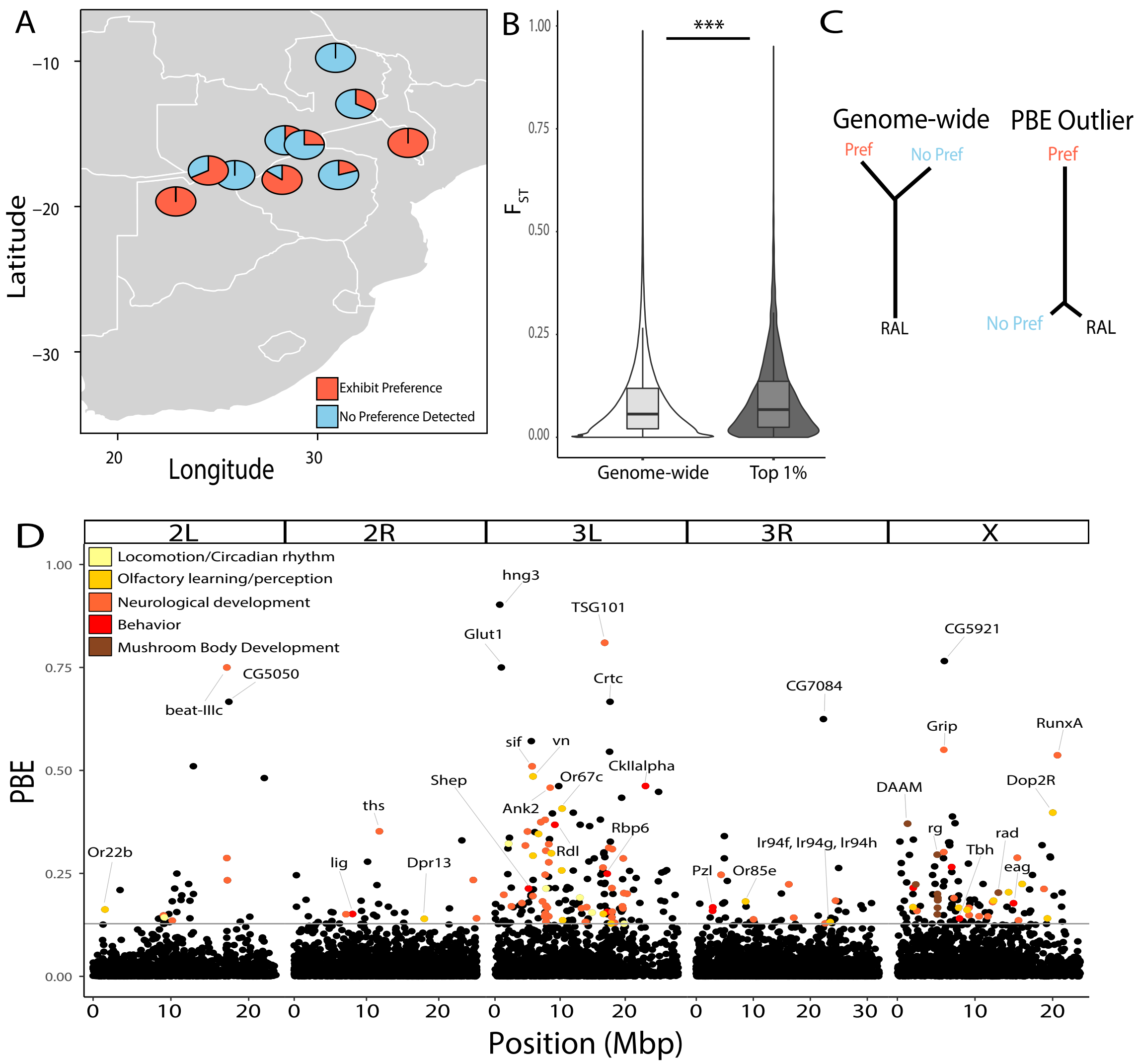
Figure S8: Courtship effort differs between males of different genotypes. Courtship effort (defined as the number of courtship attempts divided by latency) per behavioral type of male (M=Ral371, Z=ZS2) for each female in a two-way choice experiment. Z males attempt more courtings, particularly in the context of Z females.

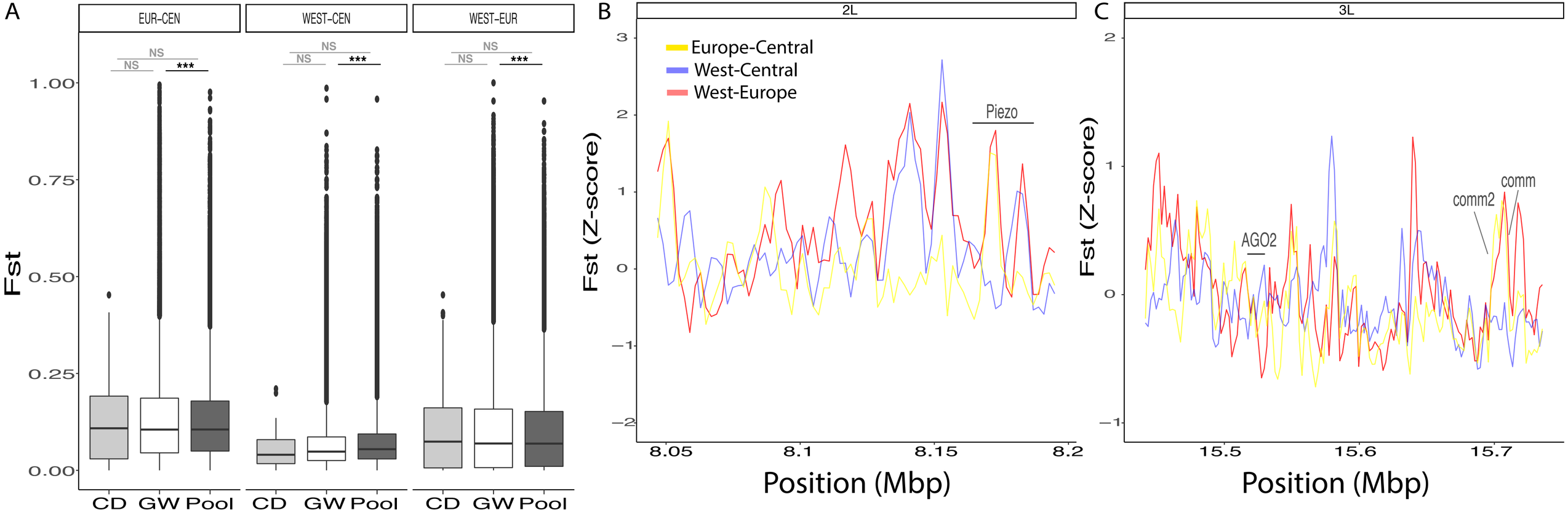

A
B
C
D
Latitude
Longitude
50
0
−100
0
100
PC2 (18.6%)
PC1 (46.7%)
0.05
0.00
−0.05
−0.10
Lineage
HD
Central
East
ET
North
OOA
South
West
OOA/North
East
West
East
Ethiopia
HD
Cen/South
Sim/Yak
0.00
0.25
0.50
0.75
1.00
Ancestry (%)


A
Pi
0.04
0.03
0.02
0.01
Cen
East
ET
HD
North
OOA
South
West
B
TajD
2
1
0
−1
−2
C
Frequency
0.75
0.50
0.25
0.00
CEN
autosomes
X
0–9
10–19
20–29
30–39
40–49
50
0–19
20–39
40–50


A
Africa & Non-African
Within-Africa
P1
P2
P3
O
Patterson's D
0.0
0.1
0.2
P2
West
HD
East
SE-US
OOA
NS
CEN
HD
SE-US
OOA
CEN
HD
P3
B
Africa & Non-African
Within-Africa
f_{dM}
0.0
0.1
0.2
P2
West
HD
East
SE-US
OOA
NS
CEN
HD
OOA
SE-US
CEN
HD
CEN
HD
P3


C
OOA
HD
AFR
OOA
AFR
HD
AFR
HD
OOA
2L
2R
3L
3R
X
Topology Weight
0.00
0.25
0.50
0.75
1.00
0
5
10
15
20
25
30
Position (Mb)


A
Latitude
−10
−20
−30
20
30
Longitude
Exhibit Preference
No Preference Detected
B
***
1.00
0.75
0.50
0.25
0.00
F_{ST}
Genome-wide
Top 1%
C
Genome-wide
Pref
No Pref
RAL
PBE Outlier
Pref
No Pref
RAL
D
2L
2R
3L
3R
X
Locomotion/Circadian rhythm
Olfactory learning/perception
Neurological development
Behavior
Mushroom Body Development
PBE
1.00
0.75
0.50
0.25
0.00
hng3
TSG101
Glut1
CG5921
CG5050
Crtc
CG7084
beat-IIIc
Grip
RunxA
sif
vn
CkIIalpha
Shep
Or67c
DAAM
Dop2R
ths
Ank2
Rbp6
rg
rad
Or22b
lr94f, lr94g, lr94h
Tbh
lig
Dpr13
Rdl
eag
Pzl
Or85e
0
10
20
30
Position (Mbp)


A
EUR-CEN
WEST-CEN
WEST-EUR
NS
NS
***
Fst
1.00
0.75
0.50
0.25
0.00
CD
GW
Pool
B
2L
Europe-Central
West-Central
West-Europe
Piezo
Fst (Z-score)
Position (Mbp)
8.05
8.1
8.15
8.2
C
3L
AGO2
comm2
comm
15.5
15.6
15.7