Patterns of and processes shaping 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: jcoug@email.unc.edu, dmatute@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

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. How genetic and phenotypic variation is partitioned across the range of *D. melanogaster*, particularly in its putative ancestral range in Subtropical Africa, remains unresolved. Here, we assess patterns of population genetic structure, admixture, mate preference, and genetic incompatibility across a global sample, including 174 new accessions from remote regions within Subtropical Africa. While almost all Out of Africa genomes correspond to a single genetic ancestry, different geographic regions within Africa contain multiple ancestries, with substantial cryptic diversity in Subtropical Africa. Admixture between distinct lineages is prevalent across the range, but admixture rates vary between lineages. Female mate choice within Subtropical Africa is highly polymorphic and behavioral types are not monophyletic. The genetic architecture of mate choice is highly polygenic, including loci associated with neurological development, behavior, olfactory perception, and learning. Finally, we discovered that many segregating putative incompatibilities likely evolved during or after expansion 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 polymorphic reproductive barriers.

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 partitioning 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

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

Post expansion, multiple historical events have also created opportunities for human-mediated admixture between genetically distinct lineages of *D. melanogaster*. It has previously been suggested that the opening of western commercial routes and the 'Scramble for Africa' facilitated hybridization between local African populations of *D. melanogaster* and invading non-African *D. melanogaster* individuals, particularly in more urban areas [25,26]. Indeed, the extent of non-African ancestry in Africa is widely variable between populations [18], with some evidence for more pronounced signatures of admixture in urban populations, [25,26]. Second, human migration associated with slavery roughly 400 years ago produced a secondary contact zone between African and non-African populations of *D. melanogaster* in the southeastern United States and the Caribbean [27–29]. However, it is unknown whether cryptic genetic lineages within Africa [16] have differentially contributed African ancestry outside of Africa [30]. Furthermore, the extent to which admixture between Out of Africa and African lineages contributes to the modern-day genetic composition of *D. melanogaster* within Africa is relatively unresolved (but see [18,25,26]). Lastly, patterns of admixture between African populations are almost entirely unexplored, despite their relevance for unraveling the evolutionary history of *D. melanogaster*.

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

107 populations is behavioral variation between populations which results in assortative
108 mating. Behavioral surveys of female mate choice within *D. melanogaster* from
109 Subtropical Africa revealed a potential case of incipient speciation within *D.*
110 *melanogaster* ([13,31–33]. At least two more cases of region-specific female mate
111 preference have also been described [26,34]). Cosmopolitan (denoted 'M') flies are
112 globally distributed, including within Subtropical Africa, while a second lineage is largely
113 restricted to Zambia and Zimbabwe (denoted 'Z'). While Z females show significant
114 preference for Z males, M females show no preference, and males of both lineages
115 court both types of females at similar rates [31,33,35]. Early attempts to map the genetic
116 basis of female preference found the trait to be highly polygenic, involving loci on all
117 major chromosome arms, and significant non-additive effects [31,32]. Despite the
118 contributions of *D. melanogaster* to the field of speciation [36–41], it remains unknown
119 whether the frequencies of these behavioral types covary with patterns of *D.*
120 *melanogaster* population structure within Subtropical Africa. Understanding if/how
121 behavioral isolation is associated with genetic structuring can allow us to determine if
122 divergence in complex, quantitative traits—such as mate choice—can contribute to
123 stable genetic divergence within a species; a hypothesis which is still widely debated
124 [42–44].
125 In addition to variation in female mate choice, negative epistatic interactions
126 between polymorphic loci, similar to hybrid incompatibilities, can cause substantial
127 variation in fitness, including the production of low fitness individuals within a species
128 [45–49]. While it has been hypothesized that these incompatibilities are geographically
129 structured across the *D. melanogaster* range [45], very little is known about the

130 geographic origins or distributions of these alleles. In particular, it remains unknown if
131 these alleles correspond to previously described behavioral and/or genetic lineages (i.e.
132 [13,16,33]). Determining how these forms of reproductive isolation correspond with
133 population structure and patterns of gene flow can contribute to our understanding of
134 the early stages of population differentiation and speciation.

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

148 **RESULTS**

149 *Diversity, divergence, and evolutionary relationships among populations of* D.
150 melanogaster

151 To understand the global distribution of diversity and population structure of *D.*

152 *melanogaster,* we combined whole genome resequence data from 420 lines of *D.*
153 *melanogaster* from around the world, including 174 newly sequenced genomes from un-
154 and under-sampled, rural locales within the proposed ancestral range in Subtropical
155 Africa. PCA revealed that genetic variation within *D. melanogaster* is mainly structured
156 between flies from Subtropical Africa and Out of Africa (OOA), as reflected by PC1
157 (which explains 46.7% of the variation). While somewhat intermediate along PC1,
158 Ethiopia, as well as East and West Africa, are much more distinct from OOA and
159 Subtropical Africa along PC2 (which explains 18.6% of the total variation). We further
160 identified eight unique genetic ancestries using *K*-means clustering in *PCAngsd* [50].
161 Four of these ancestries predominantly occur in Subtropical Africa, three largely
162 correspond to one each of Ethiopia, West Africa, and East Africa, and a final ancestry
163 type is most common in all OOA accessions (including *D. melanogaster* from North
164 Africa; Figure 1).

165 *K*-means clustering and PC analyses largely agree with our consensus
166 phylogeny with *D. simulans* and *D. yakuba* as outgroups. Of the eight ancestry types
167 identified, four correspond to largely monophyletic clades; OOA, Ethiopia, West Africa,
168 and a unique ancestry type from Subtropical Africa that comprises nine individuals
169 predominantly from Harare, Zimbabwe (denoted by gold in Figure 1). We refer to this
170 lineage as Harare-Distinct (HD). Of the remaining four ancestries, individuals from East
171 Africa cluster together in a PCA, but are split across our phylogeny, with some
172 individuals sister to West Africa, and others sister to all OOA individuals. The remaining
173 three ancestries are largely found in individuals from Subtropical Africa, but do not
174 correspond to monophyletic clades or unique sampling locales. They also do not exist in

175 pure form; the vast majority of Subtropical African individuals comprise two, and
176 sometimes three of these ancestries (denoted as yellow, red, and orange in Figure 1).
177 Although ancestry types do not correspond to monophyletic groups in these Subtropical
178 African individuals, we note that the largest and most distantly related clade of
179 Subtropical African flies has the largest proportion of one ancestry type (indicated by
180 orange in Figure 1), and largely contain samples from more remote sampling locales.
181 Combining these results, we define six genetic lineages that we use for subsequent
182 analyses: OOA, Ethiopia, East Africa, West Africa, HD, and all other Subtropical African
183 flies (which we refer to as Southern Africa).

184 Despite substantial population genetic structure, global F_{ST} estimates were
185 relatively low in almost every comparison, with the lowest global F_{ST} found between
186 East and West Africa (F_{ST}=0.027), and the highest between HD and OOA (F_{ST}=0.241;
187 Table S2). Variation in pairwise global F_{ST} is likely due to differences in nucleotide
188 diversity, as all pairwise comparisons of D_{xy} were very similar (Table S2), and the
189 populations with the highest global F_{ST} were those in which both populations also
190 showed the lowest pairwise nucleotide diversity (π; Table S2; Figure 2).

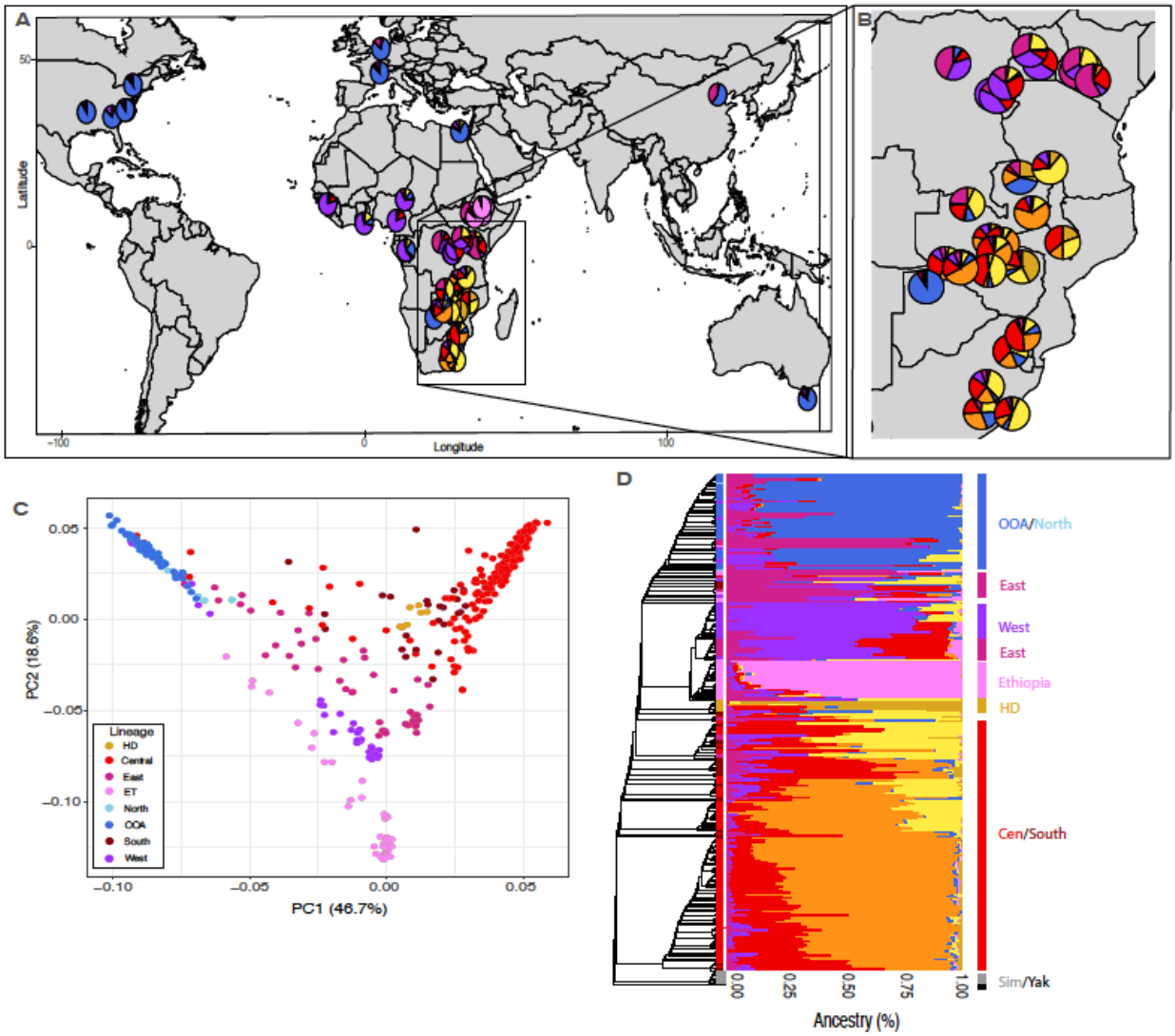


FIGURE 1. Broad sampling of *Drosophila melanogaster* from Subtropical Africa reveals substantial cryptic genetic structure. *(**A**) Geographic sampling of 420 genomes from a global distribution of D. melanogaster; pie charts represent the average ancestry determined by PCAngsd from that sampling local. (**B**) Zoom-in panel from (A), focusing on Subtropical and East African accessions. (**C**) PCA of all genomes, colors indicate the genetic lineage of the sample, defined herein. We also denote samples from North and South Africa (although they reside within OOA and Southern Africa, respectively). (**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 six genetic lineages that we identify using PCA, K-means clustering, and phylogenetic reconstruction.*

In addition to substantial cryptic genetic structure, the Southern African lineage

204 contains substantially higher levels of genetic diversity than any other genetic lineage
205 (Figure 2). This diversity is likely caused by an excess of rare alleles, as the Southern
206 African lineage also shows the most negative values of Tajima's D, and a left-skewed
207 Site-Frequency Spectrum (SFS), in line with this lineage having a much larger effective
208 population size and a recent history of population expansion (Figure 2, Figure S9; as
209 suggested in [22]). In contrast, two lineages show signals of recent population
210 contraction and lower levels of diversity: OOA (including North Africa) and HD (Figure
211 2). Tajima's D and the SFS also differ between chromosomes, although the direction of
212 these differences depended on the genetic lineage (based on an ANOVA with Type III
213 SS: chromosome arm × genetic lineage effect: F=162.9, df=20, p<0.0001). For four
214 lineages, the *X* chromosome had lower values of Tajima's D (Southern Africa, East
215 Africa, West Africa, and HD), while Ethiopia and OOA show the opposite pattern (Figure
216 2; Figure S3; Figure S9). Overall, these results suggest that both different lineages, and
217 different chromosome arms, of *D. melanogaster* have experienced different
218 demographic histories and even within Subtropical Africa, the HD lineage has a
219 significantly different demographic history than the rest of Southern Africa.

220

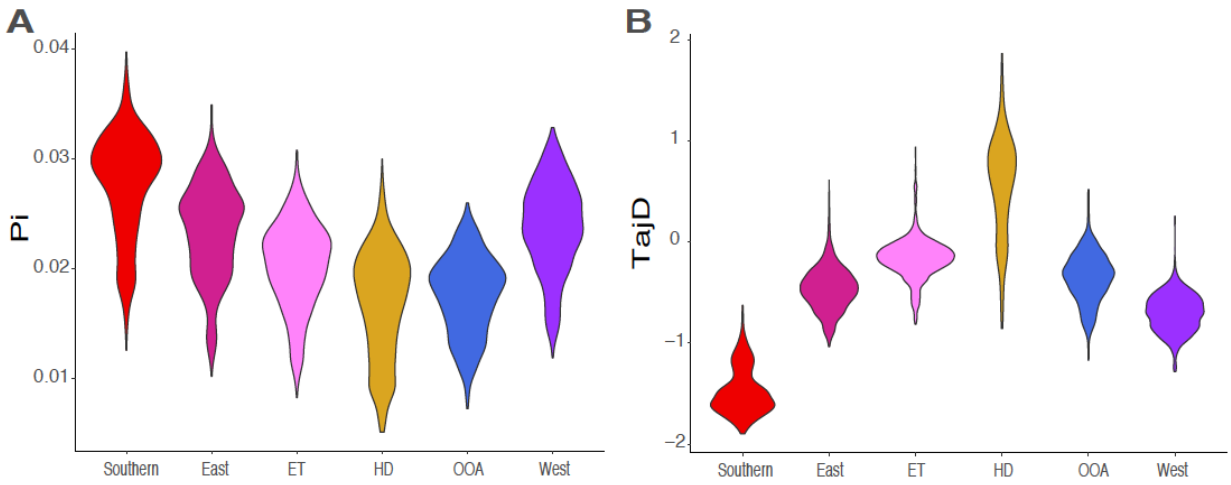


FIGURE 2. Genome-wide statistics for each genetic lineage. *Average (A) nucleotide diversity (π) and (B) Tajima's D for each genetic lineage.*

Patterns of gene flow throughout the range of D. melanogaster

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 Southern Africa, and (3) the source(s) of African ancestry in the SE United States (as proposed by [27,28]). We describe the results for each of these cases as follows.

First, we evaluated the extent of admixture among genetic lineages within Africa. We find evidence for extensive gene exchange between Southern Africa and both East and West Africa, but not between Southern Africa and either Ethiopia or HD (Table S3, Figure 3). Although gene flow between Southern and both East and West Africa is pervasive, the magnitude of gene flow is likely quite low, as only 1-3% of the genome is

238 inferred to be admixed between Southern Africa and each of East and West Africa,
239 respectively (depending on whether f_G or f_{dM} is used; Table S3, Figure 3). Similarly, we
240 find significant gene flow between HD and both East and West Africa (Table S3, Figure
241 3). Despite closer geographic proximity between most East African samples and
242 Southern African samples, introgression between West Africa and both Southern Africa
243 and HD is significantly elevated compared to introgression between East Africa and
244 either Southern Africa or HD (Figure 3; f_G between West Africa and either HD or
245 Southern Africa ranged from 0.02-0.033, while f_G between East Africa and either HD or
246 Southern Africa ranged from 0.017-0.022; we find no significant difference between
247 Southern Africa and HD in the extent of gene flow with either West or East Africa).
248 Overall, this suggests that the magnitude of admixture is not equivalent between
249 Subtropical African and other African *D. melanogaster* lineages and may not simply
250 correspond to geographic distance.

251 Second, we evaluated the extent of introgression between OOA and both
252 Southern Africa and HD. Specifically, we aimed to test whether back migration from
253 OOA may be responsible for the genetically unique HD lineage (denoted in gold in
254 Figures 1 and 2) as a test of whether urbanization has facilitated introgression in Africa.
255 Other genomic evidence suggests that this may be the case- the HD lineage shows
256 substantially reduced diversity and genomic patterns of a bottleneck relative to Southern
257 Africa, despite no geographic separation (Figure 1, Figure 2), and mimics patterns of
258 diversity and Tajima's D seen in OOA populations (Figure 2). In line with this
259 hypothesis, we find evidence of admixture between HD and OOA; f_{dM} is elevated
260 between OOA and HD when compared to OOA and Southern Africa (Figure 3), wherein

261 18-20% of OOA and HD genomes have introgressed, while only 15% of genomes
262 between OOA and Southern Africa have introgressed (depending on whether f_G or f_{dM} is
263 used; Table S3, Figure 3). However, tree topology weights generated by *twisst* further
264 reveal that the unique HD lineage does not appear to be simply a hybrid of flies from
265 Southern Africa and OOA (Figure 3C). While large proportions of the genome show
266 near complete support for topologies in which HD is sister to OOA (purple, in Figure
267 3C), the second most common topology does not place HD sister to samples from
268 Africa, but rather places all other African samples sister to OOA (teal, Figure 3C). This
269 suggests that rather than being a patchwork of African and OOA ancestry (in which we
270 expect yellow and purple topologies to be common in Figure 3C), HD carries distinct
271 genetic variation, at least in genomic regions where the topology supports all other
272 African genomes sister to OOA (i.e. the center of 2L). Extensive introgression between
273 African lineages and OOA may also contribute to these patterns. In total, this suggests
274 that while HD has likely experienced substantial introgression with OOA, it has also
275 likely experienced a unique evolutionary history from either OOA or the rest of Southern
276 Africa.

277 Last, we studied the origin of the African alleles harbored in the SE United
278 States. We find significantly elevated Patterson's *D* between SE United States and each
279 of Southern Africa, West Africa, and HD which suggests these populations have
280 contributed to *D. melanogaster* from the SE United States. We do not observe a similar
281 pattern for Ethiopia or East Africa (Table S3, Figure 3). Patterns of introgression are not
282 unique to the SE United States, as we find elevated Patterson's *D* between all other
283 OOA lines and each of Southern Africa, West Africa, and HD (Table S3, Figure 3),

284 suggesting that the contribution of these African populations might precede the split of
285 different OOA populations. However, if *D. melanogaster* from the SE United States
286 experienced a second pulse of introgression unique from other OOA populations, then
287 we predict that SE United States flies should show elevated signals of introgression with
288 at least one African source relative to other OOA flies. In line with this prediction, both f_G
289 and f_{dM} between West Africa and the SE United States are elevated relative to either f_G
290 and f_{dM} between West Africa and other OOA populations (Table S3, Figure 3), while
291 Southern Africa and HD show no difference in either f_G and f_{dM} with the SE United
292 States or other other OOA lines (Table S3, Figure 3). However, it is challenging to
293 detect regions of the genome that show unique signals of introgression between West
294 Africa and the SE United States relative to West Africa and all other OOA accessions,
295 as the landscape of f_{dM} across the genome is highly correlated between these two
296 comparisons (r^2=0.59, p<0.001; as it is between Southern Africa and OOA and
297 Southern Africa and SE United States: r^2=0.509, p<0.001, and between HD and OOA
298 and HD and SE United States: r^2=0.607, p<0.0001). Moreover, the weighted topologies
299 with West Africa as sister to all OOA versus only accessions from the SE United States
300 are highly similar (Figure S7). The lack of distinction may also be caused by rapid
301 purging of introgressed alleles, reducing any signals of introgression in a short time-
302 span [51]. In total, this suggests that all OOA populations have experienced some level
303 of gene flow with Southern Africa, West Africa, and HD, but flies from the SE United
304 States may have experienced a second, and independent pulse of West African
305 ancestry (as has been hypothesized by [27,28]).

306 We next examined how patterns of introgression vary across the genome within

307 Africa and between African and non-African lineages. For these analyses, we focus on
308 the three scenarios outlined above: (1) Introgression within Africa (as measured
309 between West Africa and Southern Africa), (2) introgression between OOA and the
310 unique HD lineage, and (3) introgression between West Africa and the SE United
311 States. In all instances, Ethiopia was used as P1 as we find little to no evidence of
312 admixture between Ethiopia and any P3 used herein (Table S3). We find that
313 chromosomes significantly differ in the extent of introgression for all comparisons (as
314 measured by f_{dM}; West-SE-US: F=47.11, df=5, p<0.001; West-South: F=18.36, df=5,
315 p<0.001; OOA-HD: F=16.01, df=5, p<0.001). For all three comparisons, the *X*
316 chromosome showed significantly higher f_{dM} values, than almost every autosome (Table
317 S4). However, definitive evidence of increased introgression on the *X*-chromosome is
318 much less apparent when using weighted topologies (Figure 3, Figure S7).

319 Many genes which may be involved in the transition to human commensalism fall
320 within the top 1% of f_{dM} windows (see Table S5 for full list). For example, these windows
321 include several genes involved in insecticide resistance, (including *Cyp6a18* (between
322 OOA-HD), *Cyp313a* and *ACE* (West-SE-US), and *Cyp12a4* and *LRR* (West-South)),
323 several genes related to metabolism, feeding behavior, and perception of and response
324 to food sources (including *happyhour*, *pbx, for,* and *Gfat1* (West-South), *NPFR, lovit,*
325 and *Ald1* (West-SE-US), and *Gr5a* (OOA-HD)), as well as genes involved in immune
326 function (*Dcr-2*, *Rab4*, *DptA*, *DptB*, and *Tl* (OOA-HD), *Npc2h*, *Npc2g*, and *ben* (West-
327 South), and *Charon* and *Ance-2* (West-Se-US)). Although causative connections are
328 still needed, it is perhaps unsurprising that alleles putatively related to human
329 commensalism may be overrepresented in three examples of potentially long-range,

and potentially human-mediated admixture.

Lastly, we find several genes associated with mating behavior fall within the top 1% of introgressed regions including *Desat2* (West-SE-US; although we note that this gene may also be involved in desiccation resistance, as is the case with other cuticular hydrocarbons [52,53]), *fru* (West-South), and *5-HT1A*, *clt*, and *Oamb* (OOA-HD). While these regions may not be involved in human commensalism *per se*, introgression at these loci may have implications for the distribution of female mate choice, and the distribution of female mate choice may be influenced by human commensalism; a possibility that we explore below.

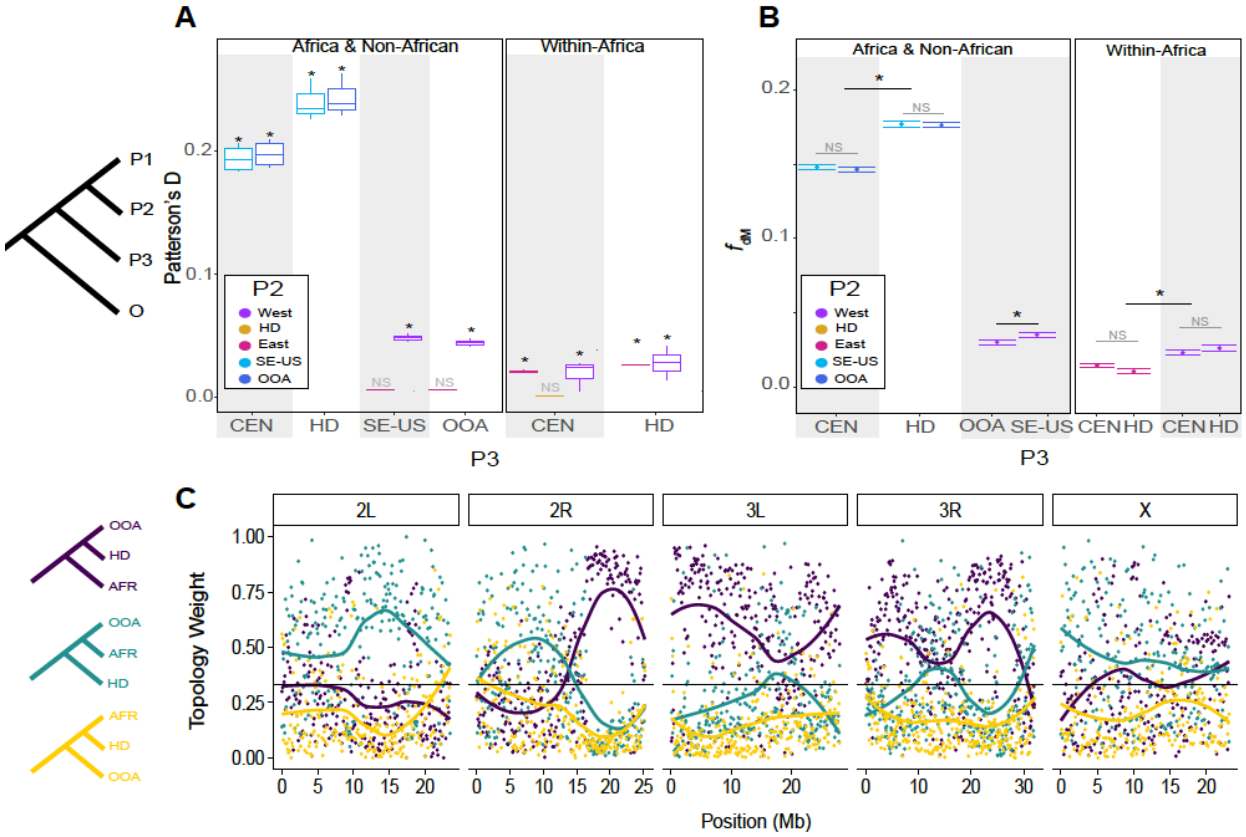


FIGURE 3: Patterns of gene flow between different African and non-African lineages of D. melanogaster. *(**A**) Boxplots of Patterson's D for all trios given the*

342 *phylogeny (((OOA, ((East, West),Ethiopia),HD),* Southern *Africa). P3 is given on the X-*
343 *axis, with the identity of the P2 population denoted by the color of the boxplot. Values of*
344 *D represent the range of values over multiple P1 populations. Comparisons that yielded*
345 *a significant Patterson's D based on a standard block jackknife procedure [54] are*
346 *denoted with an asterisk, otherwise they were deemed Non-Significant (NS). (**B**) Mean*
347 *and standard errors of 20-SNP windows of f_{dM} from across the genome for each African*
348 *genetic lineage showing significant introgression with either other African or non-Africa*
349 *lineages. The OOA lineage is split into those from the SE United States (SE-US) and all*
350 *others (OOA). Significance was determined by* ANOVA*s with Type III SS: NS= Not*
351 *Significant, * = 0.01<p<0.05. (**C**) Weighted topologies for three configurations of OOA,*
352 *HD, and Africa (denoted by colored phylogenies to the left).*
353

354

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

356 To understand if and how strong female mate preference originally described in
357 Subtropical Africa corresponds to any of the Subtropical African ancestries that we
358 identify herein, we surveyed a subset of lines from Subtropical Africa for the existence
359 of female mate choice using replicated choice experiments. Briefly, 47 focal isofemales
360 from ten sampling locales in Subtropical Africa were presented with both a Subtropical
361 Africa Z (ZS2) and OOA M (RAL371) isoline male. We quantified female mate
362 preference as a significant deviation from random choice, wherein ZS2 males were
363 more likely to copulate than RAL371 males. We find that female mate preference was
364 common in our experiment, but far from fixed across Subtropical Africa (e.g. 19/47 lines
365 had a preference for ZS2 males; Table S6, Figure S4). The frequency of lines showing
366 female mate preference varied among sampling locales, with the average proportion of
367 females exhibiting preference per population being 44% (range: 0-100%; X^2=19.04,
368 *df*=9, *p*=0.025; Figure 4; Table S6). Populations also varied in the strength of

369 preference. Some populations had nearly complete preference for Z males, and others
370 showed a mild aversion to Z males (X^2=18.63, *df*=9, *p*=0.029; Table S6; Figure S4;
371 average proportion of Z males chosen per population ranged from 0.398-0.969). Lastly,
372 we find that isolines that exhibit female mate choice are not monophyletic, nor do any
373 major PCs ascribing genetic variation among phenotyped flies separate isolines with
374 female preference from those that do not (Figure S5). Our results suggest that the
375 dynamics of mate preference are complex with Subtropical Africa; both the presence
376 and strength of female mate choice vary within and among populations, and the
377 presence of mate preference does not correspond with any obvious genetic structure.

378 We leveraged the natural diversity in female mate choice among these 47 lines
379 from Subtropical Africa to determine what regions of the genome are associated with
380 female mate choice. Given that isolines with strong preference do not show significant
381 population structure relative to isolines with no strong preference (Figure S5), we
382 instead quantified branch-specific evolution for lines with significant female mate
383 preference. We calculated the population branch excess (*PBE*) statistic [55], with
384 Subtropical African lines with strong female mate preference as the focal population and
385 Subtropical African lines with no strong female mate choice and RAL as the non-focal
386 populations. We defined the top 1% (368 windows) as outliers. These regions represent
387 all 5 chromosomal arms, although there is a significant excess of outliers on 3L and on
388 the *X* chromosome (X^2=183.01, df=4, p<0.0001). Of these 368 windows, 283 contain at
389 least one gene, and 84 of the remaining 85 windows are proximal to at least one gene
390 (of the windows in which genes are proximal, genes occur within 16.2KB on average,
391 range: 2.5-100KB). A total of 407 unique genes are included either within these

392 windows, or in close proximity (i.e. within 16.2KB on average). Gene Ontology
393 overrepresentation analyses indicate that of these 407 genes, there is an excess of
394 genes associated with learning, memory, cognition, chemotaxis, sensory perception,
395 detection of stimulus, behavior, and several categories associated with neurological
396 development (as well as other biological functions; see Table S8 for details).

397 Several individual genes are of potential interest for future study (Table S7). We
398 find thirty genes that are known to influence behavior in the top 1% of *PBE* outliers,
399 including five genes specifically involved in courtship behavior (*Shep, Rdl*, *Tbh, lig*, and
400 *eag*). Thirty-eight genes that are involved in sensory perception and sensory learning
401 are also within the top 1% of *PBE* values, particularly those involved in sensory
402 perception of smell or olfactory learning, including three odorant receptor genes (*Or22b*,
403 *Or67c*, and *Or85e*), four defective proboscis extension response genes (*dpr8*, *dpr13*,
404 *dpr14*, *dpr18*), seven ionotropic receptor genes (*Ir11a*, *Ir67b*, *Ir67c*, *Ir75a*, *Ir75b*, *Ir94f*,
405 *Ir94h*), *Dop2R* and *vn* (both involved in olfactory learning), as well as *Ank2* and *sif,*
406 which are involved in sensory perception of sound and visual perception, respectively.
407 Five genes which are known to influence male aggression fall in the top 1% of
408 behavioral *PBE* values (*Dop2R, Rpb6, CkIIalpha, Tbh,* and *Rdl*). Z males tend to be
409 substantially more aggressive during courtship than their M male counterparts,
410 particularly to Z females (Figure S8). Lastly, ten genes that are involved in mushroom
411 body development are also included in the top 1% of *PBE* values, including *Frl, DAAM,*
412 *PsGEF*, *rg*, and *rad*. Mushroom bodies play a central role in learning and memory,
413 particularly of olfactory perception, and therefore may be important in differentiating
414 male mates. Furthermore, mutants with aberrant mushroom bodies and inhibition of

415 mushroom bodies can cause virgin females to reject matings [56,57]. In total, we find
416 that *D. melanogaster* from Subtropical Africa with significant mate preference differ in
417 many regions across the genome, indicative of a polygenic basis of female mate choice.

418 We next sought to assess whether loci putatively involved in female mate choice
419 were also more highly differentiated among a broader geographic sampling of *D.*
420 *melanogaster* as a test of polygenic adaptation of this trait. Specifically, we asked
421 whether behavioral *PBE* outliers were among the most highly differentiated loci between
422 all Southern Africa lines (regardless of phenotype) and other geographic regions. We
423 find that F_{ST} among behavior outliers is elevated between Southern Africa and all other
424 genetic lineages relative to genome-wide F_{ST} (locus type (e.g. genome-wide versus
425 *PBE* outlier): F=6.426, df=1, p=0.011; genetic lineage effect: F=23799.7, df=4,
426 p<0.0001, locus type × lineage interaction: F=0.94, df=4, p=0.43, Figure 4, Figure S6).
427 However, despite this modest evidence for selection, we do not find that *PBE* outlier loci
428 are less likely to introgress than the rest of the genome, as would be expected if female
429 mate choice served as a significant barrier to introgression (f_{dM} for behavioral *PBE*
430 outliers is not significantly reduced relative to the rest of the genome between Southern
431 Africa and either West Africa: t=-0.064, df=406.27, p=0.95 or Europe: t=-1.52,
432 df=428.69, p=0.13). Furthermore, individual isolines that differ in the strength of female
433 mate preference show no difference in the history of introgression experienced (mean f_G
434 for lines with female mate choice = 0.097, mean f_G for lines with no strong female mate
435 choice= 0.112; t=-0.27, df=9.7, p=0.79). In total, this work suggests that while female
436 mate choice may show modest signals of polygenic adaptation, it is likely not a very
437 effective barrier to gene flow in natural *D. melanogaster* populations.

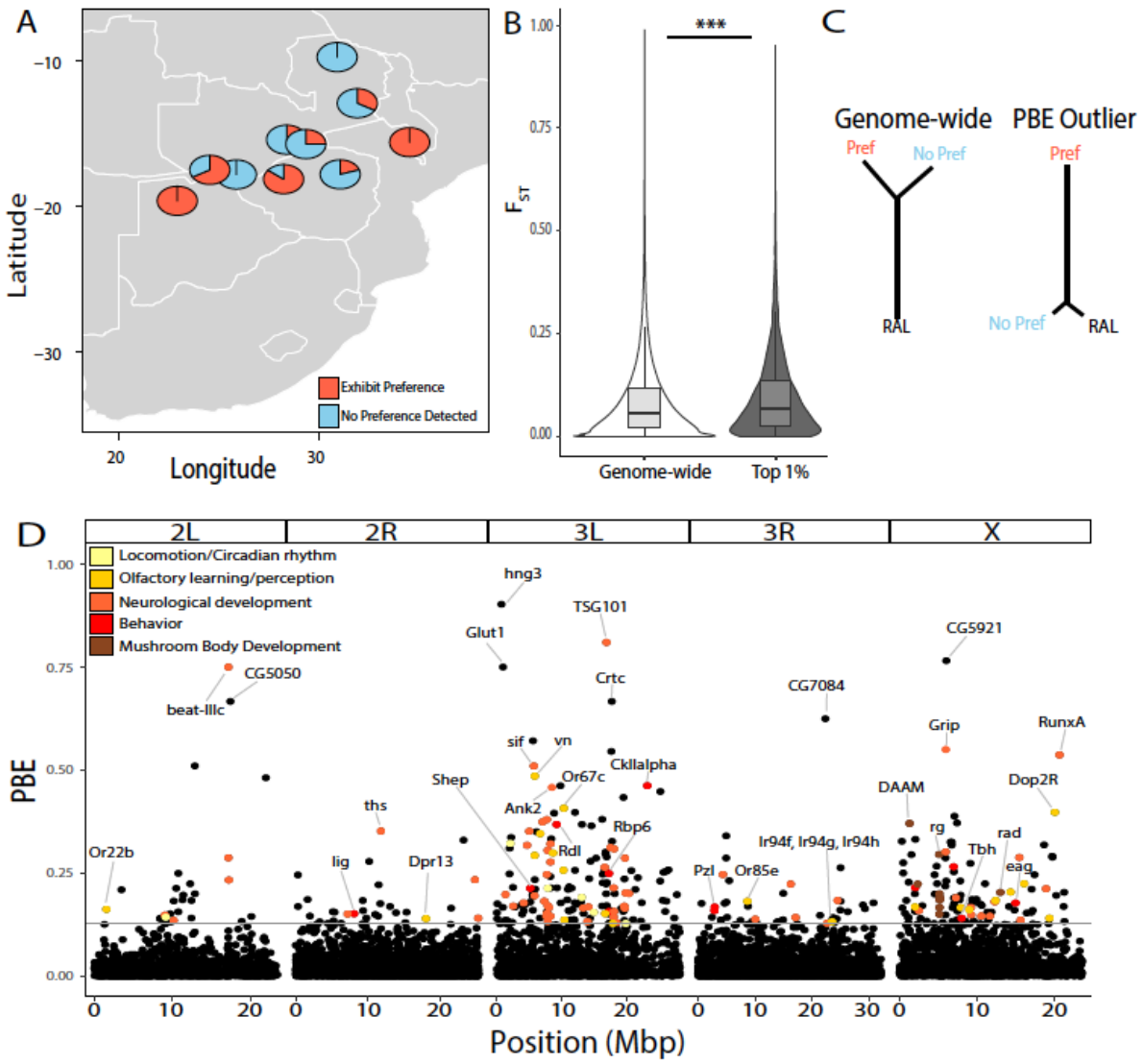


FIGURE 4: Geographic and genetic dissection of female preference in Subtropical Africa. *(**A**) Frequency of isofemale lines that show significant female mate choice in two-male choice experiments per sampling locale. (**B**) Average F_{ST} is elevated in the top 1% PBE regions between the Southern 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.*

451 *Distribution of putative incompatibilities throughout D. melanogaster*

452 We last assessed the distribution of alleles potentially involved in negative
453 epistatic fitness effects in admixed populations (herein 'incompatibility loci') throughout
454 the range of *D. melanogaster.* Although we refer to these loci as incompatibilities, they
455 include any loci that are found in repulsion of one another, and thus may include
456 traditional incompatibilities (i.e., involved in intrinsic postzygotic isolation), alleles
457 involved in ecological hybrid breakdown, as well as loci involved in assortative mating
458 [58]. We find that loci identified by [45], show slightly increased overall levels of
459 differentiation than the genome-wide average between Southern and West Africa, but
460 significantly lower differentiation than the genome-wide average between Europe and
461 each of Southern and West Africa (Figure 5). In contrast, we find no significant
462 difference in F_{ST} between loci identified by [46] and the genome-wide average for any
463 comparison. Thus, while on average, putatively incompatible loci are slightly more
464 differentiated between West and Southern Africa than other loci, they are not broadly
465 differentiated on a global scale.

466 We next assessed whether specific pairs of incompatibility loci showed elevated
467 differentiation, and used patterns of differentiation to characterize their potential
468 geographic origins. Out of 445 putative incompatibility loci identified by [45] and 45
469 putative incompatibility loci identified by [46], we identified only eight pairs of interacting
470 incompatibility alleles with high differentiation within Africa, and low differentiation
471 between Southern Africa and Europe; indicative of potentially older incompatibilities that
472 divide genetic lineages within Africa. Within these eight loci, eight unique genes are
473 included in the F_{ST} outlier windows, two of which were also found in our *PBE* behavioral

474 analysis (*Abl,* and *eag;* Table S9). In contrast, we identified 71 pairs of loci with a
475 signature of more recent origin (i.e. low differentiation in Africa, high differentiation
476 between Europe and both West and Southern Africa). We find that the F_{ST} peaks within
477 these 71 pairs of loci contain 133 unique genes (Table S9), 13 of which are also
478 behavioral *PBE* outliers. These include *DAAM, Rbp6, beat-IIIC, Doa, Lasp, luna, Eip75,*
479 *Mp, olf413, Pex7, pk, Svil,* and *eag* (we note this is a separate 1KB window than the
480 window which shows a signal of an putatively older incompatibility, described
481 immediately above). Many of these genes are involved in neurological development
482 and/or behavior (*DAAM, Rbp6, eag, beat-IIIC, Eip75, pk,* and *Mb*), and reproduction
483 (*Doa, Lasp*). Thus, many putative incompatibilities show a signature of more recent
484 evolution, and a significant proportion of these loci may also be differentiated between
485 behavioral types in Subtropical Africa.

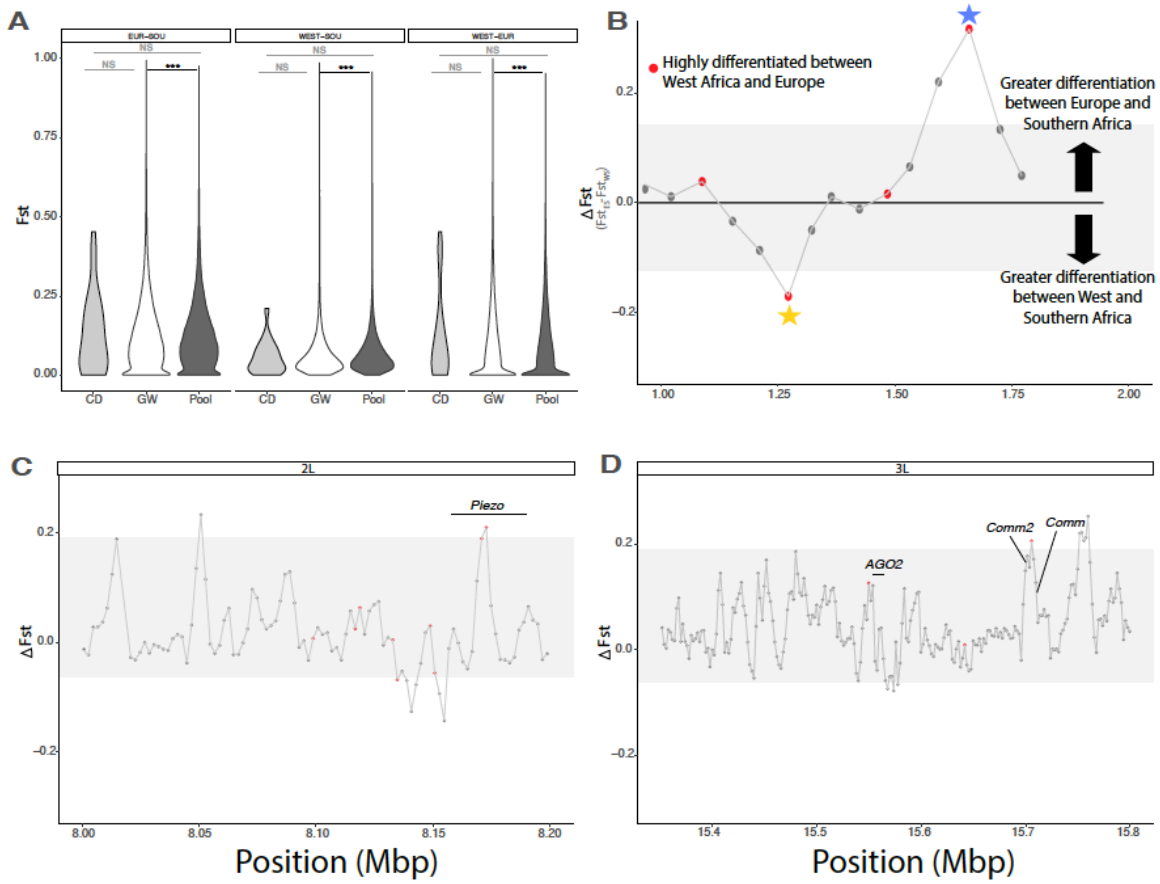


FIGURE 5: Global differentiation of putative incompatibility loci. *(**A**) Distribution of F_{ST} for incompatibility loci from CD [46], Pool [45] and genome-wide for three comparisons: Southern Africa vs Europe, West Africa vs Southern Africa, and West Africa vs Europe. (**B**) Cartoon depiction of a scan for incompatibility loci that segregate between African lineages (yellow star) or likely arose during or after the Out of Africa expansion (blue star). The Y-axis denotes the difference in F_{ST} between Europe and Southern African (ES) and West and Southern Africa (WS), with more positive values indicating that F_{ST} is greater between Europe and Southern Africa than within Africa. Red points denote loci that are also highly differentiated between Europe and West Africa. Loci that are highly differentiated between Europe and both African lineages are putatively newer incompatibilities, while loci that are highly differentiated within Africa, but shared between Southern Africa and Europe may represent older incompatibilities. (**C,D**) Two zoomed in windows representing one incompatibility pair, as identified by Pool [45]. In panel (**C**) differentiation is elevated between Europe and Southern Africa and West Africa and Europe (but not West and Southern Africa) for loci within the gene* Piezo*, indicating this allele may be more recently derived in OOA populations. In (**C**) we show the corresponding locus: again, differentiation is elevated between Europe and each of West and Southern Africa (but not within Africa) in a window containing no genes, but slight up and down stream of* comm *and* comm2*, respectively. We also highlight the approximate position of* AGO2*, the hypothesized interacting partner (based on [45]), and show there is no increased divergence in this region.*

508 **DISCUSSION**

509 The evolutionary history of genetic model systems has been the target of extensive
510 research, including *D. melanogaster* [4,5,16,18]. Nonetheless, sampling gaps across
511 critical regions of its range have left crucial aspects of its history unexplored. We
512 explore the partitioning of genetic diversity and patterns of gene flow across the global
513 range of *D. melanogaster,* including 174 samples from previously under-sampled
514 regions within the ancestral range. Furthermore, we evaluate the distribution of two
515 sources of polymorphic reproductive isolation that have been previously described:
516 mate choice and putative genetic incompatibilities. We find that flies from Subtropical
517 Africa harbor previously undescribed genetic diversity and population genetic structure.
518 We also find that *D. melanogaster* has experienced a complex history of gene flow,
519 particularly within Africa. Finally, we find that behavioral isolation and hybrid
520 incompatibilities segregating within the species have a multilayered history, but overall,
521 do not explain patterns of population genetic structure in Subtropical Africa. These
522 results contribute not only to our understanding of the natural and evolutionary history of
523 a powerful model system, but also in our understanding of how reproductive isolation
524 influences global patterns of population genetic structure. We discuss each of these
525 implications in turn.

526

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

528 *Drosophila melanogaster* individuals from its purported ancestral range of Subtropical
529 Africa show the highest genetic diversity and the lowest values of Tajima's D, in line

530 with a demographic history of population expansion and/or strong purifying selection (in
531 agreement with [13,15,16,20,22]). While these flies, originating largely from Zimbabwe,
532 Zambia, Malawi or Namibia, cluster in a genome-wide PCA, our phylogenetic analyses
533 indicate that Subtropical African individuals comprise a non-monophyletic clade that is
534 most distantly related to all other *D. melanogaster,* as would be expected if the mopane
535 forest in countries like Zambia and Zimbabwe are the likely origin of *D. melanogaster*
536 variation [4,5]. However, we also describe a strong signal of structure among individuals
537 from the ancestral range. Of the eight ancestries identified, four are most prevalent
538 within Subtropical Africa, only one of which is monophyletic in our phylogeny. This
539 monophyletic ancestry group—which we refer to as Harare Distinct (HD) —comprises a
540 distinct cluster in a whole-genome PCA (Figure 1). HD has substantially reduced
541 nucleotide diversity and elevated Tajima's D, indicative of a population bottleneck or
542 recent introgression (or less likely, balancing selection; [59]). These individuals also
543 have evidence of elevated introgression with OOA relative to Southern Africa. Given
544 that seven of nine individuals in this lineage are derived from the urban center of
545 Harare, HD may be a product of human-assisted migration of OOA individuals into
546 urban centers in Subtropical Africa. This is in agreement with previous findings based
547 on microsatellites [25], as well as previously identified patterns of human-assisted
548 migration in West Africa [26]. While our results partially agree with this hypothesis, we
549 also demonstrate that HD is unlikely to simply be an admixed population between OOA
550 and the Southern African lineage. Although much remains unknown about the HD
551 clade, the presence of previously undescribed diversity within the ancestral range of
552 arguably one of the most well studied model organisms highlights the importance of

553 thorough sampling from ancestral ranges when quantifying diversity and structure.

554 *Drosophila melanogaster* from other regions within Africa also show strong
555 population genetic structure, with West Africa, East Africa, and Ethiopia harboring three
556 distinct ancestry types, each clustering uniquely in a PCA, and West Africa and Ethiopia
557 being largely monophyletic. Individuals from these areas also have intermediate levels
558 of diversity and Tajima's D. This supports the hypothesis that these flies likely represent
559 some of the first steps of range expansion from Subtropical Africa (as suggested by [5]).
560 We also find evidence of gene flow between Subtropical Africa and both East and West
561 Africa. In contrast, Ethiopia does not show any introgression with Subtropical Africa and
562 represents a genetically unique lineage from other East Africa flies (in agreement with
563 [16,18]). Many of the Ethiopian samples herein originate from the high elevation
564 populations from the Ethiopian Highlands, which may be an environment that is not
565 permissive to migration from the lowlands [60–62]. Substantial geographic and/or
566 ecological barriers to migration may thus isolate individuals from the Ethiopian
567 Highlands from the rest of Africa.

568 In contrast, individuals from outside of Africa show substantially reduced genetic
569 complexity. All OOA flies cluster together in a genome-wide PCA, and largely comprise
570 a single ancestry. As a whole, these samples have reduced diversity and elevated
571 Tajima's D, which is suggestive of a bottleneck, likely accompanying an out of Africa
572 expansion (as has been found previously [13,20,22,24,63] and reviewed in [64]). We
573 find that all OOA individuals are sister to a subset of East African individuals, specifically
574 those from Kenya, but as a whole, OOA lines do not show evidence of introgression
575 with East Africa. It is therefore possible that OOA lineages are derived from East Africa,

576 and specifically from the ancestors of modern Kenyan flies. Additionally, we find that
577 African lineages have differentially contributed to genetic variation within OOA via post-
578 expansion introgression. We find that all OOA flies have similar levels of introgression
579 with the Southern Africa and HD lineages, while individuals from the SE United States
580 show slightly elevated levels of gene flow with flies from West Africa, relative to other
581 OOA individuals. It has previously been suggested that flies from the SE United States
582 represent a secondary contact zone between individuals of European and African
583 ancestry [16,18,27,45]. Here we show that African lineages have contributed
584 differentially to *D. melanogaster* populations outside of Africa.

585 Taken together, these broadscale analyses indicate that *D. melanogaster* has
586 substantial population genetic structure within the ancestral range in Subtropical Africa,
587 including the presence of a unique, and potential admixed lineage found in Harare,
588 Zimbabwe. Patterns of gene flow within Africa are complex, as there is at least some
589 degree of gene flow among most genetic lineages, with highland Ethiopian lines
590 remaining relatively distinct and unconnected. Gene flow between African and non-
591 African lineages is also variable, with Subtropical and West Africa exhibiting the highest
592 levels of gene flow with OOA, and West Africa specifically exhibiting a unique pulse of
593 gene flow with flies from the SE United States.

594 *Introgression is elevated in genes related to human-commensalism*

595 In addition to variation in broadscale patterns of gene flow between distinct genetic
596 lineages of *D. melanogaster,* we also find that patterns of introgression vary across the
597 genome. First, using f_{dM}, introgression is higher on the *X* chromosome than autosomes

598 for all comparisons (Table S4). This is unusual given that sex chromosomes are often
599 depleted for introgression [65–70]; indeed, within *D. melanogaster* OOA populations
600 often lack African ancestry on the *X* (for example see [45]), although several examples
601 of systems with high levels of gene flow on the *X* also exist [71–74]. Because the *X*
602 chromosome also contains an excess of alleles associated with Z/M female mate
603 choice, our discovery that the *X* chromosome exhibits the highest levels of introgression
604 has implications for the efficacy of female mate choice as a barrier to introgression, a
605 possibility that we explore below.

606 Genomic windows with the highest levels of introgression contain an excess of
607 genes associated with human commensalism, specifically genes involved in insecticide
608 resistance, metabolism/diet, and immune system function. Human commensalism is a
609 novel, and often adverse environment for organisms. Human-induced habitat change is
610 a primary driver of hybridization in many systems [75,76]. While the loss of local
611 adaptation due to introgression of domesticated traits is often considered to be a cost of
612 hybridization [77], wild populations without certain domesticated traits may simply be
613 unable to survive in human proximity. The genes we identify as outliers in introgression
614 analyses are often associated with the unique pressures of human commensalism (see
615 [78] for an overview)— insecticide resistance [5,79,80] and diet/metabolism [81–85] are
616 common targets of selection via human commensalism. Increased population densities,
617 and/or novel parasites and diseases associated with human commensalism may also
618 favor increased or novel immunological function [86]. Previous work has established
619 that *D. melanogaster* from non-human associated locales show elevated evolution of
620 some immunity-related genes [5], suggesting that the transition to human

621 commensalism in *D. melanogaster* specifically may involve allelic changes in immunity.

622 We also find that several genes involved in mate choice and courtship behavior
623 fall within the top 1% of introgressed regions. One potential explanation is that strong
624 female mate preference comes at a fitness cost in more resource-competitive urban
625 environments, potentially by limiting mating potential or delaying reproduction. Thus, it is
626 possible that the cost:benefit ratio of female choice may differ under scenarios of human
627 commensalism, ultimately favoring the loss of female choice in these environments.
628 Under this hypothesis, subsequent introgression of less-choosy alleles back into more
629 urban regions of Subtropical Africa may be favored. Although much more work is
630 needed to further test this hypothesis, previous work has demonstrated the loss of
631 female preference in more urban populations of *D. melanogaster* from Brazzaville,
632 Congo [26]. Similarly, we find that the potentially admixed urban HD clade contains
633 fewer isofemale lines with female mate choice than do the rest of our sample from
634 Subtropical Africa, despite close geographic proximity (Figure 4, Figure S5).

635

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

637 Female mate preference in Subtropical African *D. melanogaster* is a classic example of
638 incipient speciation [31–33,35,87–89]. We revisited this classic system to survey flies
639 from Subtropical Africa for the presence and strength of female mate choice, particularly
640 in previously unsampled and remote locales in Zambia, Zimbabwe, and Namibia. We
641 find that female mate choice is relatively common, but also highly polymorphic within
642 and among sampling locales. Female mate choice is also not monophyletically

643 distributed. These results are in line with previous work that found substantial variation
644 in female mate preference for a smaller number of isolines [35], as well as recent work
645 highlighting the variation in male courtship behavior within Subtropical Africa [31–
646 33,35,87–89]. Three evolutionary scenarios may explain this polymorphism: (1) female
647 mate choice is an ancestral polymorphism which is maintained in Subtropical Africa, (2)
648 strong female mate choice is the ancestral condition, and introgression from outside of
649 this region has eroded its prevalence within Subtropical Africa, or (3) female mate
650 choice is a derived character within Subtropical Africa and has either been eroded by
651 introgression or never fully fixed throughout this region. While we cannot disentangle
652 these scenarios with our current dataset, unraveling these potential scenarios has
653 implications for the stability of reproductive isolation in natural populations, as well as for
654 the potential evolutionary costs and benefits of female choosiness.

655 One potential caveat to the above female mate choice work is that we used
656 standard Z and M lines for all assays as a way to standardize the female preference
657 tests. However, one possibility is that female mate choice is more common than we
658 estimate, but preference is specific to local males (e.g. not all Z males are equivalent;
659 [87–89]). Several lines of evidence suggest that this may be true for some populations.
660 For example, the population with one of the highest levels of female preference was the
661 population from which the focal male was derived (e.g. ZS). Additionally, studies that
662 used a standard M line and a male of the same isoline as the focal female generally find
663 greater values of female mate choice [35]. Lastly, African males vary in both cuticular
664 hydrocarbon profiles and mating displays, which may be key traits involved in female
665 mate choice [89,90]. Therefore, while we present the most widespread survey of female

666 mate choice from Subtropical Africa to date, more research will be needed to more
667 precisely characterize the dynamics of female choice in this system.

668 We leveraged natural variation in the presence of female mate choice within
669 Subtropical Africa to assess genomic differences associated with female choosiness
670 and male attractiveness. Although genomic outliers exist on all major chromosome
671 arms, chromosomes 3L and the *X* are highly enriched for outliers. These results are
672 largely congruent with previous analyses [31,32], finding that the third chromosome had
673 the largest effect on both female preference and male attractiveness, including a large
674 contribution from 3L. In contrast, [31], found a much smaller contribution of the *X*
675 chromosome in single chromosome replacement experiments, but substantial epistasis
676 between the *X* chromosome and both major autosomes.

677 Behavioral *PBE* outliers are significantly enriched for genes that are involved in
678 male courtship behavior, olfactory perception, learning, memory, and neurological
679 development. Several individual genes present interesting candidates for future work,
680 including genes involved in mushroom body development and sensory perception.
681 Mushroom body ablation in virgin females results in higher rejection rates [56,57], and
682 influences male memory and courtship effort [91]. Additionally, while the precise male
683 traits controlling attractiveness are unknown, cuticular hydrocarbons likely play an
684 important role [87,89,92]; and therefore, evolution of sensory perception—particularly
685 sensory perception of smell and taste—may also help regulate female mate preference
686 (e.g. ([93]). Lastly, we find that two genes (*shep* and *eag*), which have been shown
687 directly to affect mating behavior, are highly differentiated in Subtropical African lines
688 with significant female preference. *Shep* is responsible for neuron remodeling during

689 metamorphosis, and whose loss-of-function results in increased rejection of male mates
690 in virgin females [94]. *Eag* regulates potassium permeability, and mutants of this gene
691 have been shown to influence the amount of time males spend courting [95,96]. Thus,
692 while we present some intriguing candidate genes for future functional analyses, overall
693 our work suggests that the genetic basis of mate preference is likely highly polygenic.

694 Lastly, while we find that *PBE* behavioral outliers show elevated differentiation
695 between a broader sampling of the Southern African lineage and non-Southern African
696 populations (particularly those from OOA), we also find that these behavioral *PBE*
697 outliers do not show decreased levels of introgression relative to the rest of the genome,
698 nor do lines with strong female preference show a reduced history of introgression.
699 Taken together, these results could arise from a scenario in which female mate
700 preference has experienced polygenic local adaptation, but asymmetric mate
701 preference is insufficient to noticeably dampen gene flow between mating types in
702 Subtropical Africa. In particular, female mate choice may still be maintained if the
703 genetic basis of mate choice is partially redundant (e.g. multiple alleles are sufficient to
704 cause female mate choice, and therefore the loss of any given allele by introgression
705 does not dampen female mate choice). Earlier mapping efforts [31,32] suggest the
706 existence of genetic redundancy in female mate preference and male attractiveness,
707 though more detailed genetic analyses are needed.

708

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

710 We lastly studied the distribution of previously identified putative incompatibility

711 loci across our global sampling of *D. melanogaster.* On average, these loci show slightly
712 elevated differentiation between Southern and West Africa. However, while some pairs
713 of loci have a signature of differentiation within Africa, there are substantially more pairs
714 of loci that are highly differentiated between Europe and both West and Southern Africa.
715 Moreover, we find that of the putative incompatibility loci that are highly differentiated,
716 10-25% of genes within these windows overlap with our behavioral *PBE* outliers (for
717 within Africa or between Europe and both African lineages, respectively). There are two
718 plausible explanations for the high overlap between these two datasets. First, some of
719 the putative incompatibility loci identified by [45] may be involved in assortative mating
720 [58]. Possible candidates for this scenario in our analyses include *eag,* which has been
721 shown to directly influence courtship behavior; *DAAM*, which is involved in mushroom
722 body development; and *rbp6* which influences male aggressive behavior. Second, while
723 *PBE* analyses are an effective approach to detect branch-specific evolution for a focal
724 population [5,55], by definition, this statistic cannot differentiate branch-specific
725 evolution that is specific to a trait of interest versus subtle population structure and/or
726 branch-specific evolution of non-focal traits. Therefore, if some incompatibility alleles
727 exist at higher frequencies in isolines with strong female choice relative to isolines with
728 no strong female mate choice, they too may be included in our *PBE* outlier analysis
729 (even though these alleles may not directly influence female mate choice). Regardless,
730 our results suggest that putative incompatibility alleles show meager genetic structure
731 across the range of *D. melanogaster*, and knowledge about the function of these alleles
732 can improve our understanding of barriers to gene flow within this model system.

733

734 **CONCLUSION**

735 Our work contributes to understanding of the evolutionary history of *D. melanogaster*,
736 including the distribution of genetic diversity and reproductive isolation within this model
737 organism. We combined population genomics and behavioral assays to assess how
738 genetic and phenotypic diversity is partitioned across the globe in *D. melanogaster*,
739 especially within its putative ancestral range in Subtropical Africa. While we find
740 significant population genetic structuring throughout the range, we also find that
741 polymorphic reproductive barriers are largely decoupled from broader patterns of
742 population genetic structure. Furthermore, this work provides insights into the
743 complexity of genetic changes underlying female mate choice, and highlights many
744 genes of interest for future study.

745

746

747 **MATERIALS AND METHODS**

748 *Sampling, Sequencing, and Variant Calling*

749 We created 244 new *D. melanogaster* isolines from an original 339 wild collected
750 females derived from seven novel locations in Zambia, Namibia, and Zimbabwe using a
751 similar approach to previously described efforts ([5]; see Table S1 for sampling
752 locations; see supplemental methods for details). We then extracted DNA and created
753 whole-genome libraries for each of 174 unique isofemale lines plus 32 advanced
754 generation isofemale lines collected in Malawi, Zimbabwe and Botswana (see Table S1
755 for details). Individually barcoded pools of ~10 isolines were then sequenced by the

University of North Carolina (UNC) School of Medicine, on either a single lane of an Illumina HiSeq 4000 or a Noveseq6000S4XP platform, in both cases generating paired-end 150 bp reads (see supplemental methods and Table S1 for details). These re-sequenced lines were paired with whole genome sequences for an additional 247 isolines via NCBI SRA, including 72 lines from outside of Africa and 175 from within Africa, 30 of which are derived from Subtropical Africa [16,18] (see Table S1 for details).

We generated a vcf of all 420 *D. melanogaster* plus 12 *D. simulans* and 1 *D. yakuba* genomes using a standard pipeline that followed best practices (see supplemental methods and Table S1 for details). The resultant VCF was filtered so that indels were removed, and only biallelic sites with a minimum quality score of 30, minimum coverage of 5X, minimum genotype quality of 30, a maximum of 50% missing data were kept. We additionally removed 10 individuals with poor quality genomes (e.g. less than 5X average coverage).

Lineage relationships: population structure, PCA, and phylogenetic reconstruction

To better understand the relationships among a global sampling of *D. melanogaster*, we constructed maximum likelihood (ML) phylogenies for the autosomes and *X* chromosome separately using *iqtree* version 1.6.12 [97–99]. We generated ML trees for non-overlapping 100KB windows using the model-finder and ultra-fast bootstrap approach with 1,000 bootstraps. We then used the resulting ML trees as input for *ASTRAL v5.1.1 [100]* in generating a consensus tree for the autosomes and *X* chromosome independently (Figure S2).

778 To characterize fine scale population genetic structure we performed *K*-means
779 clustering analysis and PCA using *PCAngsd* [50] for the *D. melanogaster* samples.
780 PCAngsd uses genotype likelihoods to first perform a genome-wide PCA, then assess
781 population structure with the number of ancestry types (*K*) defined as the number of
782 significant PCs + 1. We only included sites with <10% missing data, a minimum
783 mapping quality of 30, and a minimum base quality of 20. This generated a dataset of
784 461,147 sites for 420 individuals.

785 Given our phylogenetic, PC, and *K*-means clustering analysis, we identified six
786 genetic lineages which correspond to five geographic regions (all Out of Africa lines
787 (OOA); all East African lines excluding lines from Ethiopia (East), all West African lines,
788 all Ethiopian lines, nine lines from Subtropical Africa which form a unique ancestry
789 group that we refer to as Hahare-Distinct (HD), and all other Subtropical African lines
790 (Southern)). We calculated all pairwise measures of divergence and differentiation
791 between these lineages (e.g. F_{ST} and D_{xy}; defined in Figure 1) and nucleotide diversity
792 within each lineage (e.g. π) in 1KB windows across the genome using *pixy* [101] with
793 default filtering expressions (i.e. DP>=10,GQ>=20,RGQ>=20). For this analysis we
794 used an all-sites VCF to include invariant sites, and performed the same filtering above,
795 but retained sites with a maximum of 2 alleles, so as to include invariant and biallelic
796 sites. We also calculated Tajima's D using *VCFTools* [102], and estimated the Site-
797 Frequency Spectra (SFS) using *SweeD* [103] for each genetic lineage.

798 *Introgression Analyses*

799 To estimate broad patterns of gene flow between genetic lineages of *D.*

800 *melanogaster,* we first calculated Patterson's *D* and f_G (a Patterson's *D* derivative which
801 more accurately estimates the proportion of the genome experiencing introgression
802 [104]) using *Dsuite* [54] with *D. simulans* and *D. yakuba* as outgroups for all possible
803 trios, given the following phylogeny: (((OOA, ((East, West),Ethiopia)),HD), Southern).
804 Significance of Patterson's *D* was determined using a standard block jackknife
805 procedure [54]. In specific cases (outlined below) we also quantified differences in the
806 extent of introgression between genetic lineages that exhibited significant Patterson's *D*,
807 using f_{dM} in non-overlapping 20 SNP windows for focal trios (outlined below). f_{dM} is a
808 Patterson's *D* derivative that is more appropriate for windowed analysis and provides a
809 more accurate estimate of the proportion of the genome that has experienced
810 introgression than Patterson's *D [54]*. Finally, we bolstered our introgression analyses
811 by assessing heterogeneity in the genome in the relationships between potentially
812 introgressing groups by calculating tree topology weights using *twisst [105]*. For each
813 focal trio (outlined below), we calculated the topology weight at each non-overlapping
814 100KB window for trees comprising four groups: *D. simulans* as an outgroup, the two
815 potentially introgressing groups (outlined below) and the remaining African flies as a
816 single group (AFR). While f_{dM} calculates the proportion of shared derived variants
817 between non-sister lineages, *twisst* assesses the proportion of topologies that fit
818 particular phylogenetic relationships. These analyses thus provide complimentary, but
819 uniquely informative quantifications of introgression.

820 Our goals were threefold: (1) assess if female mate choice in Subtropical Africa
821 is associated with reduced gene flow with other African lineages, (2) quantify if
822 urbanization in Subtropical Africa is associated with increased levels of gene flow with

823 OOA, and (3) identify the source(s) of African ancestry in the SE United States. To
824 address each of the specific questions listed above, we used specific trios.

825 Gene flow between Southern Africa and other African lineages: we first assessed
826 patterns of gene flow between all trios involving HD or Southern Africa as P3, and
827 Ethiopia, West, and/or East Africa as P1 and P2 using Patterson's *D*. For trios that had
828 significant Patterson's *D*, we then calculated f_{dM} in 20 SNP windows across the genome
829 to evaluate the extent of introgression. To determine whether the extent of introgression
830 varied between trios, we used an ANOVA with windowed f_{dM} as the response variable,
831 and P3 (HD or Southern Africa), P2 (East or West), and their interaction as independent
832 variables with a Type III Sum of Squares using the *car* package in R [106]. In all cases,
833 Ethiopia was used as P1, as there were no scenarios in which Ethiopia demonstrated
834 significant introgression with either HD or Southern Africa (Table S3; Figure 3).

835 Urbanization and introgression with OOA: we calculated Patterson's *D* for all trios in
836 which Southern Africa (composed largely of rural individuals) or HD (composed largely
837 of urban individuals) was P3, and OOA and either Ethiopia, East, or West Africa was P1
838 or P2. Then, to determine if either HD or Southern Africa has experienced more
839 introgression with OOA, we calculated f_{dM} in 20 SNP windows across the genome for
840 both comparisons with Ethiopia as P1, OOA as P2, and either Southern Africa or HD as
841 P3. We assessed differences in windowed f_{dM} using an ANOVA with Type III SS using
842 the *car* package in R [106], with f_{dM} as the independent variable and P3 (HD or Southern
843 Africa) as the independent variable.

844 Source(s) of African ancestry in the SE United States: we performed two sets of

845 analyses. Again, we calculated Patterson's *D* with either lines from the SE United States
846 or all other OOA lines as P3, and Ethiopia, East and/or West Africa as P1 and P2. We
847 calculated Patterson's *D* with HD or Southern Africa as P3 and either lines from the SE
848 United States or all other OOA as P2, and each of Ethiopia, East and West Africa as
849 P1. As flies from the SE United States are hypothesized to represent a contact zone
850 between OOA and an unidentified African lineage, we expect that flies from the SE
851 United States should have increased levels of introgression with the source African
852 lineage(s), relative to the rest of OOA flies. Therefore, we again calculated f_{dM} in 20 SNP
853 windows across the genome for trios with significant Patterson's *D*, and asked which (if
854 any) African lineages had higher f_{dM} with the SE-United States than with OOA using
855 ANOVAs with Type III SS using the *car* package in R [106] with windowed f_{dM} as the
856 response variable and each OOA lineage (SE-United States or all other OOA lines) as
857 the independent variable.

858 Last, to assess heterogeneity in introgression across the genome, we leveraged
859 the windowed f_{dM} analyses described above and first asked if chromosomes varied in
860 the extent of introgression using an ANOVA with Type III SS in the *car* package in R
861 [106]. For each focal trio we performed these analyses with f_{dM} as the response variable
862 and chromosome arm as the independent variable. We then used pairwise *t*-tests to
863 assess significant differences between chromosome arms. To determine what genomic
864 regions were specifically more likely to introgress in each of our focal comparisons, we
865 also defined introgression outliers as windows with the top 1% of f_{dM} values.

866 *Assessing mate preference in Subtropical African samples*

867 To assess the prevalence and strength of female mate choice in Subtropical
868 Africa, we performed a series of replicated choice experiments for 47 focal isofemale
869 lines from Subtropical Africa. Briefly, each choice experiment was conducted by
870 presenting a 7-10 day old virgin female with a standard Z and M male line (ZS2 and
871 RAL371, respectively). Vials were watched continually for up to three hours, and once
872 mating began the unmated male was removed from the vial by aspiration. We
873 completed an average of 53 trials (range: 12-100) per focal isofemale line, with an
874 average of 37 trials resulting in a successful mating per isofemale (range: 8-71;
875 1,724/2,507 total trials were successful). See the supplemental methods for full
876 experimental details.

877 We assessed deviations from random mating for each isofemale line using a
878 Fisher's exact test, then assessed population-level differences in the proportion of
879 females exhibiting preference and the strength of preference using a logistic regression
880 or a generalized linear mixed effect regression with poisson distribution, respectively. In
881 both models, population was the independent variable. For the model assessing the
882 strength of preference, we also included the isoline, replicate trial date, and total
883 matings as random effects. For both of these analyses we used the *lme4* package [107]
884 and assessed significance of fixed effects using an Type III Wald's X^2 test using the
885 *Anova* function in the *car* package in R [106].

886 To test for genetic differences associated with female mate preference, we first
887 calculated the Population Branch Excess (*PBE*) statistic [55] in 1KB windows across the
888 genome with Subtropical African lines with significant female mate choice as the focal
889 population, and Subtropical African lines that do not have strong female mate choice

890 and RAL lines as the non-focal populations. *PBE* is a derivative of the population branch
891 statistic (i.e. *PBS*; [108]), but quantifies branch-specific evolution for a focal population
892 relative to two non-focal populations [55]. We then defined behavioral *PBE* outliers as
893 the top 1% of *PBE* values (368 windows). Using this list of outliers, we did a Gene
894 Ontology (GO) Enrichment Analyses with a Fisher's Exact test with an FDR threshold of
895 0.05 using *PANTHER* 16.0 [109] to assess if behavioral *PBE* outliers were enriched for
896 biological GO terms relating to behavior, sensory perception, memory, learning or
897 neurological development, each of which may have specific function in female
898 choosiness or male attractiveness [96].

899 Lastly, we aimed to determine whether female mate choice in Subtropical Africa
900 is associated with global patterns of differentiation or introgression in *D. melanogaster*
901 to assess whether female mate choice serves as a significant barrier to gene flow in
902 nature. To do this, we performed three subsequent analyses. First, we calculated
903 whether behavioral *PBE* outliers had elevated differentiation relative to the rest of the
904 genome between the Southern Africa lineage and every other genetic lineage. We used
905 the aforementioned estimates of F_{ST} calculated in 1KB windows between all Southern
906 Africa lines (regardless of phenotype) and every other genetic lineage. We then
907 performed an ANOVA with Type III SS to assess whether F_{ST} differed between locus
908 type (genome-wide versus behavioral *PBE* outlier), between comparisons (each genetic
909 lineage), and their interaction using the *car* package in R [106]. We used the *emmeans*
910 package to estimate significance of specific contrasts [110]. Second, to determine
911 whether loci potentially involved in female mate choice were associated with differential
912 patterns of introgression, we also ascertained whether windowed f_{dM} between Southern

913 Africa and each of West Africa and Europe differed between behavioral *PBE* outlier loci
914 and the rest of the genome, using the f_{dM} datasets described above. To do this, we
915 performed a *t*-test with windowed f_{dM} as the dependent variable, and locus type
916 (behavioral *PBE* outlier vs genome-wide) as the independent variable. Third, we asked
917 whether isolines that had significant female mate choice also showed an overall lower
918 history of introgression than isolines with no strong female mate choice. To do this, we
919 recalculated Patterson's *D* and f_G for trios involving OOA, West Africa, East Africa,
920 and/or Ethiopia as P1 and P2, and isolines of each behavioral type as P3. We restricted
921 these analyses to include only Southern Africa flies for which we had known behavioral
922 phenotypes, and excluded phenotyped flies from the HD lineage, as these genetic
923 lineages exhibit substantial population structure (Figure 1; Figure S5). We then
924 performed a *t*-test with f_G as the dependent variable and the identity of P3 (i.e. lines that
925 showed behavioral preference versus those that did not) as the dependent variable.

926 *Determining the global distribution of previously identified incompatibilities*

927 We characterized patterns of differentiation for loci that have previously been implicated
928 in two studies of genetic incompatibilities with *D. melanogaster* [45,46]. First, [46] used
929 a global panel of *D. melanogaster* inbred lines to create synthetically admixed
930 populations from a series of round-robin matings followed by continual inbreeding. This
931 design enabled the identification of pairs of alleles that appear less frequently than
932 expected under random mating and Mendelian segregation in their final recombinant
933 inbred line population (i.e. Genotype Ratio Distortion). Using, a similar premise, but in a
934 naturally admixed population, Pool [45] used patterns of linkage disequilibrium in the SE
935 United States to assess pairs of alleles that occur together less frequently than

936 expected based on their allele frequencies (i.e. Ancestry Disequilibrium). [45] also
937 determined that many of these loci were highly differentiated between Africa and
938 Europe, using populations from West Africa and France, respectively.

939 Elevated differentiation of these putative incompatibility alleles between West
940 Africa and France may stem from multiple evolutionary scenarios, and differentiating
941 these scenarios can help elucidate the geographic distribution and potential origins of
942 putative incompatibilities within *D. melanogaster*. Here, we aim to differentiate two
943 potential scenarios: First, putative incompatibilities between Europe and Africa may
944 have arisen with or after the Out of Africa expansion and thus may represent more
945 recently derived incompatibilities (i.e. within the last 10-23kya; [5,20,22]). Under this
946 scenario, we predict that differentiation at putative incompatibility loci should be low
947 between genetic lineages in Africa, but high between Europe and all African
948 populations. Second, it is also plausible that putative incompatibilities between Europe
949 and West Africa are very old and also exist between West Africa and Southern Africa,
950 with shared ancestry or subsequent introgression explaining allele sharing between
951 Europe and Southern Africa. Under this scenario, we expect that differentiation should
952 be high between West Africa and both Europe and Southern Africa, but relatively low
953 between Europe and Southern Africa.

954 To differentiate these scenarios, we used the F_{ST} windows from above for all
955 pairwise comparisons of Southern Africa, West Africa, and Europe. We first ask if
956 putative incompatibility loci have elevated divergence relative to the whole genome for
957 any comparison using an ANOVA with Type III SS using the *car* library in *R* [106].
958 Specifically, F_{ST} was the dependent variable, and locus type (genome-wide, loci

959 identified by [46], or loci identified by [45]), population pair (Southern Africa-West Africa,
960 Europe-Southern Africa, or West Africa-Europe), and their interaction were the
961 independent variables. We then used the *emmeans* package in R [110] to determine
962 significance for specific contrasts. Second, we assessed the history and distribution of
963 individual pairs of loci and differentiate the two evolutionary scenarios outlined above.
964 For these analyses, we identified putative incompatibility pairs in which both interacting
965 loci have the predicted patterns of differentiation. We define highly differentiated loci as
966 those with F_{ST} values within the top 2.5% of F_{ST} for that population pair.

967

968 **ACKNOWLEDGEMENTS**

969 We graciously thank the past and present members of the Matute lab, as well as Kieran
970 Samuk for thoughtful discussions on this project. This work was funded by an NIH R01
971 to DRM (R01GM121750) and an NIH R35 (R35GM124701) to BSC.

972

973 **COMPETING INTERESTS**

974 The authors declare no conflict of interest.

975

976 **DATA AVAILABILITY**

977 All new raw sequence data will be made available on the NCBI SRA upon acceptance
978 of this manuscript. Current SRA codes are noted in Table S1. Raw behavioral data will

979 be uploaded to Dryad upon acceptance of this manuscript.

980

981 **REFERENCES**

982 1. Morgan TH. 1911 Random Segregation Versus Coupling in Mendelian Inheritance.
983 *Science* **34**, 384–384.

984 2. Morgan TH. 1910 Sex Limited Inheritance in *Drosophila*. *Science* **32**, 120–122.

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

987 4. Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehermann G, Becher PG, Pool JE,
988 Stensmyr MC. 2018 Wild African *Drosophila* melanogaster Are Seasonal
989 Specialists on Marula Fruit. *Curr. Biol.* **28**, 3960–3968.

990 5. Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV,
991 Stensmyr MC, Pool JE, Larracuente A. 2020 Recurrent Collection of *Drosophila*
992 *melanogaster* from Wild African Environments and Genomic Insights into Species
993 History. *Mol. Biol. Evol.* **37**, 627–638.

994 6. Kopp A, Duncan I, Carroll SB. 2000 Genetic Control and Evolution of Sexually
995 Dimorphic Characters in *Drosophila*. *Nature* **408**, 553–559.

996 7. McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, Payre F, Stern DL.
997 2007 Morphological Evolution Through Multiple Cis-regulatory Mutations at a Single
998 gene. *Nature* **448**, 587–590.

999 8. Montooth KL, Marden JH, Clark AG. 2003 Mapping Determinants of Variation in
1000 Energy Metabolism, Respiration and Flight in *Drosophila*. *Genetics* **165**, 623–635.

1001 9. Ding Y, Berrocal A, Morita T, Longden KD, Stern DL. 2016 Natural Courtship Song
1002 Variation Caused by an Intronic Retroelement in an Ion Channel Gene. *Nature* **536**,
1003 329–332.

1004 10. Markow TA. 2015 The secret lives of *Drosophila* flies. *Elife* **4**, e06793.

1005 11. Harshman LG, Hoffmann AA. 2000 Laboratory Selection Experiments Using
1006 *Drosophila*: What Do They Really Tell Us? *Trends Ecol. Evol.* **15**, 32–36.

1007 12. White NJ, Snook RR, Eyres I. 2020 The Past and Future of Experimental
1008 Speciation. *Trends Ecol. Evol.* **35**, 10–21.

1009 13. Begun DJ, Aquadro CF. 1993 African and North American populations of
1010 *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550.

1011 14. Dieringer D, Nolte V, Schlötterer C. 2005 Population structure in African *Drosophila*
1012 *melanogaster* revealed by microsatellite analysis. *Mol. Ecol.* **14**, 563–573.

1013 15. Pool JE, Aquadro CF. 2006 History and Structure of Sub-Saharan Populations of
1014 *Drosophila melanogaster*. *Genetics* **174**, 915–929.

1015 16. Pool JE *et al.* 2012 Population Genomics of Sub-Saharan *Drosophila*
1016 *melanogaster*: African Diversity and Non-African Admixture. *PLoS Genet.* **8.**
1017 E1003080.

1018 17. Schlötterer C, Neumeier H, Sousa C, Nolte V. 2006 Highly Structured Asian
1019 *Drosophila melanogaster* Populations: A New Tool for Hitchhiking Mapping?
1020 *Genetics* **172**, 287–292.

1021 18. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA,
1022 Langley CH, Pool JE. 2015 The *Drosophila* Genome Nexus: a Population Genomic
1023 Resource of 623 *Drosophila melanogaster* Genomes, Including 197 From a Single
1024 Ancestral Range Population. *Genetics* **199**, 1229–1241.

1025 19. Adrion JR, Hahn MW, Cooper BS. 2015 Revisiting Classic Clines in *Drosophila*
1026 *melanogaster* in the Age of Genomics. *Trends Genet.* **31**, 434–444.

1027 20. Grenier JK, Roman Arguello J, Moreira MC, Gottipati S, Mohammed J, Hackett SR,
1028 Boughton R, Greenberg AJ, Clark AG. 2015 Global Diversity Lines-a Five-
1029 Continent Reference Panel of Sequenced *Drosophila melanogaster* Strains. *G3:*
1030 *Genes, Genomes, Genetics* **5**, 593–603.

1031 21. Baudry E, Viginier B, Veuille M. 2004 Non-African Populations of *Drosophila*
1032 *melanogaster* Have a Unique Origin. *Mol. Biol. Evol.* **21**, 1482–1491.

1033 22. Arguello JR, Laurent S, Clark AG, Gaut B. 2019 Demographic History of the
1034 Human Commensal *Drosophila melanogaster*. *Genome Biol. Evol.* **11**, 844–854.

1035 23. Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005 Multilocus Patterns
1036 of Nucleotide Variability and the Demographic and Selection History of *Drosophila*
1037 *melanogaster* Populations. *Genome Res.* **15**, 790–799.

1038 24. Thornton K, Andolfatto P. 2006 Approximate Bayesian Inference Reveals Evidence
1039 for a Recent, Severe Bottleneck in a Netherlands Population of *Drosophila*
1040 *melanogaster*. *Genetics* **172**, 1607–1619.

1041 25. Kauer M, Dieringer D, Schlötterer C. 2003 Nonneutral Admixture of Immigrant
1042 Genotypes in African *Drosophila melanogaster* Populations from Zimbabwe. *Mol.*
1043 *Biol. Evol.* **20**, 1329–1337.

1044 26. Capy P, Veuille M, Paillette M, Jallon JM, Vouidibio J, David JR. 2000 Sexual
1045 Isolation of Genetically Differentiated Sympatric Populations of *Drosophila*
1046 *melanogaster* in Brazzaville, Congo: The First Step Towards Speciation? *Heredity*
1047 **84**, 468–475.

1048 27. Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D. 2015 Population Genomic
1049 Analysis Uncovers African and European Admixture in *Drosophila melanogaster*
1050 Populations from the South-Eastern United States and Caribbean Islands. *Mol.*
1051 *Ecol.* **24**, 1499–1509.

1052 28. Yukilevich R, True JR. 2008 Incipient Sexual Isolation Among Cosmopolitan
1053 *Drosophila melanogaster* Populations. *Evolution* **62**, 2112–2121.

1054 29. Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S. 2013 Demographic
1055 Inference reveals African and European Admixture in the North American
1056 *Drosophila melanogaster* Population. *Genetics* **193**, 291–301.

1057 30. Caracristi G, Schlötterer C. 2003 Genetic Differentiation Between American and
1058 European *Drosophila melanogaster* Populations Could be Attributed to Admixture
1059 of African Alleles. *Mol. Biol. Evol.* **20**, 792–799.

1060 31. Hollocher H, Ting CT, Wu ML, Wu CI. 1997 Incipient Speciation by Sexual Isolation
1061 in *Drosophila melanogaster*: Extensive Genetic Divergence Without Reinforcement.
1062 *Genetics* **147**, 1191–1201.

1063 32. Ting C-T, Takahashi A, Wu C-I. 2001 Incipient Speciation by Sexual Isolation in
1064 *Drosophila*: Concurrent Evolution at Multiple Loci. *Proc. Natl. Acad. Sci. U. S. A.*
1065 **98**, 6709–6713.

1066 33. Wu CI, Hollocher H, Begun DJ, Aquadro CF, Xu Y, Wu ML. 1995 Sexual Isolation
1067 in *Drosophila melanogaster*: a Possible Case of Incipient Speciation. *Proc. Natl.*
1068 *Acad. Sci. U. S. A.* **92**, 2519–2523.

1069 34. Korol A, Rashkovetsky E, Iliadi K, Michalak P, Ronin Y, Nevo E. 2000 Nonrandom
1070 Mating in *Drosophila melanogaster* Laboratory Populations Derived from Closely
1071 Adjacent Ecologically Contrasting Slopes at 'Evolution Canyon'. *Proc. Natl. Acad.*
1072 *Sci. U. S. A.* **97**, 12637–12642.

1073 35. Hollocher H, Ting C-T, Pollack F, Wu C-I. 1997 Incipient Speciation by Sexual
1074 Isolation in *Drosophila melanogaster*: Variation in Mating Preference and
1075 Correlation Between Sexes. *Evolution* **51**, 1175.

1076 36. Sawamura K, Tomaru M. 2002 Biology of Reproductive Isolation in *Drosophila*:
1077 Toward a Better Understanding of Speciation. *Popul. Ecol.* **44**, 0209–0219.

1078 37. Barbash DA. 2010 Ninety Years of *Drosophila melanogaster* Hybrids. *Genetics*
1079 **186**, 1–8.

1080 38. Orr HA, Masly JP, Phadnis N. 2007 Speciation in *Drosophila*: From Phenotypes to
1081 Molecules. *J. Hered.* **98**, 103–110.

1082 39. Laturney M, Moehring AJ. 2012 The Genetic Basis of Female Mate Preference and
1083 Species Isolation in *Drosophila*. *Int. J. Evol. Biol.* **2012**. (doi:10.1155/2012/328392)

1084 40. Allen Orr H. 2005 The Genetic Basis of Reproductive Isolation: Insights From
1085 *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6522–6526.

1086 41. Castillo DM, Barbash DA. 2017 Moving Speciation Genetics Forward: Modern
1087 Techniques Build on Foundational Studies in *Drosophila*. *Genetics* **207**, 825–842.

1088 42. Jiggins CD, Mallet J. 2000 Bimodal Hybrid Zones and Speciation. *Trends Ecol.*
1089 *Evol.* **15**, 250–255.

1090 43. Irwin DE. 2020 Assortative Mating in Hybrid Zones Is Remarkably Ineffective in
1091 Promoting Speciation. *Am. Nat.* **195**, E150–E167.

1092 44. Coughlan JM, Matute DR. 2020 The Importance of Intrinsic Postzygotic Barriers
1093 Throughout the Speciation Process. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**,
1094 20190533.

1095 45. Pool JE. 2015 The Mosaic Ancestry of the *Drosophila* Genetic Reference Panel
1096 and the *D. melanogaster* Reference Genome Reveals a Network of Epistatic
1097 Fitness Interactions. *Mol. Biol. Evol.* **32**, 3236–3251.

1098 46. Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF. 2013 Genetic
1099 Incompatibilities are Widespread Within Species. *Nature* **504**, 135–137.

1100 47. Cutter AD. 2012 The Polymorphic Prelude to Bateson-Dobzhansky-Muller
1101 Incompatibilities. *Trends Ecol. Evol.* **27**, 209–218.

1102 48. Zuellig MP, Sweigart AL. 2018 A Two-Locus Hybrid Incompatibility is Widespread,
1103 Polymorphic, and Active in Natural Populations of *Mimulus*. *Evolution* **72**, 2394–
1104 2405.

1105 49. Sweigart AL, Mason AR, Willis JH. 2007 Natural Variation for a Hybrid
1106 Incompatibility Between Two Species of *Mimulus*. *Evolution* **61**, 141–151.

1107 50. Meisner J, Albrechtsen A. 2018 Inferring Population Structure and Admixture
1108 Proportions in Low-Depth NGS Data. *Genetics* **210**, 719–731.

1109 51. Veller C, Edelman NB, Muralidhar P, Nowak MA. 2019 Recombination, Variance in
1110 Genetic Relatedness, and Selection Against Introgressed DNA. *bioRxiv*. , 846147.
1111 (doi:10.1101/846147)

1112 52. Rouault J-D, Marican C, Wicker-Thomas C, Jallon J-M. 2004 Relations Between
1113 Cuticular Hydrocarbon (HC) Polymorphism, Resistance Against Desiccation and

Breeding Temperature; a Model for HC Evolution in *D. melanogaster* and *D. simulans*. In *Drosophila melanogaster, Drosophila simulans: So Similar, So Different* (eds P Capy, P Gibert, I Boussy), pp. 195–212. Dordrecht: Springer Netherlands.

53. Foley BR, Telonis-Scott M. 2011 Quantitative Genetic Analysis Suggests Causal Association Between Cuticular Hydrocarbon Composition and Desiccation Survival in *Drosophila melanogaster*. *Heredity* **106**, 68–77.

54. Malinsky M, Matschiner M, Svardal H. 2021 Dsuite - Fast D-Statistics and Related Admixture Evidence from VCF Files. *Mol. Ecol. Resour.* **21**, 584–595.

55. Yassin A, Debat V, Bastide H, Gidaszewski N, David JR, Pool JE. 2016 Recurrent Specialization on a Toxic Fruit in an Island *Drosophila* Population. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4771–4776.

56. Carhan A, Allen F, Armstrong JD, Hortsch M, Goodwin SF, O'Dell KMC. 2005 Female Receptivity Phenotype of Icebox Mutants Caused by a Mutation in the L1-Type Cell Adhesion Molecule Neuroglian. *Genes Brain Behav.* **4**, 449–465.

57. Charles BS. 2019 Neural Regions Affecting Female Aggression and Receptivity in *Drosophila melanogaster*. [Master's Thesis, Western University]. Western University Thesis Repository.

58. Schumer M, Brandvain Y. 2016 Determining Epistatic Selection in Admixed Populations. *Mol. Ecol.* **25**, 2577–2591.

59. Tajima F. 1989 Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **123**, 585–595.

60. Lack JB, Yassin A, Sprengelmeyer QD, Johanning EJ, David JR, Pool JE. 2016 Life History Evolution and Cellular Mechanisms Associated with Increased Size in High-Altitude *Drosophila*. *Ecol. Evol.* **6**, 5893–5906.

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

62. Lack JB, Monette MJ, Johanning EJ, Sprengelmeyer QD, Pool JE. 2016 Decanalization of Wing Development Accompanied the Evolution of Large Wings in High-Altitude *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1014–1019.

63. Singh ND, Macpherson JM, Jensen JD, Petrov DA. 2007 Similar Levels of X-linked and Autosomal Nucleotide Variation in African and Non-African Populations of *Drosophila melanogaster*. *BMC Evol. Biol.* **7**, 202.

64. Stephan W, Li H. 2007 The Recent Demographic and Adaptive History of *Drosophila melanogaster*. *Heredity* **98**, 65–68.

1150 65. Payseur BA, Krenz JG, Nachman MW. 2004 Differential Patterns of Introgression
1151 Across the X Chromosome in a Hybrid Zone Between Two Species of House Mice.
1152 *Evolution* **58**, 2064.

1153 66. Geraldes A, Carneiro M, Delibes-Mateos M, Villafuerte R, Nachman MW, Ferrand
1154 N. 2008 Reduced Introgression of the Y Chromosome Between Subspecies of the
1155 European Rabbit (Oryctolagus cuniculus) in the Iberian Peninsula. *Mol. Ecol.* **17**,
1156 4489–4499.

1157 67. Cortés-Ortiz L, Nidiffer MD, Hermida-Lagunes J, García-Orduña F, Rangel-Negrín
1158 A, Kitchen DM, Bergman TJ, Dias PAD, Canales-Espinosa D. 2019 Reduced
1159 Introgression of Sex Chromosome Markers in the Mexican Howler Monkey
1160 (Alouatta palliata × A. pigra) Hybrid Zone. *Int. J. Primatol.* **40**, 114–131.

1161 68. Turissini DA, Matute DR. 2017 Fine Scale Mapping of Genomic Introgressions
1162 Within the *Drosophila yakuba* Clade. *PLoS Genet.* **13**, e1006971.

1163 69. Garrigan D, Kingan SB, Geneva AJ, Andolfatto P, Clark AG, Thornton KR,
1164 Presgraves DC. 2012 Genome Sequencing Reveals Complex Speciation in the
1165 *Drosophila simulans* Clade. *Genome Res.* **22**, 1499–1511.

1166 70. Fraïsse C, Sachdeva H. 2020 The Rates of Introgression and Barriers to Genetic
1167 Exchange Between Hybridizing Species: Sex Chromosomes vs Autosomes.
1168 *Genetics* **217**. (doi:10.1093/genetics/iyaa025)

1169 71. Dixon G, Kitano J, Kirkpatrick M. 2018 The Origin of a New Sex Chromosome by
1170 Introgression between Two Stickleback Fishes. *Mol. Biol. Evol.* **36**, 28–38.

1171 72. Dufresnes C, Majtyka T, Baird SJE, Gerchen JF, Borzée A, Savary R, Ogielska M,
1172 Perrin N, Stöck M. 2016 Empirical Evidence for Large X-Effects in Animals with
1173 Undifferentiated Sex Chromosomes. *Sci. Rep.* **6**, 1–7.

1174 73. Macholán M, Baird SJE, Munclinger P, Dufková P, Bímová B, Piálek J. 2008
1175 Genetic Conflict Outweighs Heterogametic Incompatibility in the Mouse Hybrid
1176 Zone? *BMC Evol. Biol.* **8**, 271.

1177 74. Hartmann FE, Rodríguez de la Vega RC, Gladieux P, Ma W-J, Hood ME, Giraud T.
1178 2020 Higher Gene Flow in Sex-Related Chromosomes than in Autosomes during
1179 Fungal Divergence. *Mol. Biol. Evol.* **37**, 668–682.

1180 75. Ottenburghs J. 2021 The Genic View of Hybridization in the Anthropocene. *Evol.*
1181 *Appl.* (doi:10.1111/eva.13223)

1182 76. Thompson KA, Rieseberg LH, Schluter D. 2018 Speciation and the City. *Trends*
1183 *Ecol. Evol.* **33**, 815–826.

1184 77. Allendorf FW, Leary RF, Spruell P, Wenburg JK. 2001 The Problems with Hybrids:
1185 Setting Conservation Guidelines. *Trends Ecol. Evol.* **16**, 613–622.

1186 78. Hulme-Beaman A, Dobney K, Cucchi T, Searle JB. 2016 An Ecological and
1187 Evolutionary Framework for Commensalism in Anthropogenic Environments.
1188 *Trends Ecol. Evol.* **31**, 633–645.

1189 79. Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlotterer C.
1190 2004 World-wide Survey of an *Accord* Insertion and its Association with DDT
1191 Resistance in *Drosophila melanogaster*. *Mol. Ecol.* **13**, 2491–2504.

1192 80. Daborn PJ *et al.* 2002 A Single *p450* Allele Associated with Insecticide Resistance
1193 in *Drosophila*. *Science* **297**, 2253–2256.

1194 81. Ravinet M, Elgvin TO, Trier C, Aliabadian M, Gavrilov A, Sætre G-P. 2018
1195 Signatures of Human-Commensalism in the House Sparrow Genome. *Proc. Biol.*
1196 *Sci.* **285**, 20181246.

1197 82. Guiry E, Buckley M. 2018 Urban Rats have Less Variable, Higher Protein Diets.
1198 *Proc. Biol. Sci.* **285**. (doi:10.1098/rspb.2018.1441)

1199 83. Harris SE, Munshi-South J, Obergfell C, O'Neill R. 2013 Signatures of Rapid
1200 Evolution in Urban and Rural Transcriptomes of White-footed Mice (*Peromyscus*
1201 *leucopus*) in the New York Metropolitan Area. *PLoS One* **8**, e74938.

1202 84. Axelsson E *et al.* 2013 The Genomic Signature of Dog Domestication Reveals
1203 Adaptation to a Starch-rich Diet. *Nature* **495**, 360–364.

1204 85. Harpak A, Garud N, Rosenberg NA, Petrov DA, Combs M, Pennings PS, Munshi-
1205 South J. 2020 Genetic Adaptation in New York City Rats. *Genome Biol. Evol.* **13**.
1206 (doi:10.1093/gbe/evaa247)

1207 86. Harris SE, O'Neill RJ, Munshi-South J. 2015 Transcriptome Resources for the
1208 White-footed Mouse (*Peromyscus leucopus*): New Genomic Tools for Investigating
1209 Ecologically Divergent Urban and Rural Populations. *Mol. Ecol. Resour.* **15**, 382–
1210 394.

1211 87. Grillet M, Everaerts C, Houot B, Ritchie MG, Cobb M, Ferveur J-F. 2012 Incipient
1212 Speciation in *Drosophila melanogaster* Involves Chemical Signals. *Sci. Rep.* **2**, 1–
1213 11.

1214 88. Grillet M, Ferveur JF, Everaerts C. 2018 Behavioural Elements and Sensory Cues
1215 Involved in Sexual Isolation Between *Drosophila melanogaster* strains. *Royal*
1216 *Society Open Science* **5**. (doi:10.1098/rsos.172060)

1217 89. Jin B, Barbash DA, Castillo DM. 2020 Divergent Selection on Behavioral and
1218 Chemical Traits Contributes to Isolation Between Populations of *Drosophila*
1219 *melanogaster*. *bioRxiv*, 2020.09.28.316703. (doi:10.1101/2020.09.28.316703)

1220 90. Dembeck LM, Böröczky K, Huang W, Schal C, Anholt RRH, Mackay TFC. 2015
1221 Genetic architecture of Natural Variation in Cuticular Hydrocarbon Composition in

1222 *Drosophila melanogaster*. *Elife* **4**, e09861.

1223 91. McBride SM, Giuliani G, Choi C, Krause P, Correale D, Watson K, Baker G, Siwicki
1224 KK. 1999 Mushroom Body Ablation Impairs Short-term Memory and Long-term
1225 Memory of Courtship Conditioning in *Drosophila melanogaster*. *Neuron* **24**, 967–
1226 977.

1227 92. Coyne JA, Wicker-Thomas C, Jallon JM. 1999 A Gene Responsible for a Cuticular
1228 Hydrocarbon Polymorphism in *Drosophila melanogaster*. *Genet. Res.* **73**, 189–203.

1229 93. Combs PA, Krupp JJ, Khosla NM, Bua D, Petrov DA, Levine JD, Fraser HB. 2018
1230 Tissue-Specific cis-Regulatory Divergence Implicates eloF in Inhibiting Interspecies
1231 Mating in *Drosophila*. *Curr. Biol.* **28**, 3969–3975.e3.

1232 94. Chen D, Qu C, Bjorum SM, Beckingham KM, Hewes RS. 2014 Neuronal
1233 Remodeling During Metamorphosis Is Regulated by the alan shepard (*shep*) Gene
1234 in *Drosophila melanogaster*. *Genetics* **197**, 1267–1283.

1235 95. Griffith LC, Wang J, Zhong Y, Wu CF, Greenspan RJ. 1994 Calcium/calmodulin-
1236 dependent protein kinase II and Potassium Channel Subunit *eag* Similarly Affect
1237 Plasticity in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10044–10048.

1238 96. Greenspan RJ, Ferveur J-F. 2000 Courtship in *Drosophila*. *Annu. Rev. Genet.* **34**,
1239 205–232.

1240 97. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018 UFBoot2:
1241 Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522.

1242 98. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017
1243 ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat.*
1244 *Methods* **14**, 587–589.

1245 99. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a Fast and
1246 Effective Stochastic Algorithm for Estimating Maximum-likelihood Phylogenies. *Mol.*
1247 *Biol. Evol.* **32**, 268–274.

1248 100. Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018 ASTRAL-III: Polynomial Time
1249 Species Tree Reconstruction from Partially Resolved Gene Trees. *BMC*
1250 *Bioinformatics* **19**, 153.

1251 101. Korunes KL, Samuk K. 2021 *pixy*: Unbiased Estimation of Nucleotide Diversity and
1252 Divergence in the Presence of Missing Data. *Mol. Ecol. Resour.* **21**, 1359–1368.

1253 102. Danecek P *et al.* 2011 The Variant Call Format and VCFtools. *Bioinformatics* **27**,
1254 2156–2158.

1255 103. Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N. 2013 SweeD: Likelihood-based
1256 Detection of Selective Sweeps in Thousands of Genomes. *Mol. Biol. Evol.* **30**,

2224–2234.

104. Martin SH, Davey JW, Jiggins CD. 2015 Evaluating the use of ABBA–BABA Statistics to Locate Introgressed Loci. *Mol. Biol.*

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

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

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

108. Yi X *et al.* 2010 Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science* **329**, 75–78.

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

110. Russell L. 2018 emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3. 0.

111. dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, FlyBase Consortium. 2015 FlyBase: Introduction of the *Drosophila melanogaster* Release 6 Reference Genome Assembly and Large-Scale Migration of Genome Annotations. *Nucleic Acids Res.* **43**, D690–7.

112. Li H, Durbin R. 2009 Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Bioinformatics* **25**, 1754–1760.

113. McKenna A *et al.* 2010 The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-generation DNA Sequencing Data. *Genome Res.* **20**, 1297–1303.

114. Li H *et al.* 2009 The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **25**, 2078–2079.

115. Müller, H. 1942 Isolating Mechanisms, Evolution, and Temperature. *Biol. Symp.* **6**, 71–125.

116. Dobzhansky, T. 1982 Genetics and the Origin of Species. New York, NY: Columbia University Press.

1290 **FIGURE CAPTIONS**

1291 **FIGURE 1. Broad sampling of *Drosophila melanogaster* from Subtropical Africa**
1292 **reveals substantial cryptic genetic structure.** *(**A**) Geographic sampling of 420*
1293 *genomes from a global distribution of D. melanogaster; pie charts represent the average*
1294 *ancestry determined by PCAngsd from that sampling local. (**B**) Zoom-in panel from (A),*
1295 *focusing on Subtropical and East African accessions. (**C**) PCA of all genomes, colors*
1296 *indicate the genetic lineage of the sample, defined herein. We also denote samples*
1297 *from North and South Africa (although they reside within OOA and Southern Africa,*
1298 *respectively). (**D**) Phylogeny of all accessions with an ancestry plot denoting their*
1299 *average genomic composition of each sample. Labels to the right of the ancestry plot*
1300 *indicate how samples cluster into the six genetic lineages that we identify using PCA, K-*
1301 *means clustering, and phylogenetic reconstruction.*
1302
1303 **FIGURE 2. Genome-wide statistics for each genetic lineage.** *Average (**A**) nucleotide*
1304 *diversity (π) and (**B**) Tajima's D for each genetic lineage.*
1305
1306
1307 **FIGURE 3: Patterns of gene flow between different African and non-African**
1308 **lineages of D. melanogaster.** *(**A**) Boxplots of Patterson's D for all trios given the*
1309 *phylogeny (((OOA, ((East, West),Ethiopia),HD),* Southern *Africa). P3 is given on the X-*
1310 *axis, with the identity of the P2 population denoted by the color of the boxplot. Values of*
1311 *D represent the range of values over multiple P1 populations. Comparisons that yielded*
1312 *a significant Patterson's D based on a standard block jackknife procedure [54] are*
1313 *denoted with an asterisk, otherwise they were deemed Non-Significant (NS). (**B**) Mean*
1314 *and standard errors of 20-SNP windows of f_{dM} from across the genome for each African*
1315 *genetic lineage showing significant introgression with either other African or non-Africa*
1316 *lineages. The OOA lineage is split into those from the SE United States (SE-US) and all*
1317 *others (OOA). Significance was determined by* ANOVA*s with Type III SS: NS= Not*
1318 *Significant, * = 0.01<p<0.05. (**C**) Weighted topologies for three configurations of OOA,*
1319 *HD, and Africa (denoted by colored phylogenies to the left).*
1320
1321
1322 **FIGURE 4: Geographic and genetic dissection of female preference in Subtropical**
1323 **Africa.** *(**A**) Frequency of isofemale lines that show significant female mate choice in*
1324 *two-male choice experiments per sampling locale. (**B**) Average F_{ST} is elevated in the top*
1325 *1% PBE regions between the Southern Africa lineage (regardless of mate preference*
1326 *phenotype) and all other geographic regions. (**C**) Cartoon illustrating that the PBE*
1327 *statistic can infer branch-specific evolution of the focal population (in this case, flies with*
1328 *significant female mate preference (red, "pref"), relative to flies with no preference (blue,*
1329 *"no pref") and a third population of RAL)). (**D**) PBE of D. melanogaster lines with strong*
1330 *mate preference, measured in 1KB windows across the five major chromosome arms of*
1331 *D. melanogaster. Grey horizontal line indicates the top 1% threshold. Colored dots*
1332 *indicate a sample of genes with functions relevant to female mate preference.*
1333
1334

1335 **FIGURE 5: Global differentiation of putative incompatibility loci.** *(**A**) Distribution of*
1336 *F_{ST} for incompatibility loci from CD [46], Pool [45] and genome-wide for three*
1337 *comparisons: Southern Africa vs Europe, West Africa vs Southern Africa, and West*
1338 *Africa vs Europe. (**B**) Cartoon depiction of a scan for incompatibility loci that segregate*
1339 *between African lineages (yellow star) or likely arose during or after the Out of Africa*
1340 *expansion (blue star). The Y-axis denotes the difference in F_{ST} between Europe and*
1341 *Southern African (ES) and West and Southern Africa (WS), with more positive values*
1342 *indicating that F_{ST} is greater between Europe and Southern Africa than within Africa.*
1343 *Red points denote loci that are also highly differentiated between Europe and West*
1344 *Africa. Loci that are highly differentiated between Europe and both African lineages are*
1345 *putatively newer incompatibilities, while loci that are highly differentiated within Africa,*
1346 *but shared between Southern Africa and Europe may represent older incompatibilities.*
1347 *(**C,D**) Two zoomed in windows representing one incompatibility pair, as identified by*
1348 *Pool [45]. In panel (C) differentiation is elevated between Europe and Southern Africa*
1349 *and West Africa and Europe (but not West and Southern Africa) for loci within the gene*
1350 Piezo*, indicating this allele may be more recently derived in OOA populations. In (D) we*
1351 *show the corresponding locus: again, differentiation is elevated between Europe and*
1352 *each of West and Southern Africa (but not within Africa) in a window containing no*
1353 *genes, but slight up and down stream of* comm *and* comm2*, respectively. We also*
1354 *highlight the approximate position of* AGO2*, the hypothesized interacting partner (based*
1355 *on [45]), and show there is no increased divergence in this region.*

1356

1357

1358

1359

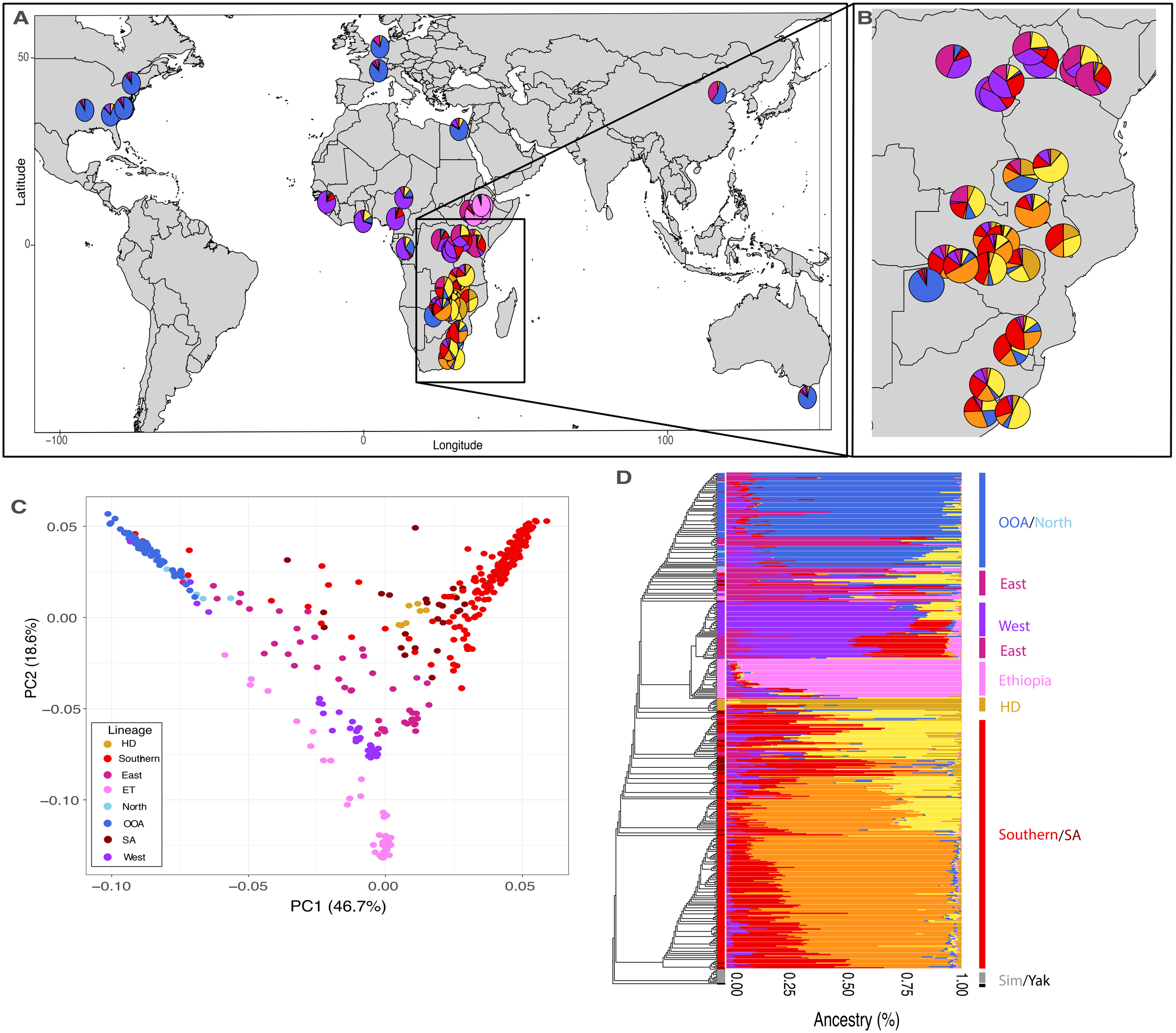
1360

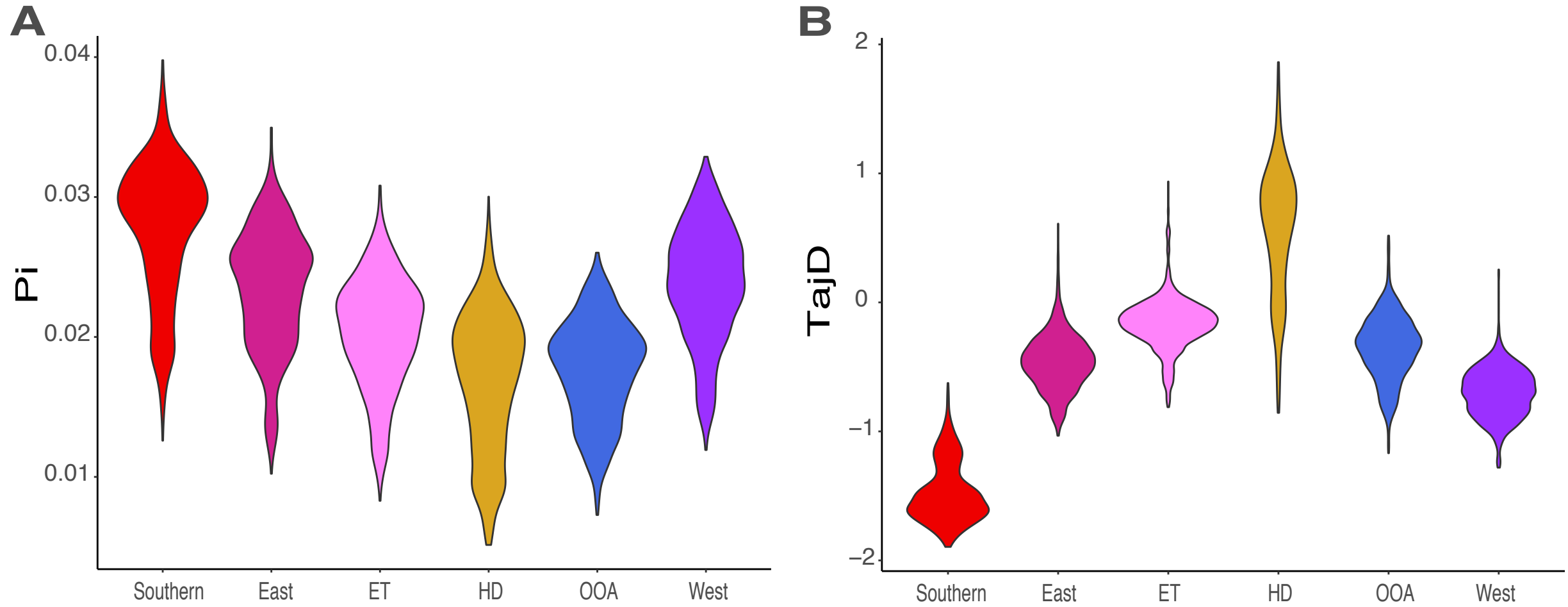
1361

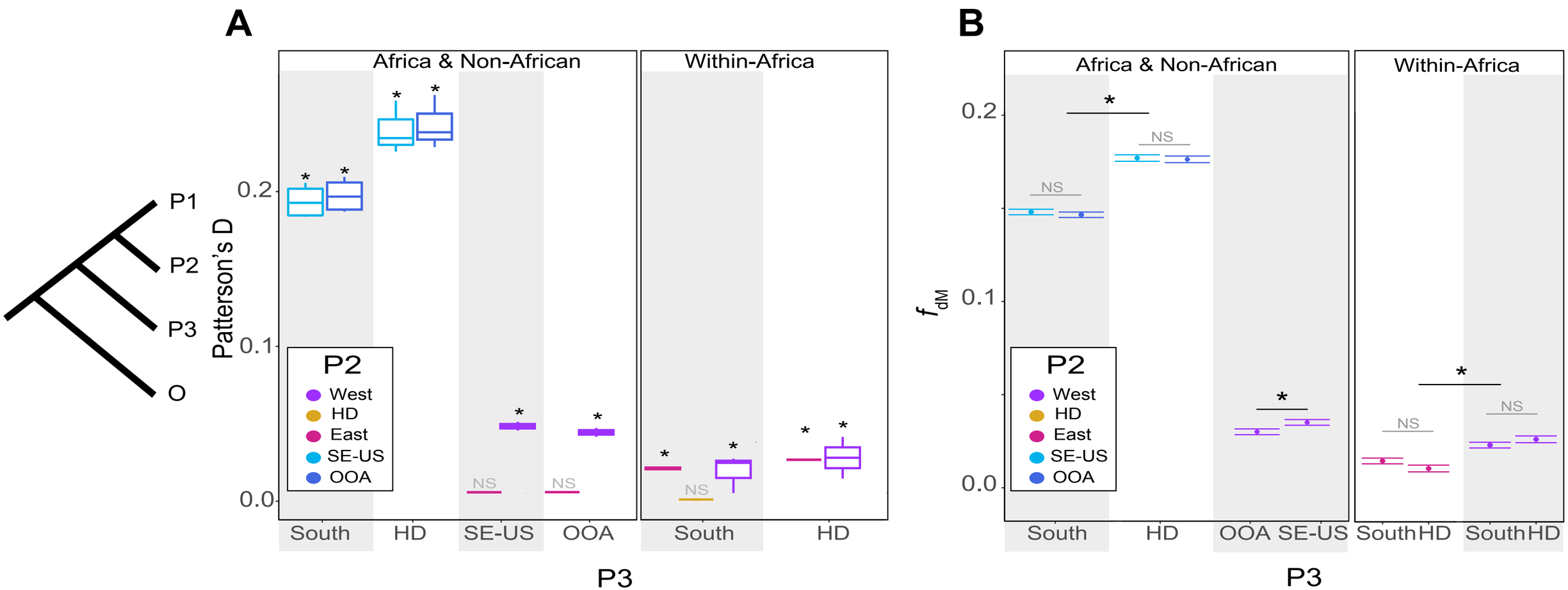
1362

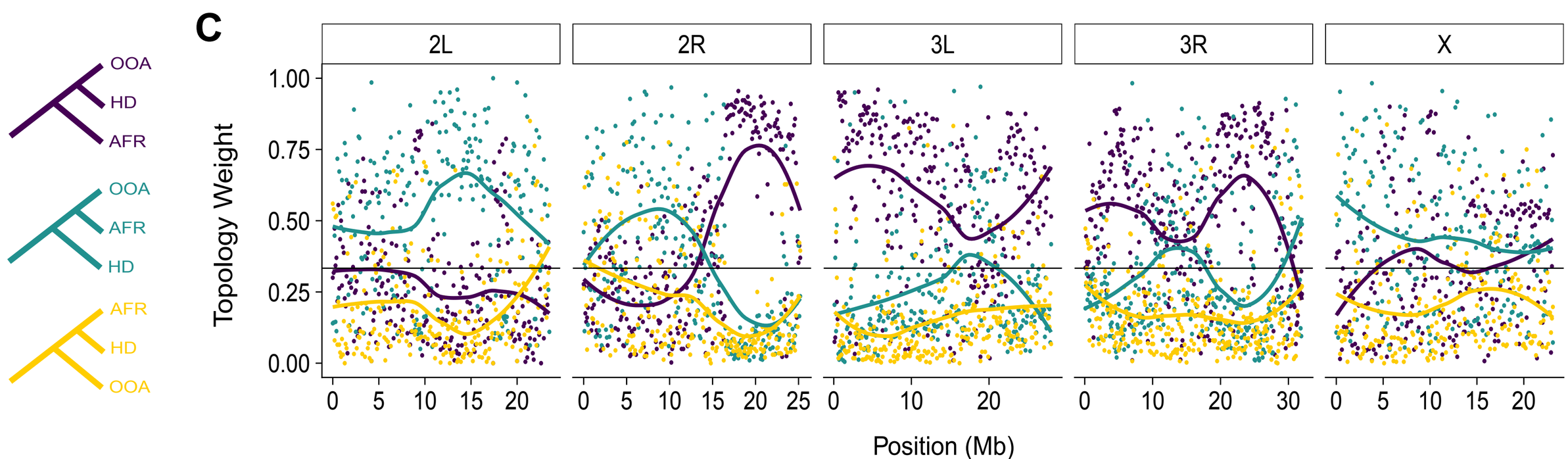
1363

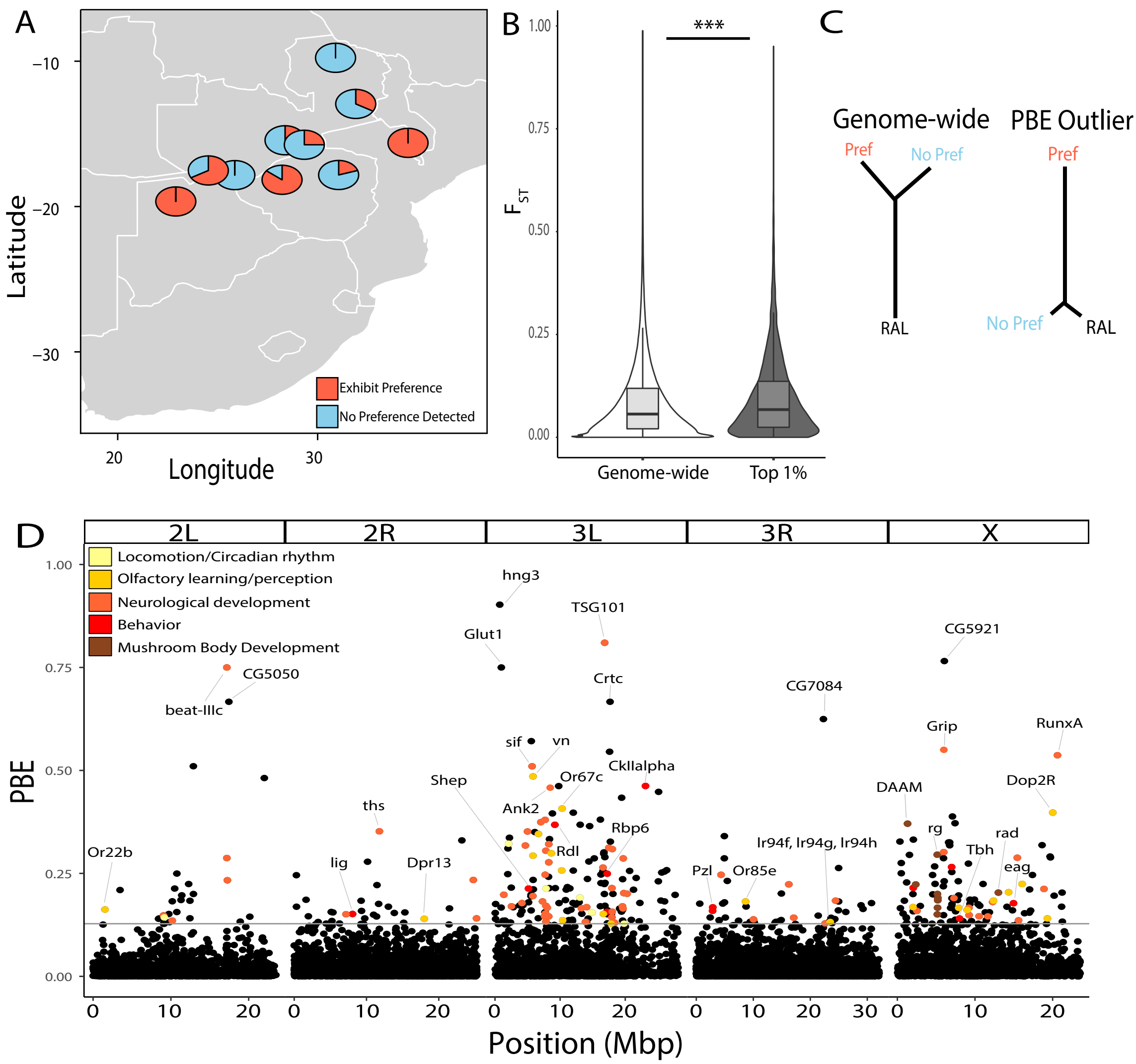
1364

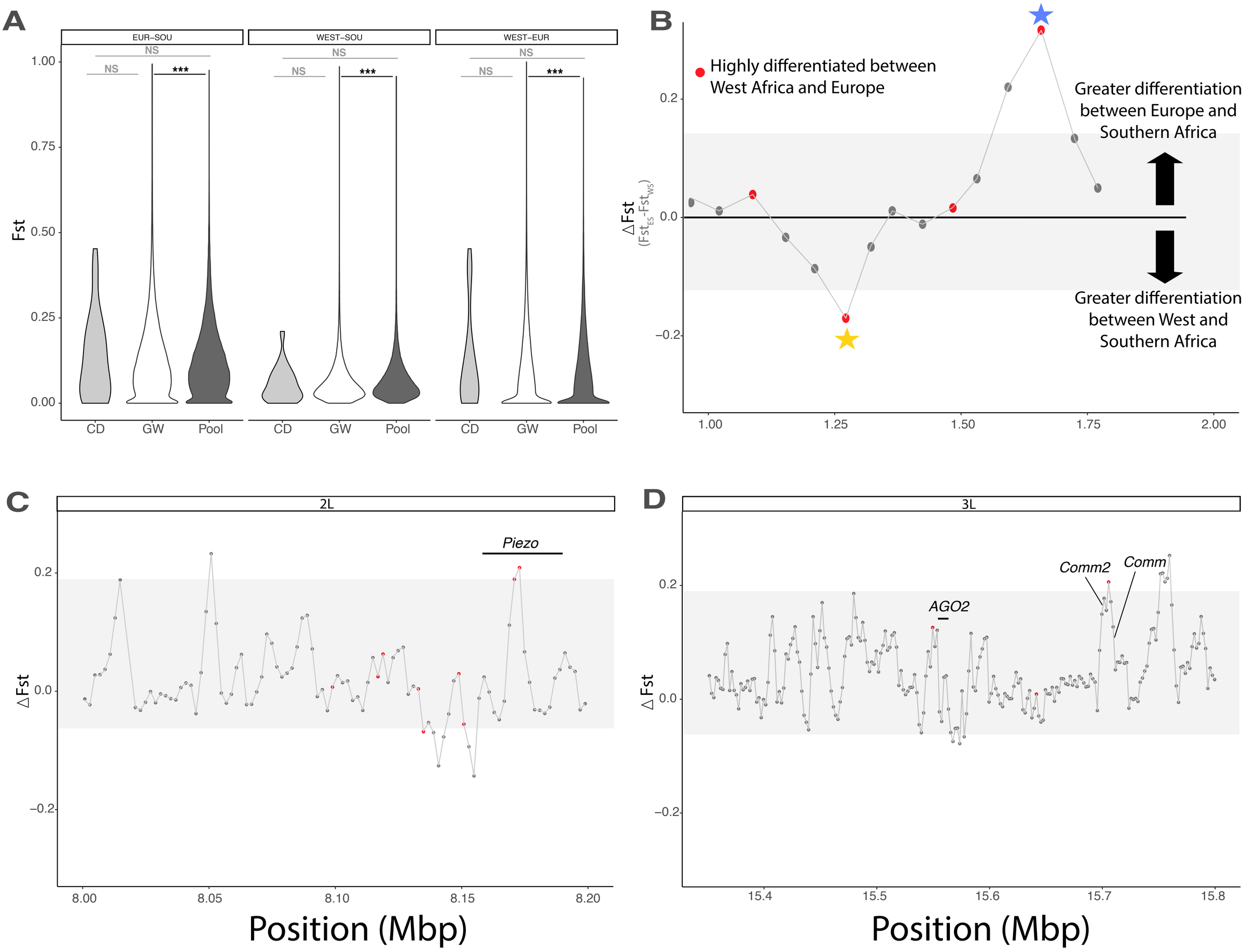

A
B
C
D
Latitude
Longitude
50
0
−100
0
100
PC2 (18.6%)
PC1 (46.7%)
0.05
0.00
−0.05
−0.10
Lineage
HD
Southern
East
ET
North
OOA
SA
West
OOA/North
East
West
East
Ethiopia
HD
Southern/SA
Sim/Yak
0.00
0.25
0.50
0.75
1.00
Ancestry (%)


A
Pi
0.04
0.03
0.02
0.01
Southern
East
ET
HD
OOA
West
B
TajD
2
1
0
−1
−2
Southern
East
ET
HD
OOA
West


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
South
HD
SE-US
OOA
South
HD
P3
B
Africa & Non-African
Within-Africa
f_{dM}
0.0
0.1
0.2
P2
West
HD
East
SE-US
OOA
NS
South
HD
OOA
SE-US
South
HD
South
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–SOU
WEST–SOU
WEST–EUR
NS
***
Fst
1.00
0.75
0.50
0.25
0.00
CD
GW
Pool
B
Highly differentiated between West Africa and Europe
Greater differentiation between Europe and Southern Africa
Greater differentiation between West and Southern Africa
ΔFst
(Fst_ES-Fst_WS)
0.2
0.0
−0.2
1.00
1.25
1.50
1.75
2.00
C
2L
Piezo
ΔFst
8.00
8.05
8.10
8.15
8.20
Position (Mbp)
D
3L
AGO2
Comm2
Comm
15.4
15.5
15.6
15.7
15.8
Position (Mbp)