1    ARTICLE

2

3    **Parallel Evolution of Cold Tolerance Within *Drosophila melanogaster***

4

5    John E. Pool*, Dylan T. Braun, Justin B. Lack[1]

6

7    Laboratory of Genetics, University of Wisconsin-Madison

8    Madison, WI, USA, 53705

9

10    *Corresponding author:

11    John E. Pool

12    Laboratory of Genetics

13    University of Wisconsin-Madison

14    425-G Henry Mall

15    Madison, WI 53706

16    Phone: 608-265-1036

17    email: jpool@wisc.edu

18

19    [1]Current address:

20    Center for Cancer Research

21    National Cancer Institute, NIH

22    Bethesda, MD 20892-1201

23

1

24    ABSTRACT

25

26          *Drosophila melanogaster* originated in tropical Africa before expanding into

27    strikingly different temperate climates in Eurasia and beyond.  Here, we show that elevated

28    cold tolerance has arisen at least three times within this species:  beyond the well-studied

29    non-African case, we show that populations from the highlands of Ethiopia and South

30    Africa have significantly increased cold tolerance as well.  We observe greater cold

31    tolerance in outbred versus inbred flies, but only in populations with higher inversion

32    frequencies.  Each cold-adapted population shows lower inversion frequencies than a

33    closely-related warm-adapted population, suggesting that inversion frequencies may

34    decrease with altitude in addition to latitude.  Using the $F_{ST}$-based "Population Branch

35    Excess" statistic (*PBE*), we found only limited evidence for parallel genetic differentiation at

36    the scale of ~4 kb windows, specifically between Ethiopian and South African cold-adapted

37    populations.  And yet, when we looked for single nucleotide polymorphisms (SNPs) with

38    codirectional frequency change in two or three cold-adapted populations, strong genomic

39    enrichments were observed from all comparisons.  These findings could reflect an

40    important role for selection on standing genetic variation leading to "soft sweeps".  One

41    SNP showed sufficient codirectional frequency change in all cold-adapted populations to

42    achieve experiment-wide significance: an intronic variant in the synaptic gene *Prosap*.

43    More generally, proteins involved in neurotransmission were enriched as potential targets

44    of parallel adaptation.  The ability to study cold tolerance evolution in a parallel framework

45    will enhance this classic study system for climate adaptation.

46    INTRODUCTION

47

48        Thermal tolerance is a critical determinant of species ranges and organismal adaptation

49    to environments.  Adaptation to cold climates and its genetic basis has been a topic of

50    particular scientific interest, including studies of humans (*e.g.* Hancock *et al.* 2010) and

51    *Drosophila* (*e.g.* Hoffmann *et al.* 2003; Ayrinhac *et al.* 2004; Svetec *et al.* 2011).  Unraveling

52    the genetic architecture and biological mechanisms underlying insect cold tolerance may

53    be especially relevant for identifying factors that currently constrain the geographic

54    distributions of medically important insects such as disease vectors.

55        Originating from a relatively warm sub-Saharan ancestral range (Lachaise *et al.* 1988;

56    Pool *et al.* 2012), *D. melanogaster* ultimately became associated with human settlements

57    and spread throughout the world, including environments much colder than its original

58    habitat.  Evidence for this species' genetic response to novel thermal environments

59    includes geographic variation in cold tolerance (Hoffmann *et al.* 2003) and the presence of

60    parallel allele frequency clines in North America and Australia (Turner *et al.* 2009).  These

61    clines provide striking examples of rapid evolution, in that North American and Australian

62    populations of *D. melanogaster* appear to have descended from the admixture of European

63    and African source populations within the past century or two (Keller 2007; Bergland *et al.*

64    2016).  Although source population ancestry shows a genome-wide latitude cline on these

65    continents (Kao *et al.* 2015; Bergland *et al.* 2016), loci showing extreme latitude clines

66    could reflect the locally adaptive reassortment of pre-existing genetic variants that differed

67    between the African and European source populations.

3

68    In contrast, the earlier expansion of *D. melanogaster* from Africa into Eurasia, which

69    occurred roughly 10,000 years ago (Thornton and Andolfatto 2006; Duchen *et al.* 2013),

70    may have involved novel genetic adaptation to cold environments.  This adaptation is most

71    commonly studied via comparisons between European and sub-Saharan African

72    populations (*e.g.* Kauer *et al.* 2002; Svetec *et al.* 2011; Božičević *et al.* 2016).  However, the

73    species' sub-Saharan exodus entailed a strong population bottleneck that complicates the

74    population genetic detection of adaptive differences between sub-Saharan and European

75    populations, by stochastically increasing the level of neutral genome-wide differentiation

76    between these populations.  We suggest that the genetic basis of European cold adaptation

77    could alternatively be studied by comparison with a population that had gone through the

78    same bottleneck, but showed less adaptation to cold climates.  Northern African

79    populations appear to be genetically similar to European populations (Dieringer *et al.*

80    2005; Lack *et al.* 2015), and in this study we compare fly strains from France and Egypt.

81    Although populations from outside sub-Saharan Africa are known to offer a range of

82    opportunities for studying cold tolerance evolution, these populations may largely reflect a

83    single origin of cold tolerance.  In contrast, sub-Saharan Africa could offer previously

84    unappreciated independent origins of cold adaptation.  Africa's most extensive highlands

85    are found in Ethiopia, featuring a consistently cool environment.  Seasonally cold

86    environments are found in South Africa due to a combination of higher altitude and

87    latitude.  We therefore study low and high altitude pairs of populations from each of these

88    countries, representing warmer and cooler environments respectively (Figure 1).

89    If confirmed, the existence of cold-tolerant populations in each of these three regions

90    would appear to indicate parallel phenotypic evolution – that is, each cold-adapted

4

91    population being founded independently by warm-adapted progenitors. *D. melanogaster* is

92    thought to originate from the southern portion of Africa (Pool *et al.* 2012), and thus cold-

93    adapted populations from this region could be the most ancient.  However, the species'

94    northward expansion within Africa would have originated far from South Africa.  Instead,

95    populations would have expanded from tropical environments to ultimately occupy

96    Ethiopia and Eurasia.  Population genomic data suggest that the species' colonization of

97    Ethiopia and Eurasia involved separate expansions, in that Ethiopian populations do not

98    appear to be the source of the out-of-Africa emigration (Pool *et al.* 2012).

99         Parallel phenotypic evolution can be explained by at least three distinct genetic

100    scenarios.  First, cold tolerance could evolve from independent mutations in each

101    population, at either the same or different genes.  Second, cold tolerance could arise from a

102    common pool of standing genetic variation present in the ancestral population.  Third, cold

103    tolerance variants could be transmitted from one population to another by migration.

104    Although we argue that cold tolerance was favored independently in each of these regions,

105    gene flow has occurred (Pool *et al.* 2012) and could contribute to cold tolerance evolution.

106         Populations in Europe and in the highlands of Ethiopia and South Africa have each

107    experienced distinct selective pressures related to their abiotic and biotic environments.

108    However, the component of each population's local adaptation that relates to cold tolerance

109    (or correlated selective pressures) could produce a genomic signal of genetically parallel

110    evolution, if the same genes tend to contribute in each population.  After documenting cold

111    tolerance evolution in these three regions, we then tested for parallel genetic

112    differentiation specific to the cold-adapted populations.  Limited evidence of parallel

113    genetic differentiation was obtained from window-based analysis of allele frequency

5

114    change.  However, a strong genomic enrichment of SNPs with codirectional allele frequency

115    change was observed between all cold-adapted populations, including SNPs in genes

116    including membrane-bound and neurotransmission-related genes.

117

118

119    RESULTS

120

121    *Cold tolerance has evolved in three geographic regions*

122         We used three pairs of population samples to search for evidence of the parallel

123    phenotypic evolution of cold tolerance within *D. melanogaster* (Figure 1).  Our assay

124    consisted of the proportion of female flies still standing after 96 hours at 4°C.  Results

125    indicated highly significant increases in cold tolerance for strains originating from colder

126    environments in France, Ethiopia, and South Africa (Table 1; Table S1).  The France strains

127    showed the greatest degree of cold tolerance.  This population represents the adaptation of

128    *D. melanogaster* to temperate Eurasian climates, which has been a rich source of biological

129    insight.  Cold-tolerant European *D. melanogaster* appear to have made substantial genetic

130    contributions to recently founded populations from the Americas and Australia as well

131    (Nunes *et al.* 2008; Bergland *et al.* 2016).  Importantly, our Egypt sample lacks most of the

132    cold tolerance present in France, in spite of these samples' low genetic differentiation (Lack

133    *et al.* 2015).  This pair of European and northern African populations therefore represents

134    an excellent comparison for studies of cold adaptation.

135         The increased cold tolerance detected in Ethiopian and South African populations

136    (Table 1) has not been previously studied, and is likely to reflect independent natural

6

137    selection in each case (Introduction).  The degree of cold tolerance in these populations is

138    less than that observed for France, which might be explained by the presence of seasonally

139    much colder environments in parts of Europe, and perhaps by the smaller geographic size

140    of cool regions in Africa (in light of the balance between local adaptation and gene flow).

141    Still, the Ethiopian and South African population comparisons are both strongly significant,

142    with greater than 20% absolute increases in the fraction of flies standing after 96 hours at

143    4°C.

144

145    *Variable Influence of Inbreeding on Cold Tolerance*

146        In a limited number of cases, we also investigated the cold tolerance of outbred flies

147    from each population.  For these within-population crosses, males from different inbred

148    strains were crossed to females from a single reference strain.  Results were surprisingly

149    population-dependent:  compared with expectations based on the average of the parental

150    inbred strains, three populations had significantly improved cold tolerance after

151    outbreeding (Egypt and the two South African samples), while the other three populations

152    showed no trend in this direction (Table 2; Table S2).  Although the numbers of crosses are

153    modest, this can not account for the divergent pattern observed; the null probability for at

154    least three populations to have a significant increase in outbred tolerance ($P < 0.05$), with

155    the rest showing no trend ($P \geq 0.5$) is quite low ($P = 0.0026$, based on one million

156    randomizations of the direction of change between outbred and midparent tolerance

157    values).  Hence, the influence of outbreeding on cold tolerance vary considerably

158    depending on the population (or based on the specific reference strain used in our crosses).

159    Thus, the effects of inbreeding on tolerance traits may depend not only on the trait

160    examined (Dahlgaard and Hoffmann 2000), but also on the genotypes being studied.  One

161    potential factor in our population-sensitive outbreeding results is inversion frequencies:

162    the three populations with increased cold tolerance after outbreeding were the also the

163    three populations with the highest frequency of inverted chromosomes (see below).  This

164    tentative correlation might be explained by associative overdominance in connection with

165    recessive deleterious alleles carried on inverted chromosomes (Ohta 1971).

166

167    *Parallel inversion frequency change in cold-adapted populations*

168        We showed above that populations in Ethiopia, France, and South Africa have

169    experienced parallel phenotypic evolution toward greater cold tolerance.  When then

170    sought to address two questions regarding the evolutionary genetic mechanisms

171    underlying parallel cold tolerance evolution.  First, is there evidence that the same genes

172    have contributed to adaptive evolution in two or more cold-adapted populations?  And is

173    there evidence that the same variants have been independently selected?

174        Before proceeding to full scans of genomic variation, we assessed whether inversion

175    polymorphisms showed evidence for parallel frequency change in cold-adapted

176    populations.  Based on PCR-testing for eight polymorphic inversions, we found a trend

177    toward lower inversion frequencies in all three cold-adapted populations (Table S3).  In

178    particular, the five most common inversions all showed lower frequency in at least two

179    cold-adapted populations (Figure 2).  Overall, fewer chromosome arms carried inversions

180    in SD relative to SP (32% $vs.$ 37%, $P = 0.26$), EF relative to EA (6% $vs.$ 17%, $P = 0.005$), and

181    FR relative to EG (16% $vs.$ 47%, $P < 0.0001$).  In order to separate inversion-linked

182    frequency changes from the potential influence of selection on specific genes and variants,

8

183    we therefore excluded all genomes carrying an inversion on a given chromosome arm from

184    the analyses described below.  These findings are concordant with past results on

185    latitudinal inversion clines (*e.g.* Lemeunier and Aulard 1992; Kapun *et al.* 2016), and

186    provide evidence for independent decreases in inversion frequency in cool high altitude

187    regions.

188

189    *Genomic window analysis of parallel genetic differentiation*

190    The population genetic signals of adaptive genetic differentiation can be variable.

191    Selective sweeps may be complete or incomplete.  Further, the outcome of positive

192    selection may be described as either a hard sweep (when a single haplotype rises in

193    frequency, due to a new mutation or potentially a rare variant) or as a soft sweep (when

194    multiple haplotypes rise in frequency, due to standing genetic variation or recurrent

195    mutation).  When statistics comparing genetic variation between two populations are

196    applied to chromosomal windows, sensitivity is highest for complete hard sweeps; and

197    depending on the parameters, significant power may persist for incomplete and/or soft

198    sweeps as well (Lange and Pool 2016).  However, soft sweeps involving many unique

199    haplotypes may lead to high genetic differentiation at the causative SNP alone (and perhaps

200    a handful of linked variants), a pattern which may be undetectable by window scans.

201    We therefore pursued both window and single SNP analyses of parallel genetic

202    differentiation.  For both analyses, we used a recently described $F_{ST}$-based statistic known

203    as *Population Branch Excess* (*PBE*; Yassin *et al.* 2016).  Like the related *PBS* (Yi *et al.* 2010),

204    *PBE* uses $F_{ST}$ values among three populations.  Whereas *PBS* simply quantifies the length of

205    the focal population's branch on a genetic distance tree, *PBE* asks whether this *PBS* value is

9

206    unexpectedly high in light of (1) $F_{ST}$ between the two non-focal populations at this locus,

207    and (2) patterns of genetic differentiation across other loci (Materials and Methods). The

208    most relevant difference between *PBE* and *PBS* is their behavior at loci that are subject to

209    high genetic differentiation between all pairs of populations (due to recurrent positive

210    selection not specific to the focal population, or other forces such as background selection).

211    *PBS* should detect such loci as outliers for high genetic differentiation, but *PBE* is designed

212    not to. Instead, *PBE* responds most strongly to cases where genetic differentiation is

213    elevated only between the focal population and each of the others. *PBE* fits our interest in

214    detecting loci that are specifically under selection in the cold-adapted populations, while

215    avoiding false signals of genetically parallel cold adaptation from loci with globally high $F_{ST}$.

216        We first evaluated *PBE* in ~4 kb windows, separately for each population pair, with the

217    ancestral range Zambia sample as the third population in each case. Note that selection in

218    either of the non-focal populations generates negative *PBE* values, rather than positive

219    outliers. We then tested for a significant overlap in (positive) *PBE* outliers between

220    different cold-adapted populations, using a permutation approach. No significant overlap

221    in window *PBE* outliers was detected when France was compared against Ethiopia (*P* =

222    0.36) or South Africa (*P* = 0.94). But when the two cold-adapted sub-Saharan populations

223    were compared, a significant overlap in *PBE* outliers was detected (*P* = 0.013). The latter

224    result reflected a moderate enrichment of 77 overlapping outlier regions compared to a

225    random expectation of 61. Thus, window *PBE* provided some evidence of parallel

226    adaptation, but only between the two sub-Saharan population pairs.

227

228

10

229    *Genomic variant analysis of parallel genetic differentiation*

230        Particularly if natural selection has acted on standing genetic variation, it is possible

231    that individual SNPs might show parallel frequency changes without strong effects on

232    linked variants.  We therefore evaluated site-specific *PBE* and tested whether SNPs with

233    strong cold vs. warm frequency differences in two different population pairs tended to be

234    "codirectional" (the same allele increased in both cold-adapted populations, relative to

235    each one's warm-adapted counterpart) rather than "antidirectional".  In contrast to the

236    window analysis described above, all three comparisons of population pairs showed a

237    genomic signal of parallel SNP frequency change (Figure 3), in that SNPs with high *PBE*

238    quantiles in two different population pairs were more likely to involve codirectional

239    frequency change.  Surprisingly, the strongest enrichment of codirectional *PBE* SNP outliers

240    did not come from Ethiopia and South Africa (the comparison that had significantly

241    overlapping window *PBE* outliers).  Instead, Ethiopia and France/Egypt showed the

242    greatest excess of codirectional *vs.* antidirectional SNPs that had *PBE* quantile below 0.05

243    in both pairs (Figure 3).  Stronger enrichments were observed for more stringent *PBE*

244    thresholds:  for example, the numbers of codirectional versus antidirectional SNPs with

245    *max*(*PBEQ*) < 0.005 was 45 vs. 8 for SD/EF, 7 vs. 1 for SD/FR, and 21 vs. 2 for EF/FR (Table

246    S4).  As expected based on linkage, some codirectional outliers were clustered, but overall

247    they were widely distributed across the analyzed genomic regions (Figure S1, Table S5).

248        If the above patterns were produced by elevated gene flow between warm-adapted

249    populations or between cold-adapted populations, then a general excess of codirectional

250    over antidirectional SNPs would be expected.  In contrast to this prediction, SNPs that were

251    not outliers in both population pairs were equally likely to be codirectional or

11

252    antidirectional in all comparisons (Figure 3).  While the stochastic variance induced by

253    population bottlenecks might modestly increase the number of codirectional outliers

254    (Tennessen and Akey 2011), genetic differentiation between the cold- and warm-adapted

255    populations of each of our pairs is low ($F_{ST}$ less than 0.001 for SD/SP, 0.035 for EF/EA, and

256    0.071 for France/Egypt).  Overall, the consistently strong enrichment of codirectional

257    outliers we observed seems likely to be a product of parallel selective pressures.

258         To generate hypotheses regarding the genetic basis of parallel adaptation, we

259    performed gene ontology (GO) enrichment analysis for each codirectional comparison.

260    Membrane-associated gene products were enriched in all three comparisons (specifically

261    "membrane part" and "integral to membrane"), and "ion channel activity" was enriched in

262    the EF/FR and FR/SD comparisons.  Other categories enriched in two comparisons

263    included "detection of stimulus", "growth factor activity", and "innate immune response"

264    (Table S6).  Intuitively, the comparison with the strongest codirectional SNP enrichment

265    (EF/FR) had the largest number of GO categories with raw permutation $P$ values below

266    0.05.  Enriched terms in this comparison included "neuropeptide receptor activity" (along

267    with other receptor, signaling, and transmembrane terms), "odorant binding", and

268    "reproduction", among others (Table S6).

269         We also performed a search for SNPs with parallel allele frequency change in all three

270    cold-adapted populations.  As with the two-pair analyses, an excess of codirectional versus

271    antidirectional $PBE$ outliers was observed (Figure 3).  Just 28 SNPs had $PBE$ quantiles

272    below 0.05 for all population pairs (Table S5), but one of these reached genome-wide

273    significance.  This variant showed striking frequency shifts in all three population pairs

274    (Figure 4), with SNP $PBE$ quantiles at or below 0.0036 in each case.  The random

12

275    probability of a three-pair codirectional SNP of this type is $(1/2)^2 \times (0.0036)^3 = 1.2 \times 10^{-8}$,

276    such that even with 431,427 tested SNPs, its estimated true positive probability is 99.5%

277    (Table S4).  This SNP on chromosome arm 2R is located at position 14,120,310 in release 6

278    of the *D. melanogaster* genome and position 10,007,815 in release 5.  It is located in the first

279    intron of *Prosap*, which encodes a predicted component of the postsynaptic density (Leibl

280    and Featherstone 2008).  Curiously, the "cold" allele at this transition polymorphism

281    appears to be the ancestral state, with the same "G" being carried by *D. sechellia*, *D.*

282    *simulans*, *D. erecta*, and *D. yakuba*.

283        There is little evidence of strong hard sweeps around *Prosap*.  The cold- and warm-

284    adapted populations studied all have fairly similar nucleotide diversity in windows across

285    this gene region (Figure S2), including the 4.5 kb window encompassing the codirectional

286    SNP noted above.  There are no obvious population differences in haplotype identity

287    (Figure S2), as might be expected after a partial sweep or a moderately soft sweep (Lange

288    and Pool 2016).  In terms of window *PBE* values, only the Ethiopian comparison shows

289    some elevation at the focal window (Figure S2).  But at this same window, all three

290    comparisons have maximum SNP *PBE* quantiles in the top 5% (Figure S2).  This pattern of

291    strong SNP differentiation without a notable linkage signal could reflect soft sweeps

292    (Pennings and Hermisson 2006), which is consistent with the non-trivial frequency of the

293    "cold" allele in the warm-adapted populations.

294        For France, the parallel divergent SNP highlighted above has the most extreme SNP *PBE*

295    across the entire ~100 kb *Prosap* region (Figure 5).  But within a 20 kb region centered on

296    this site, it is only the third and sixth most extreme SNP for South Africa and Ethiopia,

297    respectively.  Our alignments also show no evidence for a strongly codifferentiated indel in

13

298     this region.  Hence, although the allele frequency pattern at our focal SNP is dramatic,

299     further study will be needed to confirm which *Prosap* variants may have been under

300     selection in which cold-adapted populations.

301

302

303     DISCUSSION

304

305         Instances of parallel phenotypic change can shed light on the genetic predictability of

306     evolution.  We have shown that as *D. melanogaster* expanded from warm ancestral

307     environments, it evolved greater cold tolerance not only in the palearctic region, but also in

308     the highlands of Ethiopia and South Africa.  While natural selection is likely to have favored

309     cold tolerance independently in each case (Introduction), we investigated whether some of

310     the same genetic changes may have facilitated adaptation in two or more colder regions.

311         We found modest evidence of parallel genetic differentiation in short genomic windows,

312     with an excess of shared outliers specifically between the two African cold-adapted

313     populations.  These Ethiopian and South African sample locations are separated by 4,025

314     km of generally much warmer terrain, so gene flow of cold-adapted alleles between them

315     would require persisting in warm-adapted gene pools.  Both of these populations come

316     from high altitude environments (EF at 3,070 meters above sea level, SD at 2,000), so

317     parallel adaptation between them might relate not only to cold, but also to gradients such

318     as air pressure, desiccation, and ultraviolet radiation.

319         Intriguingly, our SNP-based analysis provided a much more consistent signal of parallel

320     evolution.  Overall, SNPs that changed frequency in two different cold-adapted populations

14

321    (relative to their warm-adapted paired samples) were equally likely to be codirectional

322    (with the same allele rising in each cold-adapted population) or antidirectional.  However,

323    SNPs with stronger allele frequency changes were more likely to be codirectional,

324    regardless of the cold-adapted populations compared.  It is difficult to exclude all other

325    adaptive and demographic hypotheses, but the clearest explanation for this result would be

326    if cold temperatures or correlated selective pressures favored the same variants in more

327    than one cold-adapted population.

328        The suggestion that some of the same genetic variants may have been favored in two or

329    more cold-adapted populations raises the question of why our window analysis did not

330    detect a greater number of overlapping *PBE* outliers between cold-adapted populations.

331    The window analysis may reflect a genuine contribution of distinct loci to adaptive

332    evolution in each population.  In theory, classic sweep signals could be too narrow (if

333    selection is very weak) or too old for our window analysis to detect.  However, divergence

334    among the studied populations may not be ancient enough to facilitate these explanations

335    (Pool *et al.* 2012).  Further study of the demographic history of these populations is

336    desirable, although the effects of positive and negative selection on such estimates is not

337    fully clear.  Otherwise, the contrasting window and SNP results could reflect an important

338    role for selection on standing genetic variation leading to soft sweeps.  If selection favors an

339    existing variant with non-trivial frequency, it may have recombined onto more than one

340    haplotype.  The fixation of an allele present on multiple haplotypes has far less influence on

341    linked variation (Pennings and Hermisson 2006).  If selection acted on variants with high

342    enough prior frequency, there might be little signal for window $F_{ST}$ (Lange and Pool 2016)

343    or related statistics like *PBE*.  Further study will be needed to identify the evolutionary

344     models and parameters that provide the best match to genetic differentiation at outlier loci

345     across the genome.

346         There are notable commonalities between the genes and functions identified in our

347     codirectional SNP analysis and in a recent study (Pool 2015).  That study examined African

348     and European ancestry along genomes from an admixed population of *D. melanogaster* (the

349     Drosophila Genetic Reference Panel) and found an abundance of "ancestry disequilibrium"

350     in which unlinked loci had correlated ancestry.  For example, genomes carrying an African

351     allele at a particular X-linked window might be more likely to carry an African allele in a

352     window on chromosome 2.  Of the 17 interaction "hubs" cited there (Figure 5 of Pool

353     2015), 10 of the listed genes were also detected in at least one of our codirectional

354     comparisons (*AstA-R1*, *CcapR*, *CCHa1-R*, *Fife*, *GABA-B-R1*, *norpA*, *Or65b*, *Piezo*, *Rdl*, *shakB*,

355     and *Spn*).  These genes include several with roles in neurotransmission, which is also

356     influenced by the postsynaptic density to which *Prosap* contributes (Liebl and Featherstone

357     2008).  The above genes were previously identified not only by ancestry disequilibrium,

358     but also by elevated $F_{ST}$ between European and western African genomes (Pool 2015), so

359     their detection in this study is not entirely independent.  However, our results raise the

360     possibility that similar genes and processes may have been under selection in other cold-

361     adapted populations as well.

362         Prior to this study, GO categories related to the nervous system have been enriched in a

363     variety of genome scans for positive selection in *D. melanogaster* (Langley *et al.* 2012; Pool

364     *et al.* 2012; Pool 2015).  And yet, very few empirical studies have investigated nervous

365     system evolution in *Drosophila*.  A rare example is the study of Campbell and Ganetzky

366     (2012), who found that the structure of a neuromuscular junction (a synapse between a

16

367    neuron and a muscle fiber) has evolved dramatically between *Drosophila* species, even

368    when closely related taxa were compared.  We hypothesize that parallel nervous system

369    evolution may have occurred within *D. melanogaster* during adaptation to cold

370    temperatures or a correlated selective pressure.  Such evolution may have served to

371    maintain ancestral function in a challenging environment, or it may have adaptively altered

372    behavior or other phenotypes.  Expanded study of nervous system evolution in *D.*

373    *melanogaster* is needed to clarify the molecular mechanisms and phenotypic relevance of

374    parallel adaptive changes at these genes.

375

376

377    MATERIALS AND METHODS

378

379    *Fly strains and phenotypic assays*

380       All fly population samples used here were described in previous population genomic

381    studies (Pool *et al.* 2012; Lack *et al.* 2015).  The strains used in this study were derived

382    from independent isofemale lines and inbred for eight generations prior to study.  Cold

383    tolerance experiments focused on females that were 4 to 7 days old at the outset of the

384    experiment.  Prior to cold exposure, flies were raised at 25°C on medium prepared in

385    batches of:  4.5 L water, 500 mL cornmeal, 500 mL molasses, 200 mL yeast, 54 g agar, 20

386    mL propionic acid, and 45 mL tegosept 10% (in 95% ethanol).  To control larval density,

387    flies were raised in half pint bottles with 20 virgin females and 20 males allowed to

388    oviposit for 48 hours.

389     The cold tolerance experiment involved placing flies in a 4°C chamber. For each strain,

390     a bottle containing 30-50 females was used. After 96 hours, the number of standing and

391     knocked-down flies in each glass vial was counted. Flies remaining on their legs were

392     deemed to retain mobility, while other flies were considered immobile (dead or in chill

393     coma). In our hands, we found that this 4°C assay provided more repeatable results than

394     chill coma response time or survival after brief exposure to below-freezing temperatures.

395     Most testing was done on inbred lines. For a subset of lines from each population

396     sample, we investigated cold tolerance in outbred flies as well. For each population

397     sample, a single inbred line was used as the source of virgin females for between-line

398     crosses. This "reference line" was then crossed to males from a series of other inbred lines

399     from the same population sample. The cold tolerance observed for outbred F1 flies from

400     each of these crosses was compared to expectations based on the midparent value from the

401     two parental inbred strains.

402

403     *Genomes and inversion testing*

404     An initial set of genomes from each of these six population samples was previously

405     published (Pool *et al.* 2012; Lack *et al.* 2015). Additional genomes used here (from a subset

406     of previously studied populations) were sequenced, aligned, and filtered (for

407     heterozygosity, identity by descent, and population admixture) using identical methods to

408     those described previously (Lack *et al.* 2015). These genomes are documented in an

409     accompanying article that expands the *Drosophila* Genome Nexus (Lack *et al.* 2016a).

410     Recent gene flow into Africa, which may reflect urban adaptation, was masked from this

411     data set (Pool *et al.* 2012; Lack *et al.* 2015; Lack *et al.* 2016a).

412     Our study focuses on three pairs of population samples, in which the closely-related

413     members of each pair derive from warm and cold environments, respectively.  The

414     population genetic analyses described below also draw upon a published sample of

415     genomes from Zambia (ZI), which is thought to represent an ancestral range population

416     (Pool *et al.* 2012; Lack *et al.* 2015).  Numbers of genomes analyzed for each population are

417     given in Table 1.

418     We used PCR to test inbred strains from each population sample for eight common

419     inversions (*In1A*, *In1Be*, *In2Lt*, *In2RNS*, *In3LP*, *In3RK*, *In3RMo*, and *In3RP*), using primers

420     and conditions given by Lack *et al.* (2016b).  These primers were modified from those

421     published by Corbett-Detig *et al.* (2012) to accommodate or avoid common SNPs.

422

423     *Window-based population genetic analysis*

424     Our primary interest in genomic analysis is to identify loci with strong genetic

425     differentiation between the members of a warm- and cold-adapted population pair, which

426     may represent targets of population-specific positive selection.  The primary population

427     genetic approach uses statistics based on $F_{ST}$, which quantifies allele frequency differences

428     between two populations.  By adding pairwise $F_{ST}$ values involving a third "outgroup"

429     population, *PBS* isolates the allele frequency change that is specific to one focal

430     population's history (Yi *et al.* 2010).  *PBS* will be large when natural selection has been

431     specific to the focal population, but it will also be large when all populations have long

432     branches.  In this study, our interest is to find loci that experienced positive selection in the

433     cold-adapted populations specifically, and we would like to avoid universally-selected loci

434     when testing whether different cold-adapted populations share many population genetic

19

435    outliers in common.  We therefore used a recently-described variant of *PBS* known as

436    Population Branch Excess (*PBE*; Yassin *et al.* 2016).  *PBE* quantifies the difference between

437    the observed *PBS* value and that expected based on (1) genetic differentiation between the

438    two non-focal populations at this locus, and (2) typical values for that quantity and *PBS* at

439    other loci.  Numerically,

440    $$PBE = PBS_{obs} - PBS_{exp} = PBS - \left(T^{BC} \times \frac{PBS_{med}}{T^{BC}_{med}}\right).$$

441    Here, $T^{BC}$ quantifies genetic differentiation between the non-focal populations, as $-\log(1 -$

442    $F_{ST})$, which scales our locus-specific expectations for genetic differentiation.  The ratio of

443    median values reflects the typical relationship between *PBS* and $T^{BC}$ observed genome-wide

444    (or in our case, for a given chromosome arm).  We expect *PBE* to be strongly positive when

445    selection has been specific to the focal population, but not when selection has increased all

446    branch lengths similarly.

447        We calculated *PBE* for genomic windows with sizes scaled by local diversity levels.

448    Each window contained 250 non-singleton single nucleotide polymorphisms (SNPs) in the

449    ZI population.  These windows had a median size of roughly four kilobases.  To focus on

450    regions in which selection signals should be more localized, we focused our population

451    genetic analyses on middle portions of chromosome arms where sex-averaged crossing-

452    over rates are generally above 0.5 cM / Mb (Comeron *et al.* 2012), specifically:  2.3 to 21.4

453    Mb for the X chromosome, 0.5 to 17.5 Mb for arm 2L, 5.2 to 20.8 Mb for arm 2R, 0.6 to 17.7

454    Mb for arm 3L, and 6.9 to 26.6 Mb for arm 3R.  We also evaluated *PBE* for specific SNPs, and

455    identified the maximum SNP *PBE* value for each window.  For both window and maximum

456    SNP *PBE*, we computed quantiles corresponding to the proportion of analyzed windows on

457    the same chromosome arm with an equal or greater value of *PBE*.  All window and SNP-

20

458    specific analyses included only genomes with a standard arrangement inferred for the

459    chromosomal arm of interest (Corbett-Detig *et al.* 2012).

460        We used a shift permutation approach to test whether window *PBE* outliers in different

461    populations overlapped more than expected.  Outlier windows were grouped into "regions"

462    encompassing all neighboring windows with *PBE* quantile below 0.05, interrupted by no

463    more than four non-outlier windows.  Comparing two sets of outlier regions, one set was

464    shifted within chromosome arms.  Outlier region locations were shifted in increments of

465    five windows to increase the independence of shift replicates.  Outliers were shifted a

466    minimum of 50 windows from their empirical locations.  Based on the windows present on

467    the shortest chromosome arm (2R), this allowed for a total of 827 shift permutations.

468

469    *SNP-based population genetic analysis*

470        Analogous to the study of Tennessen and Akey (2011), we tested whether SNP

471    frequency differences tend to have the same allele at elevated frequency in two or more

472    cold-adapted populations.  Specifically, we assigned a SNP quantile to the *PBE* value for

473    each of a population pair's analyzed SNPs.  When then tested whether "codirectional" SNPs

474    (those with the same allele at elevated frequency in both cold-adapted populations relative

475    to their warm-adapted counterparts) were more common than "antidirectional" SNPs in

476    which opposite alleles were at higher frequency in the two cold-adapted populations.  Non-

477    directional SNPs (those with no frequency change in one or both pairs) were excluded from

478    this analysis, so the null expectation is that 50% of SNPs should be codirectional by chance.

479    If parallel selective pressures have raised the frequency of certain variants in more than

480    one cold-adapted population, we might observe that SNPs with high *PBE* in both population

21

481   pairs are more likely to be codirectional than antidirectional.  Here, we focused on the

482   comparison of SNPs with *PBE* quantile below 0.05 in both population pairs versus those

483   with *PBE* quantile above 0.1 in one or both pairs.  We also performed a codirectional

484   analysis with all three population pairs, using the same data filters outlined above.  Here,

485   the proportion of codirectional SNPs is expected to be one quarter by chance, rather than

486   one half.

487       Within each pair of cold- and warm-adapted populations, we excluded sites with less

488   than ten called alleles in the cold-adapted population, or less than five called alleles in each

489   warm-adapted population.  We restricted the analysis to SNPs with an average population

490   frequency of the minor allele of at least 10% between the two populations.  We then

491   created a merged set of SNPs that met our sampling and frequency criteria for each of two

492   population pairs.  Although SNPs with intermediate *versus* rare frequency in the ancestral

493   population might have different distributions of *PBE* under neutrality, we would still expect

494   both classes to be codirectional and antidirectional with equal probability.

495       We performed gene ontology enrichment analysis on SNPs identified as codirectional

496   outliers, separately for each of the three analyses comparing two population pairs (EF/FR,

497   EF/SD, and FR/SD with respect to the cold-adapted populations).  Outlier SNP lists were

498   pared to remove sites within 10 kb of each other, excluding the site with the least extreme

499   *PBE* quantile in either population pair – in other words, retaining the site with the lower

500   max($PBEQ_1$, $PBEQ_2$).  For exonic SNPs (coding or UTR), the genes with overlapping exons

501   were noted.  For non-exonic SNPs, the genes with the nearest proximal and distal exons

502   were noted.  Each outlier SNP could only implicate a specific GO category once, and each

503   gene could only be counted once.  Permutations (50,000) were performed in which each

22

504    outlier SNP on the pared list was randomly relocated within the analyzed region of the

505    same chromosome arm (thus accounting for variability in gene length).  For each GO

506    category, a *P* value indicated the proportion of permutation replicates in which an equal or

507    greater number of genes was implicated.

508

509    ACKNOWLEDGMENTS

512 REFERENCES

513 Ayrinhac A, Debat V, Gibert P, Kister A-G, Legout H, Moreteau B, Vergilino R, David JR. 2004.

514     Cold adaptation in geographical populations of *Drosophila melanogaster*: phenotypic

515     plasticity is more important than genetic variability. Func Ecol. 18:700–706.

516 Bergland AO, Tobler R, Gonzalez J, Schmidt P, Petrov D. 2016. Secondary contact and local

517     adaptation contribute to genome-wide patterns of clinal variation in *Drosophila*

518     *melanogaster*. Mol Ecol. 25:1157–1174.

519 Božičević V, Hutter S, Stephan W. 2016. Population genetic evidence for cold adaptation in

520     European *Drosophila melanogaster* populations. Mol Ecol. 25:1175–1191.

521 Campbell M, Ganetzky B. 2012. Extensive morphological divergence and rapid evolution of

522     the larval neuromuscular junction in *Drosophila*. Proc Natl Acad Sci USA. 109:E648-E655.

523 Cavalli-Sforza L. 1969. Human diversity. Proc 12th Int Congr Genet. 2:405-416.

524 Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in

525     *Drosophila melanogaster*. PLoS Genet 8:e1002905.

526 Corbett-Detig RB, Cardeno C, Langley CH. 2012. Sequence-based detection and breakpoint

527     assembly of polymorphic inversions. Genetics 192:131-137.

528 Dahlgaard J, Hoffmann AA. 2000. Stress resistance and environmental dependency of

529     inbreeding depression in *Drosophila melanogaster.* Cons Biol. 14:1187-1192.

530 DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis AA, Del Angel G,

531     Rivas MA, Hanna M, *et al.*, 2011. A framework for variant discovery and genotyping using

532     next-generation DNA sequencing data. Nature Genet. 43: 491–498.

533 Dieringer D, Nolte V, Schlötterer C. 2005. Population structure in African *Drosophila*

534     *melanogaster* revealed by microsatellite analysis. Mol Ecol. 14:563–573.

535   Duchen P, Živković D, Hutter S, Stephan W, Laurent S. 2013. Demographic inference reveals

536      African and European admixture in the North American *Drosophila melanogaster*

537      population. Genetics. 193:291-301.

538   Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, Sukernik

539      R, Utermann G, Pritchard J, Coop G, *et al*. 2010. Human adaptations to diet, subsistence,

540      and ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci USA

541      107:8924–8930.

542   Hoffmann AA, Sørensen JG, Loeschcke V. 2003. Adaptation of *Drosophila* to temperature

543      extremes: bringing together quantitative and molecular approaches. J Therm Biol.

544      28:175–216.

545   Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D. 2015. Population genomic analysis

546      uncovers African and European admixture in *Drosophila melanogaster* populations from

547      the south-eastern United States and Caribbean Islands. Mol Ecol. 24:1499–1509.

548   Kapun M, Fabian DK, Goudet J, Flatt T. 2016. Genomic evidence for adaptive inversion

549      clines in *Drosophila melanogaster*. Mol Biol Evol. 33:1317-1336.

550   Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis

551      version 7.0 for bigger datasets. Mol Biol Evol. In Press.

552   Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M. 1988. Historical

553      Biogeography of the Drosophila melanogaster Species Subgroup. In: Hecht MK, Wallace B,

554      Prance GT, editors. *Evolutionary Biology*. Boston, MA: Springer. p. 159–225.

555   Lack JL, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH,

556      Pool JE. 2015. The Drosophila genome nexus: a population genomic resource of 623

25

557     *Drosophila melanogaster* genomes, including 197 from a single ancestral range

558     population. Genetics. 199:1229-1241.

559     Lack JL, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. 2016. A thousand fly genomes:

560     version 1.1 of the *Drosophila* Genome Nexus. Mol Biol Evol., co-submitted with this article.

561     Lack JB, Monette MJ, Johanning EJ, Sprengelmeyer QD, Pool JE. 2016. Decanalization of wing

562     development accompanied the evolution of large wings in high-altitude *Drosophila*. Proc

563     Natl Acad Sci USA 113:1014-1019.

564     Lange JD, Pool JE. 2016. A haplotype method detects diverse signatures of local adaptation

565     from genomic sequence variation. Mol Ecol. In Press.

566     Lemeunier F, Aulard S. 1992. Inversion polymorphism in *Drosophila melanogaster*. In: CB

567     Krimbas, JR Powell, editors. *Drosophila* inversion polymorphism. Boca Raton (FL): CRC

568     Press. p. 339–405.

569     Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler

570     transform. Bioinformatics 26:589–595.

571     Liebl FLW, Featherstone DE. 2008. Identification and investigation of *Drosophila*

572     postsynaptic density homologs. Bioinform Biol Insights 2:369–381.

573     Lunter G, Goodson M. 2010. Stampy: a statistical algorithm for sensitive and fast mapping

574     of Illumina sequence reads. Genome Res. 18:821–829.

575     Ohta T. 1971. Associative overdominance caused by linked detrimental mutations. Genet

576     Res. 18:277-286.

577     Pennings PS, Hermisson J. 2006. Soft sweeps III: the signature of positive selection from

578     recurrent mutation. PLoS Genet. 2:e186.

579    Pool JE.  2015. The mosaic ancestry of the *Drosophila* Genetic Reference Panel and the *D.*

580        *melanogaster* reference genome reveals a network of epistatic fitness interactions.Mol

581        Biol Evol. 32:3236-3251.

582    Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P,

583        Emerson JJ, Saelao P, Begun DJ, *et al*. 2012. Population genomics of sub-Saharan

584        *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet.

585        8:e1003080.

586    Svetec N, Werzner A, Wilches R, Pavlidis P, Alvarez-Castro JM, Broman KW, Metzler D,

587        Stephan W. 2011. Identification of X-linked quantitative trait loci affecting cold tolerance

588        in *Drosophila melanogaster* and fine mapping by selective sweep analysis. Mol Ecol.

589        20:530–544.

590    Thornton KR, Andolfatto P. 2006. Approximate Bayesian inference reveals evidence for a

591        recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*.

592        Genetics. 172:1607–1619.

593    Turner TL, Levine MT, Eckert ML, Begun DJ. 2008. Genomic analysis of adaptive

594        differentiation in *Drosophila melanogaster*. Genetics. 179:455–473.

595    Yassin A, Debat V, Bastide H, Gadesweski N, David JR, Pool JE. 2016. Repeated ecological

596        specialization and incipient speciation in an island *Drosophila* population. Proc Natl Acad

597        Sci USA 113:4771-4776.

598    Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N,

599        Korneliussen TS, *et al.* 2010. Sequencing of 50 human exomes reveals adaptation to high

600        altitude. Science. 329:75–78.

01 **Table 1.** *D. melanogaster* inbred lines from three distinct regions show elevated cold tolerance (measured as the percentage of female flies

02 still standing or moving after 96 h at 4°C). To illustrate climate differences, estimates of the "midpoint temperature" (the average of 24

03 monthly average highs and lows) and the minimum monthly average low are given. *P* values show that within each pair of populations, the

04 sample originating from the higher altitude/latitude shows significantly greater cold tolerance.

05

| Population Sample | Genomes | Latitude | Altitude (meters) | Midpoint Temp. (°C) | Min. Month Low (°C) | Lines Tested | Mobility, 96 h at 4°C | U Test *P* Value |
|---|---|---|---|---|---|---|---|---|
| Siavonga, Zambia (ZI) | 197 | -16.5 | 530 | 25 | 12 | 0 | -- | |
| Phalaborwa, S.Africa (SP) | 37 | -23.9 | 350 | 22 | 8 | 19 | 26.8% | 0.0007 |
| Dullstroom, S.Africa (SD) | 81 | -25.4 | 2000 | 13 | 1 | 50 | 48.8% | |
| Gambella, Ethiopia (EA) | 24 | 8.2 | 525 | 28 | 20 | 9 | 13.1% | 0.0003 |
| Fiche, Ethiopia (EF) | 69 | 9.8 | 3070 | 11 | 2 | 46 | 36.3% | |
| Cairo, Egypt (EG) | 32 | 30.1 | 25 | 22 | 8 | 24 | 28.0% | <0.0001 |
| Lyon, France (FR) | 96 | 45.8 | 175 | 12 | -1 | 67 | 82.9% | |

606     **Table 2.**  The relationship between inbred and outbred cold tolerance differs strongly

607     population.  Here, outbred cold tolerance (measured as the percentage of female flies s

608     standing or moving after 96 h at 4°C) is compared with the midparent values expected

609     based on the average cold tolerance of the parental inbred strains.  $P$ values indicate st

610     advantages of outbred flies for EG, SD, and SP, but no hint of such an effect for the othe

611     populations.  These same three populations have the highest inversion frequencies (Fi

612     2).

613

| Population | Midparent Mobility | Outbred Mobility | Number of Crosses | Outbred Greater | Binomial $P$ Value |
|---|---|---|---|---|---|
| S. Africa SP | 11.8% | 50.1% | 6 | 6 | 0.016 |
| S. Africa SD | 51.5% | 73.0% | 6 | 6 | 0.016 |
| Ethiopia EA | 12.8% | 11.8% | 4 | 2 | 0.688 |
| Ethiopia EF | 53.4% | 50.6% | 12 | 4 | 0.927 |
| Egypt EG | 28.3% | 41.8% | 11 | 9 | 0.033 |
| France FR | 91.4% | 92.4% | 7 | 4 | 0.500 |

614

615
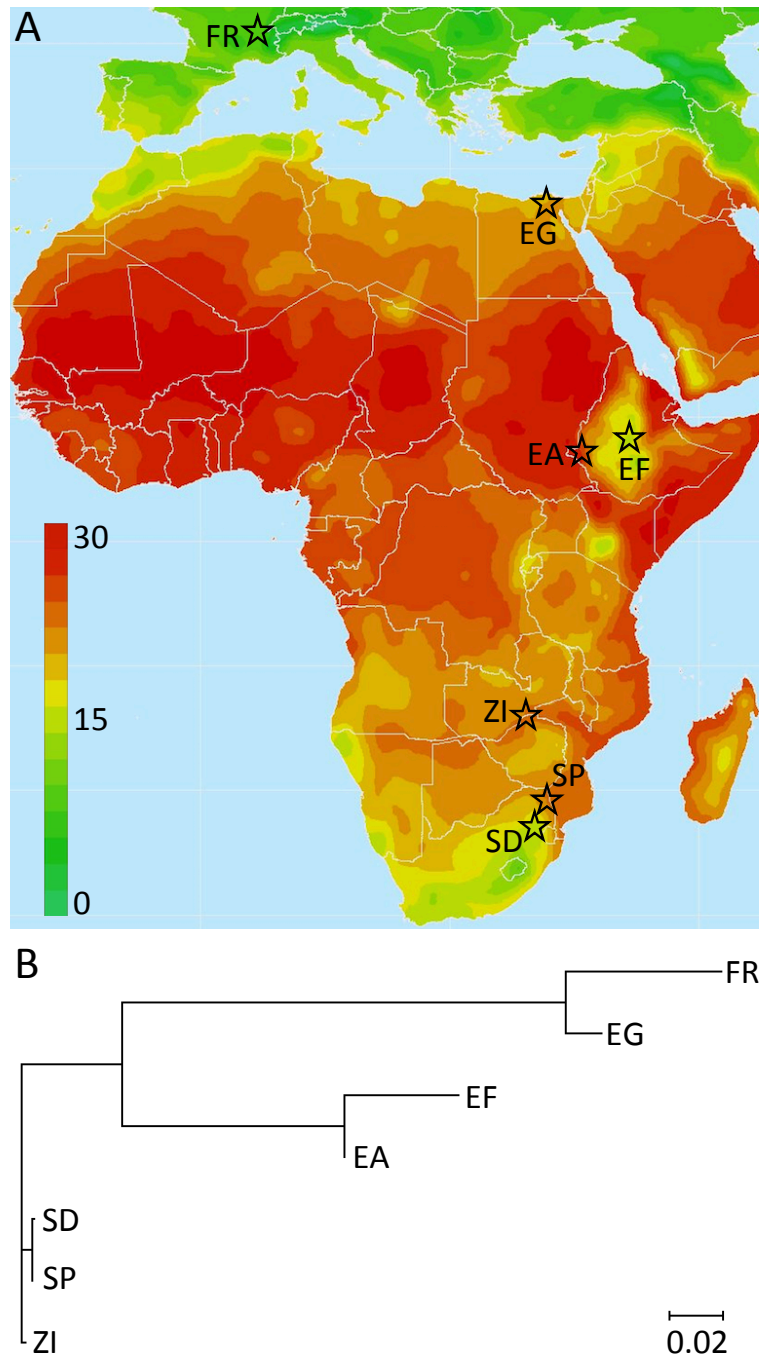
**Figure 1.** (A) Geographic locations and thermal environments of the studied populations are depicted. We compare three pairs of cold-/warm-adapted populations: France/Egypt and highland/lowland populations from Ethiopia and from South Africa. The origin of the ancestral range Zambia sample used in population genomic analyses is shown as well. The source map indicating average temperature (°C) is courtesy of The Nelson Institute, Center for Sustainability and the Global Environment, UW-Madison. (B) A neighbor-joining tree, based on transformed genome-wide average $F_{ST}$ values (Cavalli-Sforza 1969) and rooted along the branch to ZI, was generated using MEGA 7.0.16 (Kumar *et al.* 2016).

30
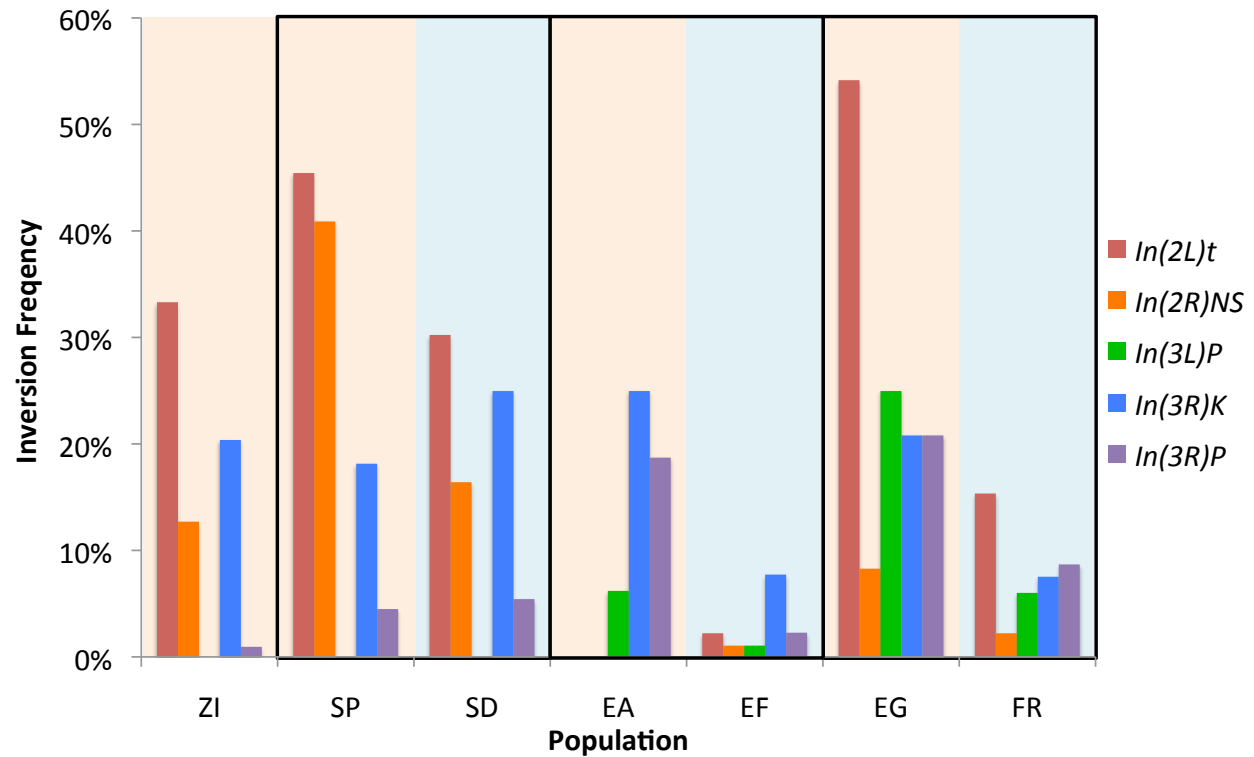
**Figure 2.** The frequencies of the five most common inversion polymorphisms tend to decrease in cold-adapted populations (blue background) relative to closely-related warm-adapted populations (tan background in same box). Further results of this inversion PCR testing are given in Table S3.

31

629
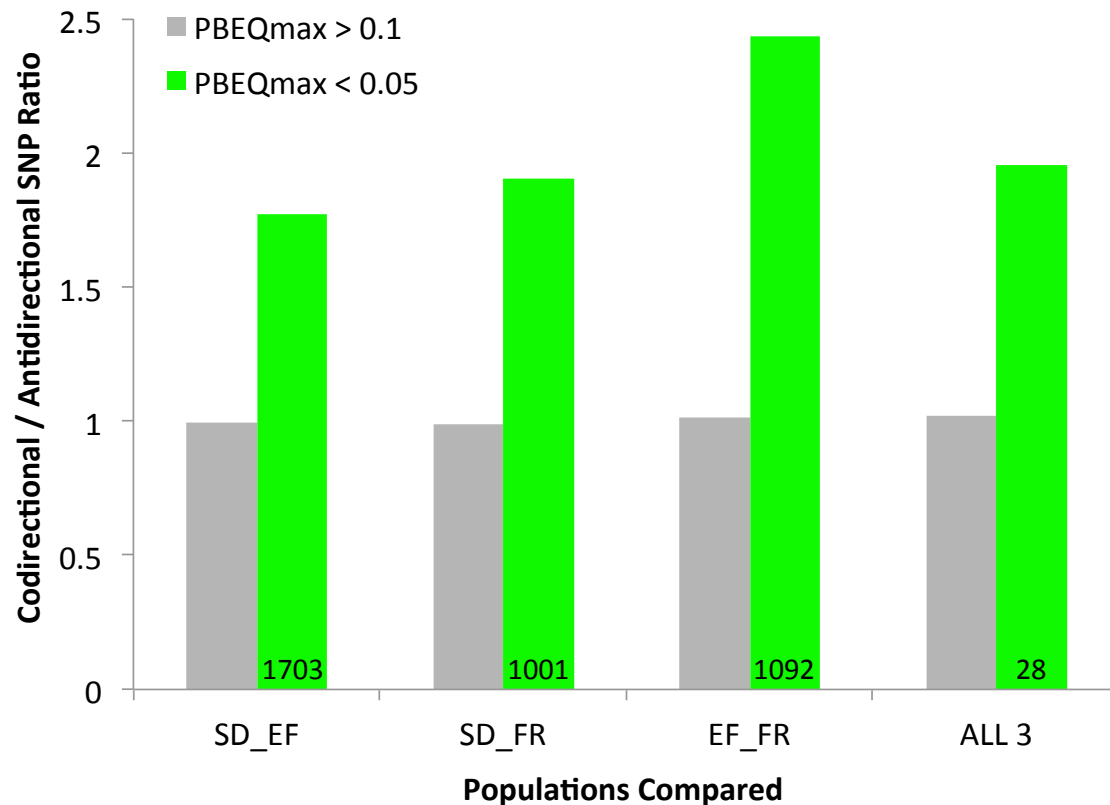
**Figure 3.** The genomic enrichment of SNPs with parallel frequency differences between cold-adapted and warm-adapted populations is illustrated. A SNP is classified as codirectional if the same allele is at higher frequency in cold-adapted populations compared to their warm-adapted counterparts. The ratio of codirectional to antidirectional SNPs is depicted for two classes of sites. Green bars indicate variants with high *PBE* values in each population pair considered (SNP *PBE* quantile below 0.05), and the total number of such codirectional SNPs is shown. Gray bars are for SNPs with *PBE* quantile above 0.1 for at least one of the population pairs. Joint *PBE* outliers consistently show an excess of codirectional SNPs, while non-outliers show very similar numbers of codirectional and antidirectional SNPs. Note that for the comparison including all three warm and cold populations pairs (right), ratios were multiplied by three to account for the expectation that only one quarter of SNPs should be codirectional by chance.
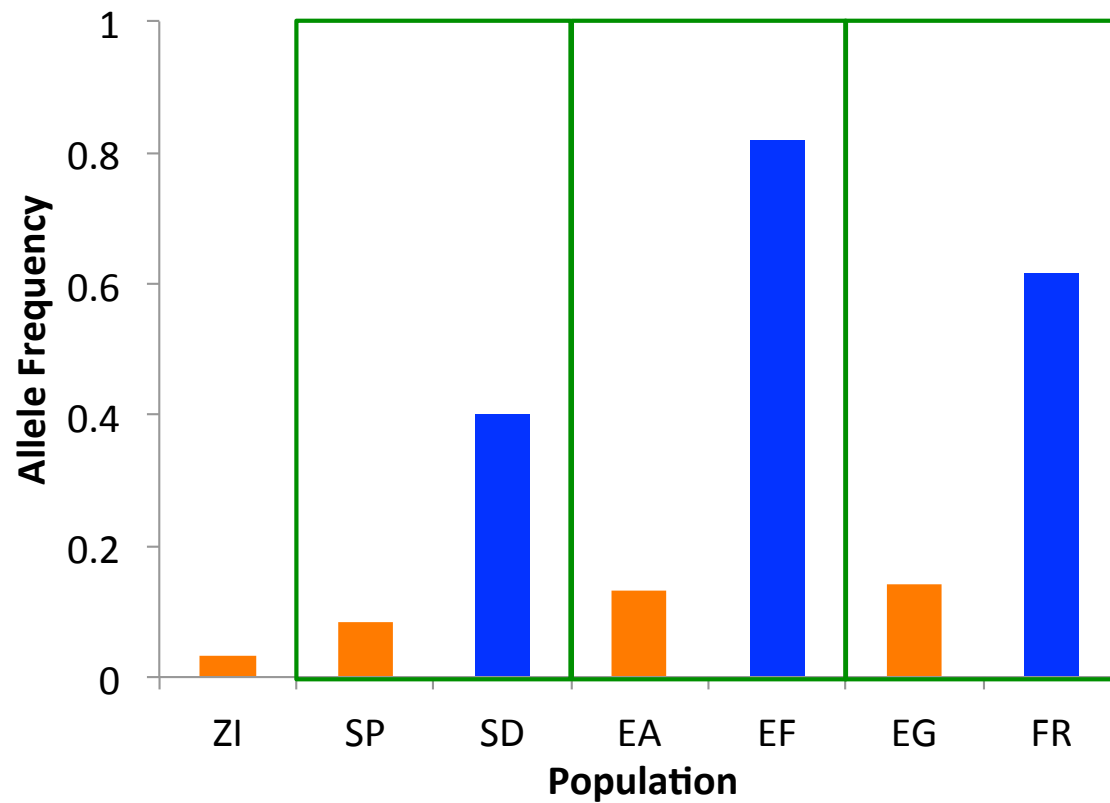
**Figure 4.** An intronic variant at *Prosap* displays codirectional allele frequency change in three cold-adapted populations (blue bars).
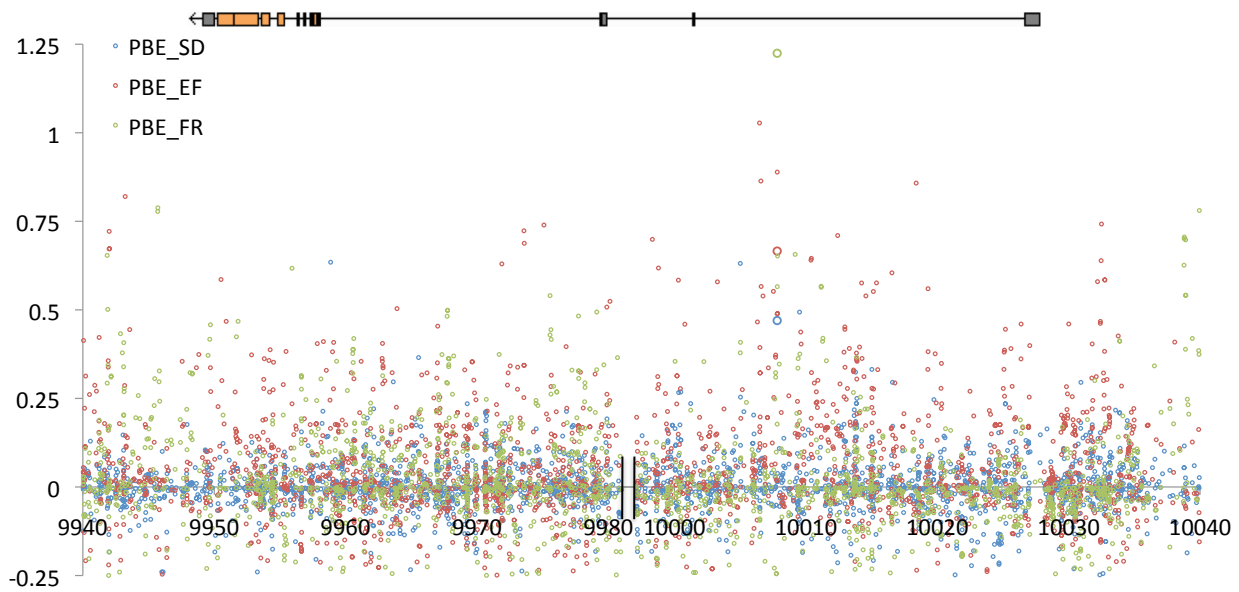
**Figure 5.** SNP *PBE* across the *Prosap* region is shown for three cold-adapted populations. While our focal SNP (larger circles near 10,008 kb) is the most differentiated for France (green), a handful of other SNPs are more extreme for Ethiopia (red) and South Africa (blue). The top diagram depicts the coding (orange) and non-coding (gray) exons of *Prosap*. The x-axis shows kilobase position along chromosome arm 2R according to release 5 of the *D. melanogaster* reference genome. A break in this axis, 11 kb away from the focal SNP, indicates a 16 kb region in which the reference genome contains two transposable elements (roo and HMS-Beagle) and our reference alignments contain no called sites.