1   **Article**
2   **Discoveries**
3
4   **Inter- and intra-specific genomic divergence in *Drosophila montana* shows evidence for**
5   **cold adaptation**
6

7   Darren J. Parker[1,2,3], R. Axel W. Wiberg[2], Urmi Trivedi[4], Venera I. Tyukmaeva[1], Karim Gharbi[4,5],
8   Roger K. Butlin[6], Anneli Hoikkala[1], Maaria Kankare[1*], and Michael G. Ritchie[2*]
9

10   1. Department of Biological and Environmental Science, University of Jyväskylä, P.O.
11   Box 35, FI-40014, Jyväskylä, Finland
12   2. Centre for Biological Diversity, School of Biology, University of St Andrews, Fife, KY16
13   9TH, UK.
14   3. Department of Ecology and Evolution, University of Lausanne, Biophore, Lausanne
15   1015, Switzerland
16   4. Edinburgh Genomics, School of Biological Sciences, University of Edinburgh,
17   Edinburgh, UK
18   5. Present address: Earlham Institute, Norwich Research Park, Norwich, UK
19   6. Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, U.K;
20   and Department of Marine Sciences, University of Gothenburg, 405 30 Göteborg,
21   Sweden

22   *Corresponding authors: Maaria Kankare: s.maaria.kankare@jyu.fi, Michael Ritchie: mgr@st-
23   andrews.ac.uk
24
25
26

27 **ABSTRACT**

28 The genomes of species that are ecological specialists will likely contain signatures of
29 genomic adaptation to their niche. However, distinguishing genes related to ecological
30 specialism from other sources of selection and more random changes is a challenge. Here
31 we describe the genome of *Drosophila montana*, which is the most extremely cold-adapted
32 Drosophila species. We use branch tests to identify genes showing accelerated divergence in
33 contrasts between cold- and warm adapted species and identify about 250 genes that show
34 differences, possibly driven by a lower synonymous substitution rate in cold-adapted
35 species. We look for evidence of accelerated divergence between *D. montana* and *D. virilis*,
36 a previously sequenced relative, and do not find strong evidence for divergent selection on
37 coding sequence variation. Divergent genes are involved in a variety of functions, including
38 cuticular and olfactory processes. We also re-sequenced three populations of *D. montana*
39 from its ecological and geographic range. Outlier loci were more likely to be found on the X
40 chromosome and there was a greater than expected overlap between population outliers
41 and those genes implicated in cold adaptation between Drosophila species, implying some
42 continuity of selective process at these different evolutionary scales.
43

44 **KEY WORDS**

45 Cold tolerance, *Drosophila montana*, ecological adaptation, comparative genomics

46

47 **BACKGROUND**

48

49 Comparative genomic analyses provide new insights into our understanding of evolutionary
50 processes by helping to identify genes contributing to adaptive divergence (Ellegren 2008;
51 Radwan and Babik 2012). If strong divergent selection due to environmental adaptation or
52 social interactions, such as sexual selection, act as "barrier loci" by influencing species
53 isolation, then identifying them can help to understand the process of speciation (Nosil et al.
54 2009; Smadja and Butlin 2011). However, accurately identifying such genes is a considerable
55 challenge (Noor and Bennett 2009; Cruickshank and Hahn 2014; Ravinet et al. 2017; Wolf
56 and Ellegren 2017).

57

58 Comparative genomic analyses are often hampered by a poor understanding of the sources
59 of selection contributing to species divergence (Ravinet et al. 2017; Wolf and Ellegren 2017).
60 Even when some of the sources of selection seem clear they are often complex and
61 multifaceted, greatly complicating our ability to identify the genetic basis of adaptations.
62 One approach to this problem is to apply comparative genomic techniques to species with
63 distinct ecological specialisations. Several studies have been made of such ecological
64 specialists, including: cactophilic *Drosophila* (Matzkin et al. 2006; Smith et al. 2013), Asian
65 longhorn beetles with specialized feeding habits (McKenna et al. 2016), climate-mediated
66 adaptations in honey bees (Chen et al. 2016), and adaptation to high altitude in humans (Foll
67 et al. 2014). These have successfully identified some associations, but such studies are still
68 relatively few, hindering our general understanding of the genomic landscape of adaptation.
69 Here we describe the genome of *Drosophila montana*, a widely-distributed northern
70 member of the virilis group of *Drosophila*, which shows unique adaptations to seasonally
71 varying environmental conditions prevailing at high latitudes and altitudes. *D. montana* is the
72 most cold-tolerant *Drosophila* species known (Kellermann et al. 2012; Andersen et al. 2015).
73 Their cold tolerance or hardiness involves multiple adaptations, including both a high general
74 resistance to cold and a strong inducible cold acclimation response (Vesala and Hoikkala
75 2011), as well as a robust photoperiodic diapause (Tyukmaeva et al. 2011), which all
76 contribute to its ability to survive through cold and dark winters. The daily and seasonal
77 activity patterns of *D. montana*, and the interactions and neurochemistry of the core
78 circadian clock genes behind these patterns, differs from those of more temperate species
79 such as *D. melanogaster* (Kauranen et al. 2012; Kauranen et al. 2016; Tapanainen et al.
80 2018). These features have likely played an important role in allowing *D. montana* to colonize
81 and persist in high-latitude environments (Terhzaz et al. 2015; Kauranen et al. 2016;
82 Menegazzi et al. 2017).

83 *D. montana* belongs to the *virilis* group of *Drosophila*, which comprises 13 species or
84 subspecies divided into two clades, the virilis and montana phylads, the latter being further
85 split into three lineages (Spicer and Bell 2002). These phylads are thought to have diverged in
86 South Asia during the Early Miocene, after which both of them entered the New World by
87 way of Beringia (Throckmorton 1982). The virilis phylad is constrained mostly within the
88 temperate zone, while the montana phylad has expanded into a variety of habitats and
89 spread to higher latitudes (Throckmorton 1982). Divergence of the two phylads has been
90 estimated to have occurred 7 (Ostrega 1985) to 11 (Spicer and Bell 2002) million years ago,

91   while the North American, European and Asian *D. montana* populations have diverged
92   within the last 450,000 to 900,000 years (Mirol et al. 2007). Interestingly, conspecific *D.*
93   *montana* populations have been shown to diverge in traits that play a role in ecological
94   adaptation (e.g. Lankinen et al. (2013) and Tyukmaeva et al. (2015)), male sexual cues and
95   female preferences (e.g. Klappert et al. (2007)), and also to show sexual and post-mating
96   pre-zygotic reproductive barriers (Jennings et al. 2014). Information on potential candidate
97   genomic regions and genes for traits involved in cold adaptation and sexual selection has
98   been accumulated through QTL analyses (Schäfer et al. 2010; Tyukmaeva et al. 2015),
99   microarray (Vesala et al. 2012; Salminen et al. 2015), transcriptome (Parker et al. 2015;
100 Kankare et al. 2016; Parker et al. 2016), and RNAi (Vigoder et al. 2016) studies.

101

102 Here we aim to identify genes showing evidence of divergent selection linked to cold
103 adaptation by contrasting the genomes of species and populations from different climatic
104 conditions. These analyses were conducted at three levels. Firstly, we classified Drosophila
105 species with well annotated genomes into cold-tolerant and non-cold-tolerant species and
106 used branch tests to identify genes evolving differently between these contrasts. Secondly,
107 we compared *D. montana* with its more temperate relative *D. virilis*. Finally, we compared
108 three divergent populations of *D. montana* from different geographic regions. Such a multi-
109 level approach allows us to identify genes that show recurrent divergence associated with
110 climatic differences between species and populations. Such genes are likely to be particularly
111 important for thermal adaptation, giving insight into the genes and functional processes
112 involved in the evolution of cold tolerance in insects more generally. Thus, our results thus
113 give a novel insight into genomic patterns of selection-driven divergence at different
114 evolutionary scales, in addition to providing a well-annotated genome for a uniquely cold
115 adapted insect species.
116
117
118
119

120 **RESULTS**

122 **Genome Sequencing and Assembly**

124 The assembled *D. montana* genome (**Table 1**) has a total length of 183.6 Mb, which falls
125 within the range seen for *Drosophila* species (111-187 Mb), and is similar to that of *D. virilis*
126 (172 Mb), a close relative of *D. montana* with a sequenced genome. CEGMA identified 238
127 complete orthologs (96%) and 244 partial orthologs (98%) of the 248 CEGMA proteins and
128 BUSCO identified 2457 genes as complete (92%) and failed to identify only 46 (1.7%).
129 RepeatMasker identified that 14.4% of the assembly was composed of repeat elements, the
130 major classes of which were: Simple repeats (4.5%), LTR elements (4.3%), Unclassified
131 (2.9%), and LINEs (1.9%) (**Supplementary Figure 1**). The total percentage of repeat elements
132 identified was around half of that found for related *Drosophila* species (*D. virilis* = 25.9%, *D.*
133 *mojavensis* = 23.8%, and *D. grimshawi* = 26.1%) likely reflecting the problem of assembling
134 repetitive regions with short reads.

136 For the genetic map construction, the final dataset contained 5,858 polymorphic SNPs. The
137 median depth of the SNPs in the final dataset was 52.4 and the average missing data rate
138 was 0.003. The initial analysis formed five major linkage groups (as expected since *D.*
139 *montana* has five chromosomes in total). Chromosome number was assigned by blasting
140 genes assigned to the linkage groups to the *D. virilis* genome, which have been localised to
141 chromosomes and is largely syntenic with *D. montana* (Schäfer et al. 2010). While the
142 analysis showed clear linkage groups, the order of markers was not totally resolved, likely
143 due to lack of recombination events among F2 progeny (**Supplementary Figure 2**). The
144 tentative scaffold order and position are given in **Supplementary Table 1**. Using this map we
145 were able to anchor approximately one third of the genome assembly to chromosomes. To
146 validate our linkage map we examined coverage of anchored scaffolds. X-linked regions were
147 found to have approximately half the coverage of autosomal regions, as expected since the
148 reference genome was produced from male-only samples (**Supplementary Figure 3**).

150 **Between-species comparisons identify genes showing accelerated divergence between**
151 **cold- and warm adapted species**

153 Across the 13 *Drosophila* species we found 250 genes that had significantly different rates of
154 evolution (ω) in cold- and non-cold-tolerant species (**Fig. 1, Supplemental Table 2**). dS was
155 on average lower for cold-tolerant species than for non-cold-tolerant species while dN was
156 very similar (**Fig. 2**, **Supplementary Table 3**). ω was on average greater for cold-tolerant
157 species, probably driven by generally lower values of dS in these species (**Supplementary**
158 **Table 3**). 203 and 47 genes showed higher values of ω for cold-tolerant and for non-cold-
159 tolerant species, respectively (**Fig. 3**). Genes with elevated ω in cold-tolerant species were
160 enriched for 23 GO terms (Biological Processes: Molecular Functions: Cellular Components =
161 6:10:7) (FDR < 0.1) (**Supplementary Table 4**), which semantically cluster into the following
162 categories: response to drug, male courtship behaviour, olfaction, ion-channel activity, and
163 developmental processes (**Fig. 4**). Of genes with elevated ω in non-cold-tolerant species we
164 identified 50 enriched GO terms (Biological Processes: Molecular Functions: Cellular

165 Components = 34:3:13) (FDR < 0.1) (**Supplementary Table 5**), which semantically cluster into
166 the following categories: proteasome-mediated ubiquitin-dependent protein catabolic
167 process, reproductive processes, response to fungus, animal organ morphogenesis,
168 regulation of biological and cellular processes (**Fig. 4**). Moreover, DAVID identified 11
169 functional group clusters for genes with significantly higher ω in cold-tolerant species
170 (**Supplementary Table 6**) including: Nucleotide-binding, Olfaction, Transmembrane proteins,
171 Neural development, Leucine-rich repeat containing proteins, GTPase / GTP binding,
172 Cytoskeleton / Microtubule, and Ion Transport. Finally, DAVID identified 3 functional group
173 clusters for genes with significantly higher ω in cold-tolerant species (**Supplementary Table
174 7**) including: Calcium-binding EGF domain containing proteins, Transmembrane proteins, and
175 Cytoskeleton / Microtubule.
176
177 **Comparison of *D. montana* and *D. virilis***
178
179 We estimated ω (dN/dS) for each of the one-to-one orthologs between *D. montana* and *D.
180 virilis* (**Supplemental Table 2**). No genes had a ω significantly greater than 1 after filtering
181 and multiple-test correction. Comparison of mean ω for several candidate gene sets (genes
182 involved in immune function, reproduction, and cold tolerance) found that none of the
183 candidate genes sets differed significantly from the genomic background (**Fig. 5**). By ranking
184 genes by ω we identified GO terms enriched in genes with relatively high and low ω. For
185 those with high ω we identified 23 enriched GO terms (Biological Processes: Molecular
186 Functions: Cellular Components = 10:4:9) (FDR < 0.1) (**Supplementary Table 8**). Semantic
187 clustering of these GO terms shows that they fall into the following categories: Reproduction,
188 detection of chemical binding / olfaction, amino sugar metabolism, and chitin binding (**Fig.
189 6**). DAVID identified 9 functional group clusters (**Supplementary Table 9**) including 2 related
190 to chitin production and 2 related to olfactory functions, congruent with the findings from
191 the single GO term enrichment analysis (above). In addition, DAVID also identified 2 clusters
192 involved in: immune defence (C-type lectin domain carrying genes, and Fibrinogen related
193 genes), Transcription factor binding, and a cluster containing genes with either a CAP
194 (cysteine-rich secretory protein) or SCP (Sperm-coating protein) domain. We identified 662
195 enriched GO terms for genes with low ω between *D. montana* and *D. virilis* (Biological
196 Processes: Molecular Functions: Cellular Components = 485:80:97) (FDR < 0.1). As expected
197 for genes with very low ω the enriched GO terms are consistent with housekeeping roles in
198 the cell (cell cycle control, cell communication, cell developmental process etc.), which are
199 expected to be under strong purifying selection (**Supplementary Table 10, Supplementary
200 Figures 4-6**).
201
202 **Genes showing significant between-population divergence are enriched for functional
203 processes associated with cold adaptation**
204
205 Significant outlier SNPs were found in, or within 1kb of, 1801, 155, and 1387 genes (from
206 pairwise comparisons between Colorado: Oulanka, Colorado: Vancouver, and Oulanka:
207 Vancouver respectively) (see supplementary material for detail on SNP numbers). 10 genes
208 overlapped between all the three pairwise comparisons (**Supplementary Figure 7,
209 Supplementary Table 11**). Although this is a relatively small number of genes, it is

210 significantly greater than expected by chance (p = 0.00013). By ranking genes by q-value we
211 could identify GO terms enriched in genes with high divergence for each population
212 comparison (Colorado: Oulanka = 74 (Biological Processes: Molecular Functions: Cellular
213 Components = 27:29:18) (**Supplementary Table 12**), Colorado: Vancouver = 66 (Biological
214 Processes: Molecular Functions: Cellular Components = 19:28:19) (**Supplementary Table 13**),
215 Oulanka: Vancouver = 91 (Biological Processes: Molecular Functions: Cellular Components =
216 37:39:14) (**Supplementary Table 14**). As with genes, there was a significant overlap of
217 enriched GO-terms between population comparisons (N = 22, p = 1.74 x 10^{-79},
218 **Supplementary Figure 7, Supplementary Table 15**). Semantic clustering of GO terms (**Fig. 7**)
219 and functional clustering (**Supplementary Table 16-18**) showed that the dominant terms
220 include: membrane components, ion transport, small molecule binding, and neuron /
221 synaptic associated terms.
222
223 Interestingly, outlier SNPs were not randomly distributed throughout the genome (**Fig. 8 and**
224 **Supplementary Figure 8**). There was a significant excess of outlier SNPs on the X-
225 chromosome in all pairwise comparisons (Colorado: Oulanka – Chi-squared = 3,029.4, d.f. =
226 4, p < 0.01; Colorado: Vancouver - Chi-squared = 31.9, d.f. = 4, p < 0.01; Oulanka: Vancouver -
227 Chi-squared = 2477.7, d.f. = 4, p < 0.01). These results held when the proportion of the total
228 genome length of each chromosome was taken to calculate the expected numbers of SNPs.
229
230 **Genes showing divergence between species and between populations overlap**
231
232 We examined whether genes showing significant divergence between populations were the
233 same as those showing higher rates of evolution between cold-tolerant and non-cold-
234 tolerant species. We found 68 genes that had both an elevated rate of evolution between
235 species and significant divergence in at least one population comparison (**Supplementary**
236 **Table 19**). This is significantly greater than we would expect by chance (Fisher's exact test =
237 1.447, p = 0.0006) and implies that genes under divergent selection within species are also
238 more likely to diverge between species. The functions of these genes mirror those enriched
239 in each of the separate comparisons (transmembrane transport / ion transport (9/68), sexual
240 reproduction (16/68), and neurological system process/neurogenesis (15/68)), implicating
241 these genes' involvement in similar differences in cold adaptation and reproduction between
242 populations and species.
243

Discussion

Ecological studies with *Drosophila montana* have shown that it is able to thrive at high latitudes due to a number of adaptations including the evolution of increased cold tolerance and reproductive diapause. By sequencing the genome of this species we were able to use comparative genomics to identify genes and functional processes that differ between *D. montana* and its less cold-adapted relatives. We find evidence for selection acting on neuronal, membrane-transport and ion-transport related genes at both the inter- and intra-specific levels. These findings likely result from selection for an ability to overwinter under harsh environmental conditions, as these processes have clear links to both increased cold tolerance and reproductive diapause.

Genome assembly and features

We assembled the *D. montana* genome using a combination of Illumina paired-end reads and mate-pair reads. We annotated 13,683 genes, which is comparable to other Drosophila species that have been sequenced (Clark et al. 2007). 10,898 of these genes (80 %) were then assigned to a *D. virilis* ortholog, comparable to the number of orthologs identified between *D. melanogaster* and *D. simulans*. Together with the high BUSCO and CEGMA scores, this suggests that the genic component of the assembled genome is largely complete and successfully annotated.

Inter- and intra- specific comparisons show evidence for cold adaptation

Firstly, in the comparison between multiple *Drosophila* species, we identified 250 genes with an elevated rate of evolution between cold-tolerant and non-cold-tolerant *Drosophila* species. Interestingly, the increased rate of evolution was biased toward cold-tolerant species, with 77% of these genes showing a higher rate of evolution in these species. Secondly, we compared *D. montana* to its sequenced relative *D. virilis*. Although *D. montana* and *D. virilis* are both relatively cold-tolerant species, *D. montana* is significantly more cold-tolerant than *D. virilis* (Vesala et al. 2012), and *D. montana* is also more desiccation tolerant (Kellerman et al. 2012). In addition, unlike *D. virilis*, *D. montana* females enter reproductive diapause in late summer, which further increases their chances to survive over the cold season and produce progeny in spring (Watabe 1983). However, genes in the *D. montana* genome showed little evidence for divergent selection when compared to *D. virilis*, most genes showing evidence for purifying selection. Finally, we compared *D. montana* populations from Oulanka, Colorado and Vancouver. These populations face quite different abiotic and biotic conditions throughout the year, and hence can be expected to vary in several traits affecting flies' life-cycle and stress tolerances. We identified many SNPs that show significant divergence between the three populations; the number of divergent SNPs was smallest between Colorado and Vancouver populations reflecting the likely later divergence times of these populations. Although no divergent SNPs were shared between all population comparisons, when SNPs were grouped by gene, we found evidence for a significant number of overlapping genes. Divergent SNPs were overrepresented in the X chromosome which often shows elevated rates of evolution due to a combination of effects

289   including a smaller effective population size, increased efficacy of selection in hemizygous
290   males, and sexual antagonism. However, as some of the populations are known to differ in
291   sexual behaviour and post-mating pre-zygotic reproductive barriers (Klappert et al. 2007;
292   Jennings et al. 2014), as well as ecological adaptations, it is not possible to distinguish among
293   the multiple possible sources of any divergent selection on X-linked SNPs.

294

295   In most of these comparisons the genes with elevated dN/dS or F_{ST} were enriched for
296   functional processes previously demonstrated as important in cold adaptation (see below).
297   In particular changes to membrane components and ion transport, as well as in the
298   neurological system were heavily represented in our enrichment analyses in all comparisons.
299   In addition, we also found enrichment of many small-molecule binding terms, but the
300   specific terms enriched tended to be more varied across the different comparisons. Finally,
301   several comparisons were also enriched for many reproduction-associated terms, which are
302   unlikely to be linked to cold adaptation per se. We discuss each of these functional groups
303   below.

304

305   **Functional processes enriched in inter- and intra- specific comparisons**

306

307   *Cellular membranes*

308

309   The composition of the cell membrane is critical for maintenance of cellular function in sub-
310   optimal temperatures (Hazel 1995; Koštál et al. 2003) with changes to cell membrane
311   viscosity shown to be a critical component of cold acclimation in many species (Hazel 1995),
312   including *D. melanogaster* (Cooper et al. 2014). We found enrichment of many terms
313   associated with membrane structure (e.g. intrinsic component of membrane, integral
314   component of membrane, plasma membrane, transmembrane region, etc.) across all our
315   comparisons, providing further evidence for the importance of adjusting cell membrane
316   structure to better survive in cold environments. In addition to these terms, we also found
317   enrichment of other key processes that likely contribute to the functioning of cell
318   membranes at low temperatures. The most important of these are functions associated with
319   cellular ionic balance (e.g. ion channel activity, transmembrane transporter activity, calcium
320   transport, ion binding). Many of the mechanisms involved in the maintenance of cellular ion
321   balance are known to be temperature specific (Heitler et al. 1977; Kivivuori et al. 1990).
322   Failure to maintain the ionic balance of cells leads to metabolic perturbations which can
323   cause a wide range of negative consequences, including cellular damage and even death
324   (Hochachka 1986; Koštál et al. 2004). One class of cells particularly affected by low
325   temperature are neurons (Montgomery and Macdonald 1990; Janssen 1992; Robertson and
326   Money 2012) which are particularly susceptible to cold injury (Hochachka and Somero 2002).
327   In line with this we also found enrichment of several terms related to neuron function (cell
328   projection membrane, dendrite membrane, signal transducer activity, etc.). Finally, we
329   observed that membrane, ion transport, and neuronal terms often functionally clustered
330   together, showing that changes to each of these functions are in fact interrelated. Taken
331   together this suggests that the adjustment of cell membranes for increased cold tolerance is
332   complex, requiring changes to many genes to improve cellular functioning at low
333   temperatures.

334

335 *Small-molecule binding*

336

337 We observed enrichment of many small-molecule binding terms (small-molecule binding,
338 ATP-binding, kinase, nucleotide-binding, nucleotide phosphate-binding, carbohydrate
339 derivative binding, ribonucleotide binding, anion binding, etc.), both in the population and in
340 the multi-species comparisons. At low temperatures the activity levels of many reactions are
341 reduced meaning that during cold adaptation there is selection to adjust chemical reactions
342 to work better in cold environments (Margesin 2017). In particular ATP-binding and
343 associated terms were enriched in most of our comparisons suggesting that adjustments to
344 ATP-binding may be particularly important for cold adaptation. This finding is supported by
345 the fact that low temperatures adversely affect ATP metabolism across a broad range of taxa
346 (Napolitano and Shain 2004; Morrison and Shain 2008), including freeze-tolerant species like
347 terrestrial earthworm (*Enchytraeus albidus*) that are able to survive winters in a frozen state
348 (Boer et al. 2017).

349

350 *Reproduction*

351

352 Genes involved in reproduction typically show faster rates of divergence than other genes
353 (Swanson and Vacquier 2002; Clark et al. 2006). Consistent with this we find reproductive-
354 associated terms (male courtship behaviour, single organism reproductive process,
355 reproductive process) are enriched at each comparison level. Different species of Drosophila
356 (including *D. montana* and *D. virilis*) are known to vary for a number of reproductive traits
357 and so this finding is not too surprising. Interestingly, the only pairwise population
358 comparison that shows enrichment for reproductive-associated terms (reproductive
359 structure development, gonad development) is between Colorado and Vancouver. Although
360 all populations show some evidence of reproductive isolation, crosses between Colorado
361 and Vancouver showed the highest proportion of non-developing eggs (Jennings et al. 2014).
362 Moreover, although the exact cause of non-developing eggs is unknown, one possibility is
363 that it could be due to a negative interaction between sperm and the female reproductive
364 tract. Some support for this idea comes from examining the top differentiated genes
365 between Colorado and Vancouver which include the transcription factor *ken and barbie*
366 which has a major role in the development of genitalia of *D. melanogaster* (Lukacsovich et al.
367 2003).

368

369 **Functional processes enriched in specific-comparisons**

370

371 Although we observe many terms related to cold tolerance common to each of our
372 comparisons (described above), we also observe enrichment of several other functional
373 processes which are restricted to one or two of our comparisons. Of these two (Olfaction
374 and cuticular processes) are of particular interest due to their potential link to cold
375 adaptation and are discussed below.

376

377

378

379 *Cuticular related processes*

380

381 Cuticular and chitin related processes show an extensive enrichment in genes showing
382 elevated dN/dS between *D. montana* and *D. virilis,* but not in the multi-species or population
383 comparisons. Changes to the cuticle are linked to increased cold and desiccation tolerance in
384 insects (Gibbs 2002; Dennis et al. 2015) and in particular to enhancing the stress resistance
385 of the cuticle during diapause (Li and Denlinger 2009; Benoit 2010). This is particularly
386 interesting as *D. montana*, unlike *D. virilis*, has a reproductive diapause meaning the changes
387 we observe in cuticular related genes may have resulted from selection for increased stress
388 resistance to help *D. montana* successfully overwinter. This idea is consistent with the fact
389 that cuticular related processes are only found in the *D. montana – D. virilis* comparison, as
390 this is the only one of our comparisons that directly compares non-diapausing and
391 diapausing capable groups.

392

393 *Olfaction*

394

395 Drosophila flies have various kinds of olfaction-driven behaviors including the location of
396 food and mates (Amrein 2004; Libert et al. 2007) and the genomic repertoire of olfactory loci
397 is correlated with environmental variation (Gardiner et al. 2008). A cold environment may
398 affect the perception of olfactory signals as the detection of odorants at low temperatures is
399 more difficult due to the reduced concentration of olfactory cues in the air. Previous work in
400 *D. melanogaster* has shown that the sensitivity of the olfactory system increases in response
401 to cold temperature (e.g. Dalton (2000)), and that this change is accompanied by a change in
402 expression in olfactory genes (Riveron et al. 2009; Riveron et al. 2013). Since both sexual and
403 non-sexual olfactory signals are likely to be affected by colder temperatures, we hypothesize
404 that the changes in olfaction-related genes we observe in the present study are a product of
405 adaptation to living in a colder environment as well as of sexual selection to distinguish
406 conspecific flies from the heterospecific ones. Olfaction related terms were enriched in both
407 species-level comparisons, but not population comparisons.

408

409 **Population and species divergence at common loci**

410

411 Phenotypic variation in similar traits between and within species may or may not arise from
412 the same genes even when selection processes are similar (Wittkopp et al. 2009). Here we
413 find that genes which show divergence between populations were also more likely to show
414 elevated differences between species. The functions of these genes mirror those enriched in
415 each of the separate comparisons (transmembrane transport / ion transport sexual
416 reproduction and neurological system processes), implicating these genes' involvement in
417 similar differences in cold adaptation and reproduction between populations and species.
418 Although any of these genes may be important in cold adaptation, one gene in particular,
419 *Task6*, stands out as an interesting candidate. *Task6* encodes a subunit of two-pore domain
420 potassium (K2P) channels, which are important in setting the membrane potential and input
421 resistance of neurons in Drosophila (Döring et al. 2006). Temperature impacts a cell's ability
422 to maintain ionic balance, and in particular loss of potassium ion balance has been shown to
423 cause membrane depolarization, induction of chill-coma, and cell death (Andersen et al.

424   2015; Andersen et al. 2017). As such the changes we observe in *Task6* may be involved in
425   thermal adaptation of species and populations.
426
427   **Conclusion**
428
429   *D. montana* is an exceptional species of Drosophila in terms of cold adaptation, as well as a
430   species used for studies of behavioural variation and reproductive isolation. Here we report
431   the first description of its genome. Although there are few strong signals of divergent
432   selection on coding sequence variation, especially with its closest available relative, contrasts
433   between cold-adapted species and intraspecific population sequencing suggest that the
434   genome contains a clear signal of selection for cold tolerance. We identify many genes
435   potentially important in adaptation and speciation in this ecological specialist species.
436
437

MATERIALS AND METHODS

Samples and Sequencing

Genomic DNA for the *D. montana* reference genome was extracted from an inbred isofemale line originating from Vancouver, Canada (Can3F9) in summer 2003. This line was inbred via full-sib matings for 37 generations, relaxed for 9 generations and maintained on malt food (Lakovaara 1969) at 19°C in constant light. Quality checked DNA extracted from 210 males using a Gentra Puregene Tissue Kit (Qiagen) was used to produce 3 libraries with different insert sizes: 200bp, 400bp and 3,000bp. The 200bp and 400bp libraries were sequenced using an Illumina HiSeq 2000 at Edinburgh Genomics to produce paired-end reads (101 + 101bp). The 3,000bp library was sequenced using an Illumina MiSeq at The Centre for Genomic Research, University of Liverpool to produce mate-pair reads (101+101bp). This strategy produced 65107854 paired-end reads for the 200bp library, 25618163 paired-end reads for the 400bp library and 19020110 mate-pair reads for the 3000bp library. Reads from the 200bp and 400bp libraries were trimmed using scythe (Buffalo 2014) to remove adaptors and sickle (Joshi and Fass 2011) to quality trim reads (bases with phred quality of <20 were trimmed from the tail end of each read). Reads from the 3,000bp library were trimmed in the same manner, with the addition of a linker sequence removal step.

An initial assembly using reads from the 200bp and 400bp libraries was made using CLC assembly cell (4.0.12). Contigs from this were then blasted (blastN) to two subsets of NCBI's nt database (arthropod and bacteria) with a e-value threshold of 1×10^{-40}. Bit scores of blast hits from the arthropod and bacterial databases were compared for each contig, and any with a higher bit score for bacteria than arthropods were considered to be contaminants (**Supplementary Figure 9**). Reads were mapped to contigs identified as contaminants using BWA (v. 0.7.12) (Li and Durbin 2009) and then the unmapped reads were assembled using CLC assembly cell (4.0.12) (default options, minimum contig length = 200bp). Contigs were then scaffolded using the 3,000bp mate pair library using SSPACE-BASIC-2.0. This assembly contained 68950 scaffolds (N50 = 39341). This assembly was then further screened for contaminants using DeconSeq (v. 0.4.3) (Schmieder and Edwards 2011). Bacterial (2786) and viral (4359) genomes were downloaded from NCBI on January 20th 2016 and used as the contamination databases in DeconSeq along with the human genome (hg38). The *D. melanogaster* (r6.09) and *D. virilis* (r1.05) genomes were used as retention databases. DeconSeq identified 5208 scaffolds as contaminants, which were removed from our assembly. We then used this assembly for all subsequent analyses. To assess the completeness of our genome assembly we used the CEGMA analysis pipeline (v. 2.4) (Parra et al. 2007; Parra et al. 2009) which identifies the presence of 248 conserved eukaryote genes, and the BUSCO pipeline (v.1.22) (Simão et al. 2015) which identifies the presence of 2675 conserved arthropod genes.

Genome Annotation

Full details of the genome annotation are given in the supplementary methods. Briefly, we used the Maker2 pipeline (Holt and Yandell 2011) to first mask putative repeats within the genome, and then used *ab initio* gene predictors SNAP and AUGUSTUS, and gene evidence (from proteins homology and RNA-seq data) to generate gene predictions. Gene predictions from Maker2 were reciprocally blasted to proteins from *D. virilis* (r1.2) with the following cutoffs: e-value < 3 x10^{-13}, query cover > 60% to give reciprocal best blast hits (RBBH). Orthologs for *D. melanogaster*, *D. sechellia*, *D. simulans*, *D. erecta*, *D. yakuba*, *D. ananassae*, *D. persimilis*, *D. pseudoobscura*, *D. willistoni*, *D. mojavensis*, and *D. grimshawi* were then obtained from FlyBase using *D. virilis* FlyBase numbers. Genes without a single ortholog for each species were discarded from multi-species selection analyses (below).

Linkage map construction

For the genetic map construction, we selected 192 samples from a previous QTL study (Tyukmaeva et al. 2015), which consisted of two families (four parent individuals and their F2 progeny, females only). We used RAPiD Genomics' (Florida, USA) facilities to develop a set of oligonucleotide probes for 13,975 selected regions in the largest scaffolds of the *D. montana* genome. These probes were used to capture sequence these target loci with 100bp single end reads using an Illumina HiSeq 2000. A resulting SNP dataset was cleaned with Genotype Checker to eliminate possible errors in pedigree/genotyping (Paterson and Law 2011). The R/qtl package (Broman et al. 2003) was used to construct a genetic linkage map after discarding any polymorphic loci that were heterozygous for both parents, duplicated markers, markers showing segregation distortion, and individuals with fewer than 2000 markers. Reads from the 200bp and 400bp genome reference libraries were mapped back to anchored scaffolds using BWA (v. 0.7.12) (Li and Durbin 2009). Multi-mapping reads were discarded. Since the genome reference libraries were produced from males, X linked regions should have half the coverage of autosomal regions, we used the coverage of these scaffolds to validate our linkage map.

Selection Analyses

Multispecies analysis

13 species with fully annotated genome sequences available were divided into cold-tolerant and non-cold-tolerant ones; six species with a knockdown temperature <3°C (Kellermann et al. 2012; MacMillan et al. 2015) were classified as cold-tolerant, the remainder as non-cold-tolerant **Fig. 1**). This approach for classifying species was taken to a maximise the power of PAML's branch tests (see below). To identify genes showing elevated signatures of selection in these species we extracted the longest CDS (N = 5,619) for each ortholog and codon-aligned them using PRANK (v.140110) (Löytynoja and Goldman 2005). Sequences were then analysed in codeml from the PAML (v4.8) package (Yang 1997; Yang and Bielawski 2000). Two models were compared; the "null" model (clock = 0; fix_omega = 0, model = 0, NSSites = 0) which assumes a single common value for ω with an alternative model (clock = 0; fix_omega

525 = 0, model = 2, NSSites = 0) which assumes one value of omega for all the cold-tolerant
526 species and a separate value of omega for the non-cold-tolerant species. Nested models
527 were compared using a likelihood ratio test and p-values corrected for multiple testing using
528 a Bonferroni correction. Additionally, results were filtered to exclude sequences with dN, dS
529 or ω > 10. This comparison tests whether there is a different rate of molecular evolution in
530 cold-tolerant species compared to non-cold-tolerant species.
531
532 **Pairwise analysis**
533
534 To identify protein-coding genes with elevated signatures of selection we estimated pairwise
535 ω (dN/dS) for each gene that had a reciprocal best blast hit (RBBH) to a *D. virilis* gene. The
536 longest coding sequence of each gene and its RBBH ortholog were codon-aligned using
537 PRANK (v.140110) (Löytynoja and Goldman 2005), before estimating ω using codeml in PAML
538 (v. 4.8) (Yang 1997; Yang and Bielawski 2000). To determine if any genes showed ω > 1, we
539 compared genes using a Bayesian estimation of ω in codeml (runmode = -3, model = 0,
540 NSsites = 0) (Angelis et al. 2014) with default priors. The p-values were corrected for multiple
541 testing using a strict Bonferroni correction. We further filtered to exclude any genes where
542 estimates of *dN, dS* or ω were greater than 10.
543
544 We then compared mean ω values in several candidate gene sets (genes involved in immune
545 function, reproduction, and cold tolerance) against the genomic background. Genes were
546 classified into two 'immune' classes firstly using the GO term 'immune response' from
547 FlyBase (version 6.05) and secondly using orthologs of genes identified as being involved in
548 immune function by Sackton et al. (2007). Next, genes connected to reproduction were
549 classified into several reproductive classes following Haerty et al. (2007): sex and
550 reproduction related genes (SRR), female reproductive tract (FRTP) and seminal fluid
551 proteins (SFP). Finally, cold tolerance genes were classified into two classes with genes
552 differentially expressed in response to cold in *D. montana* and in *D. virilis* (Parker et al. 2015).
553 Parker et al. (2015) found that from the differentially expressed genes, 42 were the same in
554 both species but 550 were different, allowing genes to be classified into 'cold tolerance
555 same' and 'cold tolerance different' groups.
556
557
558 **Population resequencing**
559
560 For population comparisons we used *D. montana* flies from 3 populations: Oulanka (Finland;
561 66°N), Crested Butte, Colorado (USA; 39°N) and Vancouver (Canada; 49°N). These
562 populations were established from the progenies of fertilized females collected in the
563 summer of 2008 in Oulanka and Vancouver, and in the summer of 2009 in Colorado.
564 Population cages were set up using 20 F3 generation individuals from approximately 20
565 isofemale lines for each population. Population cages were maintained at 19°C in constant
566 light (for more details see Jennings et al. (2011)). In March 2013 Genomic DNA was extracted
567 from a pool of 50 females for each population and sequenced at Beijing Genomics Institute
568 using an Illumina HiSeq 2000 to produce paired-end reads (90 + 90bp, insert size = 500bp).
569

570   Sequencing produced 84938118 paired-end reads for Colorado and 82663801 for Oulanka.
571   Two runs for Vancouver resulted in 303365095 reads. Reads were quality trimmed (leading
572   or trailing bases with a phred score of <20, or if two consecutive bases had an average phred
573   score of <32 the read was trimmed at this point) and screened for adaptor sequence using
574   trimmomatic (v. 0.30) (Bolger et al. 2014). Reads containing adaptor sequence or that had a
575   length of less than 85 bp after quality trimming, were discarded. Since coverage depth can
576   influence the estimation of allele frequency (Zhu et al. 2012), reads for Vancouver were
577   randomly sampled prior to mapping to the mean number of reads from Colorado and
578   Oulanka. Reads were mapped to the genome assembly using BWA (v. 0.7.12) (Li and Durbin
579   2009). Reads with a mapping quality of <20 were then removed, and an mpileup file was
580   produced using samtools (v. 0.1.19) (Li et al. 2009). From this, a sync file was produced using
581   PoPoolation2 pipeline (v 1.201) (Kofler et al. 2011). Outlier detection was performed on the
582   raw read count data with BayeScan v. 2.1 (Foll and Gaggiotti 2008; Foll et al. 2010; Fischer et
583   al. 2011), which performs comparably alongside other outlier methods in several simulation
584   studies (Pérez-Figueroa et al. 2010; Vilas et al. 2012; Villemereuil et al. 2014). SNPs were
585   filtered to include only sites with a minimum coverage of 25 and a maximum coverage of 93
586   (corresponding to the median 10th and 90th percentiles of the population coverage
587   distributions). At the same time, SNPs were only considered if the minor allele had a read
588   count > 4 across all populations. BayeScan was run with 5 pilot runs of 1,000 iterations each
589   followed by a main run of 2,000 iterations, a thinning interval of 10 and a burn in of 1,500.
590   Additionally, three pairwise runs of BayeScan were performed with the same parameters as
591   above. The three pairwise analyses compared Colorado to Vancouver, Vancouver to Oulanka,
592   and Colorado to Oulanka populations, respectively.

593

594   **Functional Enrichment**

595

596   To examine functional enrichment of genes for the species level selection analyses and
597   population level F_{ST} scans, we used GOrilla (Eden et al. 2009). For the pairwise selection
598   analyses genes were ranked by ω (from high to low and low to high). For the multispecies
599   selection analyses, we ranked genes by p-value and direction so that genes with the lowest
600   p-values and a higher ω in cold-tolerant species were at the top, and genes with lowest p-
601   values and higher ω in non-cold-tolerant species were at the bottom, allowing us to identify
602   enriched GO terms for genes showing elevated ω in cold-tolerant species. To examine GO
603   terms for genes showing elevated ω in non-cold-tolerant species the list order was simply
604   reversed. For population level analyses genes were ranked by the most significantly
605   differentiated SNP occurring within 1kb, 10kb, or 100kb of a gene for each population.
606   Results from GOrilla were then visualised using ReviGO (Supek et al. 2011), using the January
607   2017 version of Gene Ontology.

608

609   We used DAVID (v6.8) (Huang et al. 2009a; Huang et al. 2009b) to identify enriched
610   functional groups of genes. A functional group was considered to be significantly enriched if
611   its enrichment score (the geometric mean (in -log scale) of the p-values of the GO terms in
612   the group) was >1 (p < 0.1). For the pairwise selection analyses we identified functional
613   clusters for genes occurring in the top and bottom 10% of genes for ω estimates. For the
614   multispecies selection analyses we identified functional clusters for genes that showed a

615   significantly (FDR < 0.1) higher omega in cold-tolerant species or in non-cold-tolerant species

616   separately. For population level analyses we identified functional clusters for genes

617   containing (within 1kb) significantly differentiated SNPs for each population.

618

619   To take advantage of the superior annotation of *D. melanogaster* (Tweedie et al. 2009), we

620   used *D. melanogaster* orthologs for all of the above function enrichment analyses. For the

621   DAVID analyses the 'background' list used was the subset of *D. melanogaster* genes available

622   for each analysis.

623

Acknowledgements

Data availability

This project has been deposited at NCBI under the BioProject accession PRJNA312336. The accession number for the assembly is LUVX00000000. Raw reads were deposited in the SRA under the following accession numbers: mate-pair reads: SRX1604922, paired-end reads: SRX1602883, SRX1602879, population resequencing reads: SRX1625831, SRX1625832, SRX1625834.

Authors' contributions

DJP, AH, MK, and MGR designed the study. KG, UT and RKB conceived the sequencing strategy. DJP and MK performed DNA extractions and original quality checking. DJP, RAWW, UT, and VI, performed the bioinformatics analyses. DJP, AH, MK, and MGR drafted the manuscript, with input from all authors. All authors agreed on the final version of the manuscript.

REFERENCES

Amrein H. 2004. Pheromone perception and behavior in Drosophila. Curr. Opin. Neurobiol. 14:435–442.

Andersen JL, Manenti T, Sørensen JG, MacMillan HA, Loeschcke V, Overgaard J. 2015. How to assess Drosophila cold tolerance: chill coma temperature and lower lethal temperature are the best predictors of cold distribution limits. Funct. Ecol. 29:55–65.

Andersen MK, Jensen SO, Overgaard J. 2017. Physiological correlates of chill susceptibility in Lepidoptera. J. Insect Physiol. 98:317–326.

Angelis K, Dos Reis M, Yang Z. 2014. Bayesian estimation of nonsynonymous/synonymous rate ratios for pairwise sequence comparisons. Mol. Biol. Evol. 31:1902–1913.

Benoit JB. 2010. Water Management by Dormant Insects: Comparisons Between Dehydration Resistance During Summer Aestivation and Winter Diapause. In: Aestivation. Progress in Molecular and Subcellular Biology. Springer, Berlin, Heidelberg. p. 209–229.

Boer TE, Roelofs D, Vooijs R, Holmstrup M, Amorim MJ. 2017. Population-specific transcriptional differences associated with freeze tolerance in a terrestrial worm. Ecol. Evol.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Broman KW, Wu H, Sen Ś, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890.

Buffalo V. 2014. Scythe. Available at https://github.com/vsbuffalo/scythe

Chen C, Liu Z, Pan Q, Chen X, Wang H, Guo H, Liu S, Lu H, Tian S, Li R. 2016. Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp. Mol. Biol. Evol. 33:1337–1348.

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203–218.

Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. Reproduction 131:11–22.

Cooper BS, Hammad LA, Montooth KL. 2014. Thermal adaptation of cellular membranes in natural populations of Drosophila melanogaster. Funct. Ecol. 28:886–894.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.

Dalton P. 2000. Psychophysical and behavioral characteristics of olfactory adaptation. Chem. Senses 25:487–492.

Dennis AB, Dunning LT, Sinclair BJ, Buckley TR. 2015. Parallel molecular routes to cold adaptation in eight genera of New Zealand stick insects. Sci. Rep. 5:13965.

Döring F, Scholz H, Kühnlein RP, Karschin A, Wischmeyer E. 2006. Novel Drosophila two-pore domain K+ channels: rescue of channel function by heteromeric assembly. Eur. J. Neurosci. 24:2264–2274.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10:48.

698 Ellegren H. 2008. Comparative genomics and the study of evolution by natural selection.
699   Mol. Ecol. 17:4586–4596.
700 Fischer MC, Foll M, Excoffier L, Heckel G. 2011. Enhanced AFLP genome scans detect local
701   adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). Mol.
702   Ecol. 20:1450–1462.
703 Foll M, Fischer MC, Heckel G, Excoffier L. 2010. Estimating population structure from AFLP
704   amplification intensity. Mol. Ecol. 19:4638–4647.
705 Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for
706   both dominant and codominant markers: a bayesian perspective. Genetics 180:977–
707   993.
708 Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L. 2014. Widespread signals of convergent
709   adaptation to high altitude in Asia and America. Am. J. Hum. Genet. 95:394–407.
710 Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. 2008. Drosophila chemoreceptor
711   gene evolution: selection, specialization and genome size. Mol. Ecol. 17:1648–1657.
712 Gibbs AG. 2002. Lipid melting and cuticular permeability: new insights into an old problem. J.
713   Insect Physiol. 48:391–400.
714 Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG,
715   Wolfner MF, Civetta A, et al. 2007. Evolution in the fast lane: rapidly evolving sex-
716   related genes in *Drosophila*. Genetics 177:1321–1335.
717 Hazel JR. 1995. Thermal adaptation in biological membranes - is homeoviscous adaptation
718   the explanation? Annu. Rev. Physiol. 57:19–42.
719 Heitler WJ, Goodman CS, Fraserrowell CH. 1977. Effects of temperature on threshold of
720   identified neurons in locust. J. Comp. Physiol. 117:163–182.
721 Hochachka PW. 1986. Defense strategies against hypoxia and hypothermia. Science 231:234–
722   241.
723 Hochachka PW, Somero GN. 2002. Temperature. In: Biochemical Adaptation: Mechanism and
724   Process in Physiological Evolution. New York: Oxford University Press. p. 290–449.
725 Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database
726   management tool for second-generation genome projects. BMC Bioinformatics
727   12:491.
728 Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward
729   the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37:1–13.
730 Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large
731   gene lists using DAVID bioinformatics resources. Nat. Protoc. 4:44–57.
732 Janssen R. 1992. Thermal influences on nervous system function. Neurosci. Biobehav. Rev.
733   16:399–413.
734 Jennings JH, Mazzi D, Ritchie MG, Hoikkala A. 2011. Sexual and postmating reproductive
735   isolation between allopatric *Drosophila montana* populations suggest speciation
736   potential. BMC Evol. Biol. 11:68.
737 Jennings JH, Snook RR, Hoikkala A. 2014. Reproductive isolation among allopatric *Drosophila
738   montana* populations. Evolution 68:3095–3108.
739 Joshi NA, Fass JN. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for
740   FastQ files. Available at https://github.com/najoshi/sickle.

741 Kankare M, Parker DJ, Merisalo M, Salminen TS, Hoikkala A. 2016. Transcriptional differences
742   between diapausing and non-diapausing *D. montana* females reared under the same
743   photoperiod and temperature. PloS One 11:e0161852.
744 Kauranen H, Ala-Honkola O, Kankare M, Hoikkala A. 2016. Circadian clock of *Drosophila*
745   *montana* is adapted to high variation in summer day lengths and temperatures
746   prevailing at high latitudes. J. Insect Physiol. 89:9–18.
747 Kauranen H, Menegazzi P, Costa R, Helfrich-Förster C, Kankainen A, Hoikkala A. 2012. Flies in
748   the north: locomotor behavior and clock neuron organization of *Drosophila*
749   *montana*. J. Biol. Rhythms 27:377–387.
750 Kellermann V, Loeschcke V, Hoffmann AA, Kristensen TN, Flojgaard C, David JR, Svenning J-C,
751   Overgaard J. 2012. Phylogenetic constraints in key functional traits behind species'
752   climate niches: patterns of desiccation and cold resistance across 95 *Drosophila*
753   species. Evolution 66:3377–3389.
754 Kivivuori L, Lehti S, Lagerspetz KYH. 1990. Effect of temperature-acclimation on thermal-
755   dependence and hysteresis of the resting membrane-potential of the stretch-
756   receptor neuron in crayfish *Astacus astacus* (L). J. Therm. Biol. 15:9–14.
757 Klappert K, Mazzi D, Hoikkala A, Ritchie MG. 2007. Male courtship song and female
758   preference variation between phylogeographically distinct populations of *Drosophila*
759   *montana*. Evolution 61:1481–1488.
760 Kofler R, Pandey RV, Schlötterer C. 2011. PoPoolation2: identifying differentiation between
761   populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics
762   27:3435–3436.
763 Koštál V, Berková P, Šimek P. 2003. Remodelling of membrane phospholipids during transition
764   to diapause and cold-acclimation in the larvae of *Chymomyza costata*
765   (Drosophilidae). Comp. Biochem. Physiol. B Biochem. Mol. Biol. 135:407–419.
766 Koštál V, Vambera J, Bastl J. 2004. On the nature of pre-freeze mortality in insects: water
767   balance, ion homeostasis and energy charge in the adults of *Pyrrhocoris apterus*. J.
768   Exp. Biol. 207:1509–1521.
769 Lakovaara S. 1969. Malt as a culture medium for Drosophila species. Drosoph. Inf Serv
770   44:128.
771 Lankinen P, Tyukmaeva VI, Hoikkala A. 2013. Northern *Drosophila montana* flies show
772   variation both within and between cline populations in the critical day length
773   evoking reproductive diapause. J. Insect Physiol. 59:745–751.
774 Li A, Denlinger DL. 2009. Pupal cuticle protein is abundant during early adult diapause in the
775   mosquito Culex pipiens. J. Med. Entomol. 46:1382–1386.
776 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler
777   transform. Bioinformatics 25:1754–1760.
778 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
779   others. 2009. The sequence alignment/map format and SAMtools. Bioinformatics
780   25:2078–2079.
781 Libert S, Zwiener J, Chu X, VanVoorhies W, Roman G, Pletcher SD. 2007. Regulation of
782   Drosophila life span by olfaction and food-derived odors. Science 315:1133–1137.
783 Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of
784   sequences with insertions. Proc. Natl. Acad. Sci. U. S. A. 102:10557–10562.

785   Lukacsovich T, Yuge K, Awano W, Asztalos Z, Kondo S, Juni N, Yamamoto D. 2003. The *ken and*
786   *barbie* gene encoding a putative transcription factor with a BTB domain and three
787   zinc finger motifs functions in terminalia development of Drosophila. Arch. Insect
788   Biochem. Physiol. 54:77–94.
789   MacMillan HA, Ferguson LV, Nicolai A, Donini A, Staples JF, Sinclair BJ. 2015. Parallel
790   ionoregulatory adjustments underlie phenotypic plasticity and evolution of
791   Drosophila cold tolerance. J. Exp. Biol. 218:423–432.
792   Margesin R. 2017. Psychrophiles: From Biodiversity to Biotechnology. London, UK: Springer
793   Matzkin LM, Watts TD, Bitler BG, Machado CA, Markow TA. 2006. Functional genomics of
794   cactus host shifts in *Drosophila mojavensis*. Mol. Ecol. 15:4635–4643.
795   McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, Mitchell RF, Waterhouse RM,
796   Ahn S-J, Arsala D. 2016. Genome of the Asian longhorned beetle (*Anoplophora*
797   *glabripennis*), a globally significant invasive species, reveals key functional and
798   evolutionary innovations at the beetle–plant interface. Genome Biol. 17:227.
799   Menegazzi P, Dalla Benetta E, Beauchamp M, Schlichting M, Steffan-Dewenter I, Helfrich-
800   Förster C. 2017. Adaptation of circadian neuronal network to photoperiod in high-
801   latitude European drosophilids. Curr. Biol. 27:833–839.
802   Mirol PM, Schäfer MA, Orsini L, Routtu J, Schlötterer C, Hoikkala A, Butlin RK. 2007.
803   Phylogeographic patterns in *Drosophila montana*. Mol. Ecol. 16:1085–1097.
804   Montgomery JC, Macdonald JA. 1990. Effects of temperature on nervous system:
805   implications for behavioral performance. Am. J. Physiol. - Regul. Integr. Comp.
806   Physiol. 259:R191–R196.
807   Morrison BA, Shain DH. 2008. An AMP nucleosidase gene knockout in *Escherichia coli*
808   elevates intracellular ATP levels and increases cold tolerance. Biol. Lett. 4:53–56.
809   Napolitano MJ, Shain DH. 2004. Four kingdoms on glacier ice: convergent energetic
810   processes boost energy levels as temperatures fall. Proc. R. Soc. Lond. B Biol. Sci.
811   271:S273–S276.
812   Noor MA, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the
813   role of restricted recombination in maintaining species. Heredity 103:439.
814   Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic
815   divergence. Mol. Ecol. 18:375–402.
816   Ostrega MS. 1985. Restriction endonuclease analysis of the relatedness of *D. montana* and *D.*
817   *virilis* lines. Drosoph. Inf Serv 61:132–133.
818   Parker DJ, Ritchie MG, Kankare M. 2016. Preparing for Winter: the Transcriptomic Response
819   Associated with Different Day Lengths in *Drosophila montana*. G3
820   GenesGenomesGenetics:g3.116.027870.
821   Parker DJ, Vesala L, Ritchie MG, Laiho A, Hoikkala A, Kankare M. 2015. How consistent are
822   the transcriptome changes associated with cold acclimation in two species of the
823   *Drosophila virilis* group? Heredity 115:13–21.
824   Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in
825   eukaryotic genomes. Bioinformatics 23:1061–1067.
826   Parra G, Bradnam K, Ning ZM, Keane T, Korf I. 2009. Assessing the gene space in draft
827   genomes. Nucleic Acids Res. 37:289–297.
828   Paterson T, Law A. 2011. Genotypechecker: an interactive tool for checking the inheritance
829   consistency of genotyped pedigrees. Anim. Genet. 42:560–562.

830   Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A. 2010. Comparing
831   three different methods to detect selective loci using dominant markers. J. Evol. Biol.
832   23:2267–2276.

833   Radwan J, Babik W. 2012. The genomics of adaptation. Proc. R. Soc. Lond. B Biol. Sci.
834   279:5024–5028.

835   Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, Noor MAF, Mehlig B,
836   Westram AM. 2017. Interpreting the genomic landscape of speciation: a road map
837   for finding barriers to gene flow. J. Evol. Biol. 30:1450–1477.

838   Riveron J, Boto T, Alcorta E. 2009. The effect of environmental temperature on olfactory
839   perception in *Drosophila melanogaster*. J. Insect Physiol. 55:943–951.

840   Riveron J, Boto T, Alcorta E. 2013. Transcriptional basis of the acclimation to high
841   environmental temperature at the olfactory receptor organs of *Drosophila*
842   *melanogaster*. BMC Genomics 14:259.

843   Robertson RM, Money TG. 2012. Temperature and neuronal circuit function: compensation,
844   tuning and tolerance. Curr. Opin. Neurobiol. 22:724–734.

845   Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007. Dynamic
846   evolution of the innate immune system in Drosophila. Nat. Genet. 39:1461–1468.

847   Salminen TS, Vesala L, Laiho A, Merisalo M, Hoikkala A, Kankare M. 2015. Seasonal gene
848   expression kinetics between diapause phases in *Drosophila virilis* group species and
849   overwintering differences between diapausing and non-diapausing females. Sci. Rep.
850   5.

851   Schäfer MA, Mazzi D, Klappert K, Kauranen H, Vieira J, Hoikkala A, Ritchie MG, Schlötterer C.
852   2010. A microsatellite linkage map for *Drosophila montana* shows large variation in
853   recombination rates, and a courtship song trait maps to an area of low
854   recombination. J. Evol. Biol. 23:518–527.

855   Schmieder R, Edwards R. 2011. Fast Identification and Removal of Sequence Contamination
856   from Genomic and Metagenomic Datasets. PLOS ONE 6:e17288.

857   Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing
858   genome assembly and annotation completeness with single-copy orthologs.
859   Bioinforma. Oxf. Engl. 31:3210–3212.

860   Smadja CM, Butlin RK. 2011. A framework for comparing processes of speciation in the
861   presence of gene flow. Mol. Ecol. 20:5123–5140.

862   Smith G, Fang Y, Liu X, Kenny J, Cossins AR, Oliveira CC, Etges WJ, Ritchie MG. 2013.
863   Transcriptome-wide expression variation associated with environmental plasticity
864   and mating success in cactophilic *Drosophila mojavensis*. Evolution 67:1950–1963.

865   Spicer GS, Bell CD. 2002. Molecular phylogeny of the *Drosophila virilis* species group
866   (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA
867   genes. Ann. Entomol. Soc. Am. 95:156–161.

868   Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO Summarizes and Visualizes Long Lists of
869   Gene Ontology Terms. PLOS ONE 6:e21800.

870   Swanson WJ, Vacquier VD. 2002. Reproductive Protein Evolution. Annu. Rev. Ecol. Syst.
871   33:161–179.

872   Tapanainen R, Parker DJ, Kankare M. 2018. Photosensitive Alternative Splicing of the
873   Circadian Clock Gene timeless Is Population Specific in a Cold-Adapted Fly,
874   Drosophila montana. G3 GenesGenomesGenetics 8:1291–1297.

875   Terhzaz S, Teets NM, Cabrero P, Henderson L, Ritchie MG, Nachman RJ, Dow JA, Denlinger DL,
876   Davies S-A. 2015. Insect capa neuropeptides impact desiccation and cold tolerance.
877   Proc. Natl. Acad. Sci. U. S. A. [Internet]. Available from:
878   http://www.pnas.org/content/early/2015/02/12/1501518112.short
879   Throckmorton LH. 1982. The virilis species group. In: Ashburner M, Carson HL, Thompson JN,
880   editors. Genetics and Biology of Drosophila. Vol. 3b. London, UK: Academic Press. p.
881   227–296.
882   Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, Millburn G, Osumi-
883   Sutherland D, Schroeder A, Seal R, et al. 2009. FlyBase: enhancing *Drosophila* Gene
884   Ontology annotations. Nucleic Acids Res. 37:D555–D559.
885   Tyukmaeva VI, Salminen TS, Kankare M, Knott KE, Hoikkala A. 2011. Adaptation to a
886   seasonally varying environment: a strong latitudinal cline in reproductive diapause
887   combined with high gene flow in *Drosophila montana*. Ecol. Evol. 1:160–168.
888   Tyukmaeva VI, Veltsos P, Slate J, Gregson E, Kauranen H, Kankare M, Ritchie MG, Butlin RK,
889   Hoikkala A. 2015. Localization of quantitative trait loci for diapause and other
890   photoperiodically regulated life history traits important in adaptation to seasonally
891   varying environments. Mol. Ecol. 24:2809–2819.
892   Vesala L, Hoikkala A. 2011. Effects of photoperiodically induced reproductive diapause and
893   cold hardening on the cold tolerance of *Drosophila montana*. J. Insect Physiol.
894   57:46–51.
895   Vesala L, Salminen TS, Laiho A, Hoikkala A, Kankare M. 2012. Cold tolerance and cold-
896   induced modulation of gene expression in two *Drosophila* virilis group species with
897   different distributions. Insect Mol. Biol. 21:107–118.
898   Vigoder FM, Parker DJ, Cook N, Tournière O, Sneddon T, Ritchie MG. 2016. Inducing Cold-
899   Sensitivity in the Frigophilic Fly Drosophila montana by RNAi. PloS One 11:e0165724.
900   Vilas A, Pérez-Figueroa A, Caballero A. 2012. A simulation study on the performance of
901   differentiation-based methods to detect selected loci using linked neutral markers. J.
902   Evol. Biol. 25:1364–1376.
903   Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE. 2014. Genome scan methods
904   against more complex models: when and how much should we trust them? Mol.
905   Ecol. 23:2006–2019.
906   Watabe H. 1983. Photoperiodic responses in the *Drosophila virilis* species group (Diptera,
907   Drosophilidae) from Japan. Kontyo 51:628–634.
908   Wittkopp PJ, Stewart EE, Arnold LL, Neidert AH, Haerum BK, Thompson EM, Akhras S, Smith-
909   Winberry G, Shefner L. 2009. Intraspecific polymorphism to interspecific divergence:
910   genetics of pigmentation in Drosophila. Science 326:540–544.
911   Wolf JBW, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of
912   speciation. Nat. Rev. Genet. 18:87–100.
913   Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood.
914   Comput. Appl. Biosci. 13:555–556.
915   Yang ZH, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends
916   Ecol. Evol. 15:496–503.
917   Zhu Y, Bergland AO, González J, Petrov DA. 2012. Empirical validation of pooled whole
918   genome population re-sequencing in *Drosophila melanogaster*. PloS One 7:e41901.
919
920

921   **Table 1 | Summary statistics of *D. montana* genome assembly**

922

Metric	Value
Total assembled length (bp)	183585048
Scaffolds (n)	63742
Scaffold N50 (bp)	40647
Largest scaffold (bp)	515352
GC content (%)	40.57
Number of predicted gene models	13683
Number of predicted gene models with RBBH to *D. virilis* genome	10898
CEGMA pipeline analysis (% complete/partial)	95.97/98.39
BUSCO (% complete/missing)	91.85/1.72

923

924

925

926
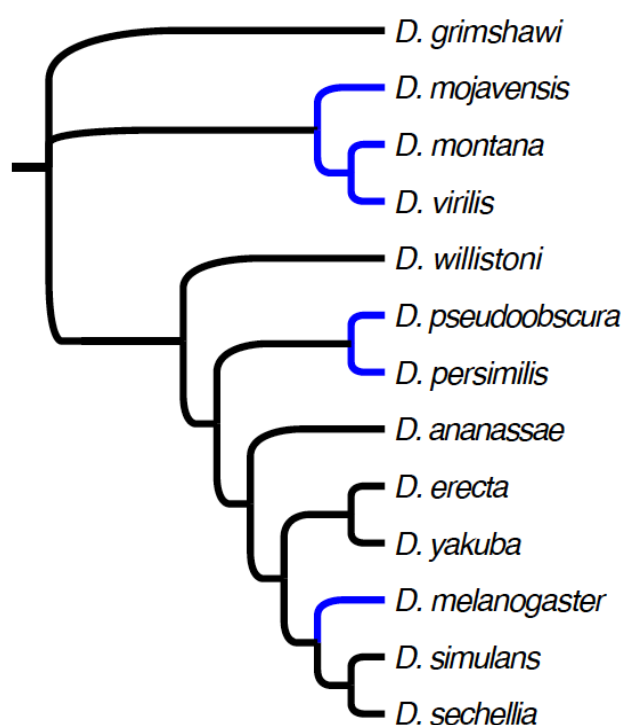
927

928 **Figure 1 |** Tree used for multi-species PAML analyses. Cold-tolerant species (species that
929 have a knockdown temperature of <3°C) are shown in blue (data from Kellermann et al.
930 (2012) and MacMillan et al. (2015)).
931
932

933
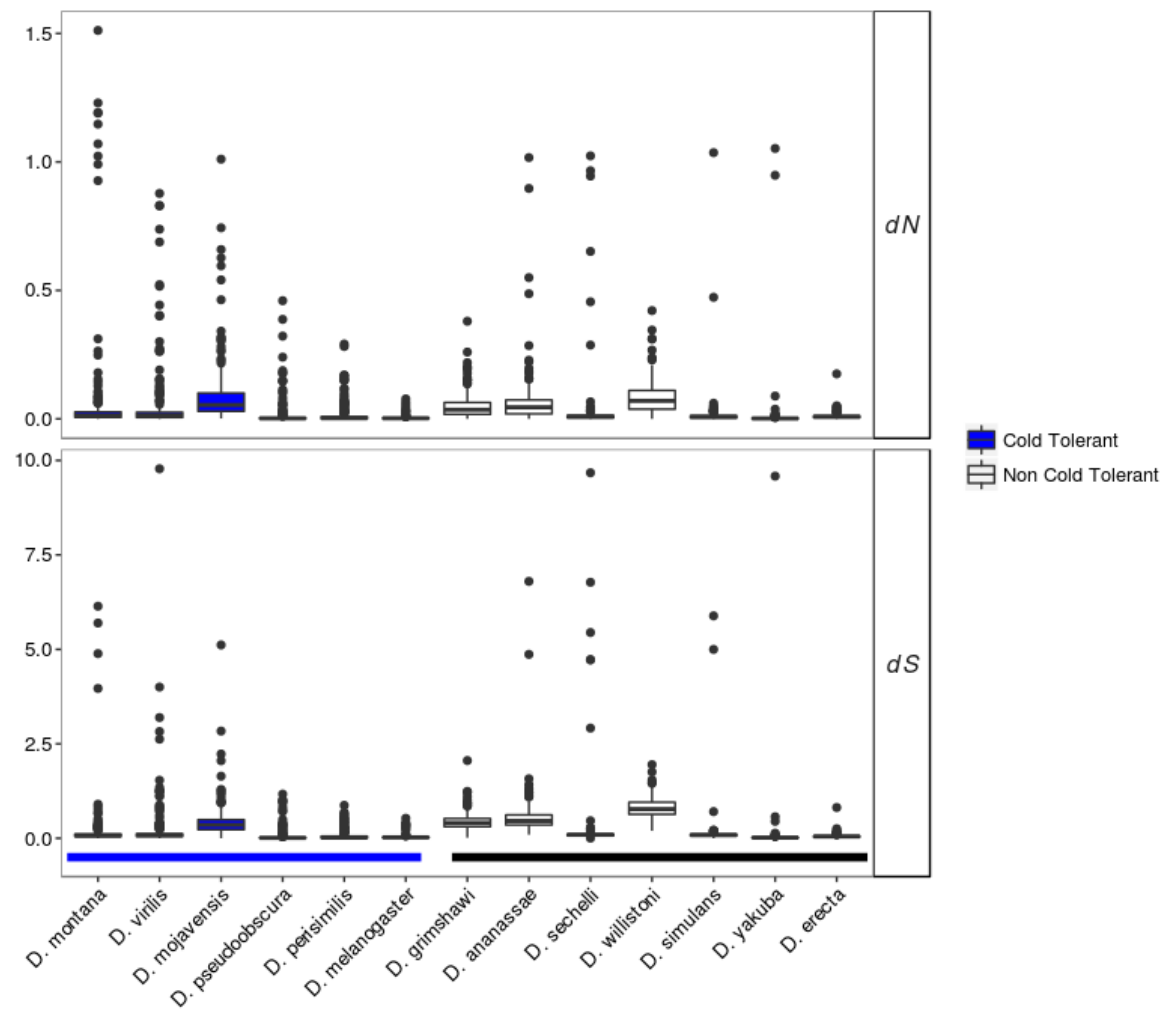
934

935 **Figure 2 |** Distributions of dN and dS estimates for each of the 250 genes from 13 *Drosophila*
936 species with significant differences in omega between cold-tolerant and non-cold-tolerant
937 species.

938
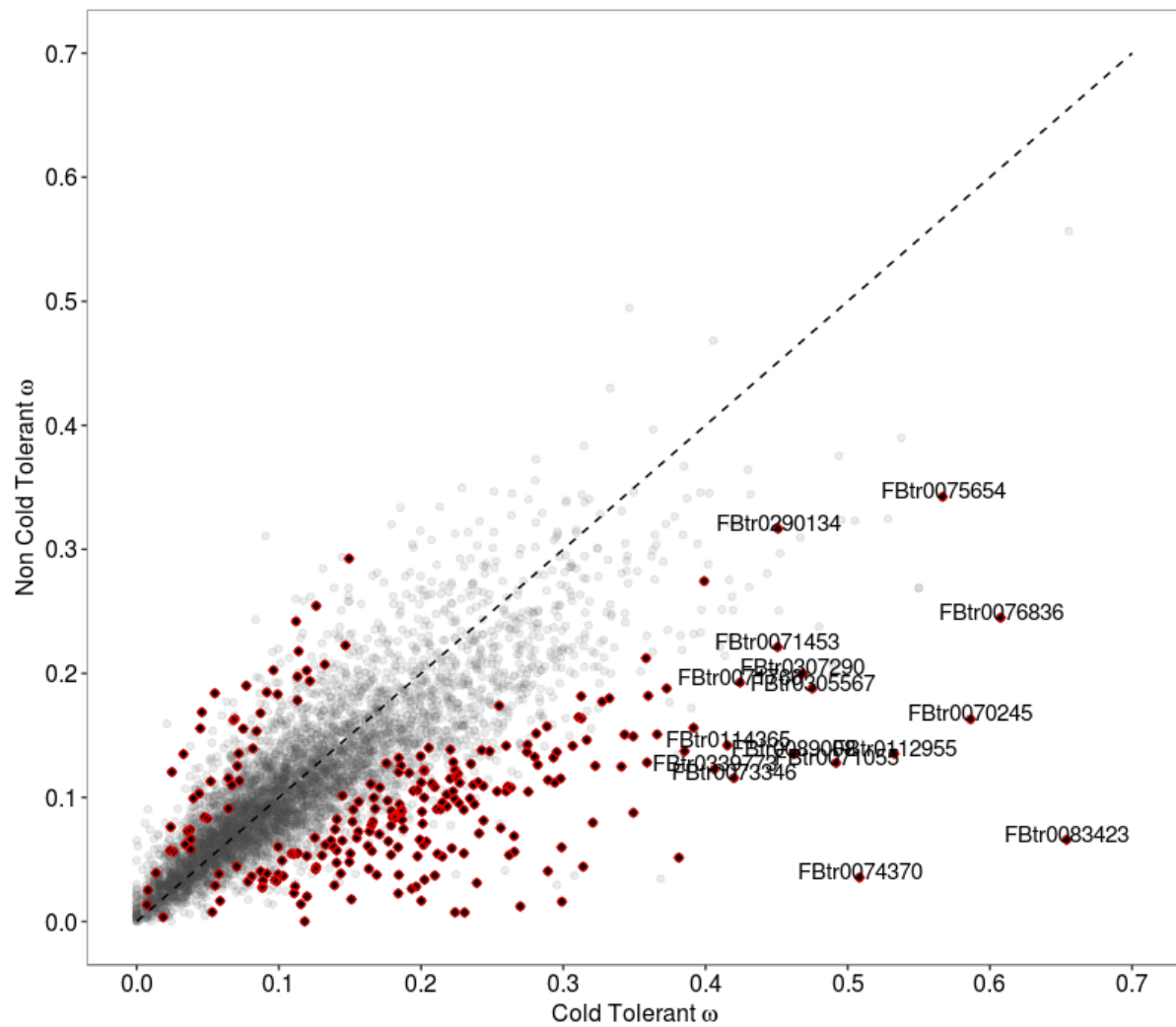
Figure 3 | The relationship between values of omega estimated for 5,619 genes in cold-tolerant and non-cold-tolerant species of Drosophila. 250 genes with significantly different estimates of omega are shown in black with red outline. Diagonal line indicates the 1-1 diagonal, points below the diagonal line show elevated levels of omega in cold-tolerant species compared to non-cold-tolerant species, while points above the diagonal show elevated levels of omega in non-cold-tolerant species.

945

946
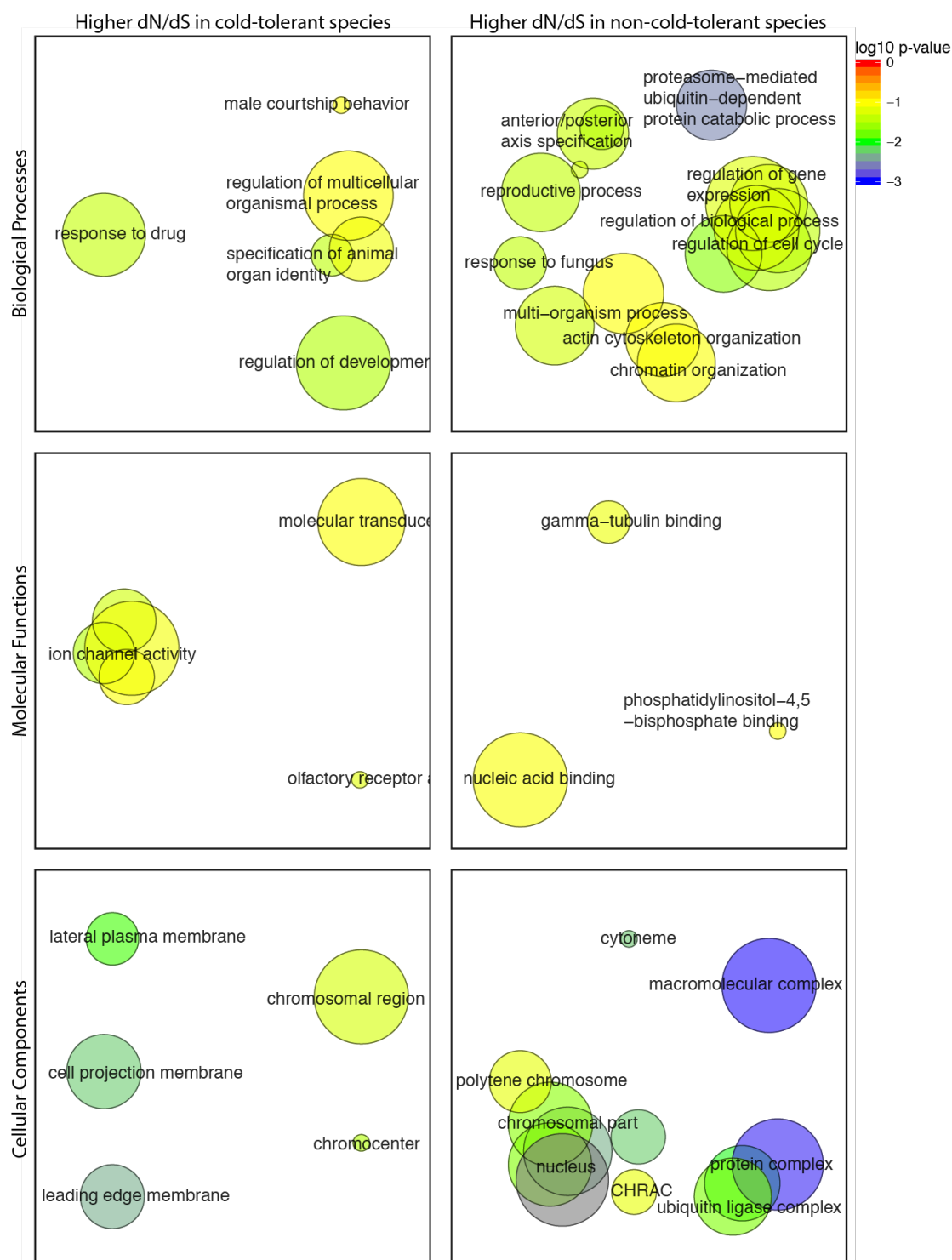
Figure 4 | Semantic clustering of significantly (FDR < 0.1) enriched GO-terms for genes showing significantly elevated dN/dS in cold-tolerant or non-cold-tolerant species. Circle size corresponds to the number of genes annotated to the term in the reference database. Circle colour indicates $\log_{10}$ FDR of the GO term
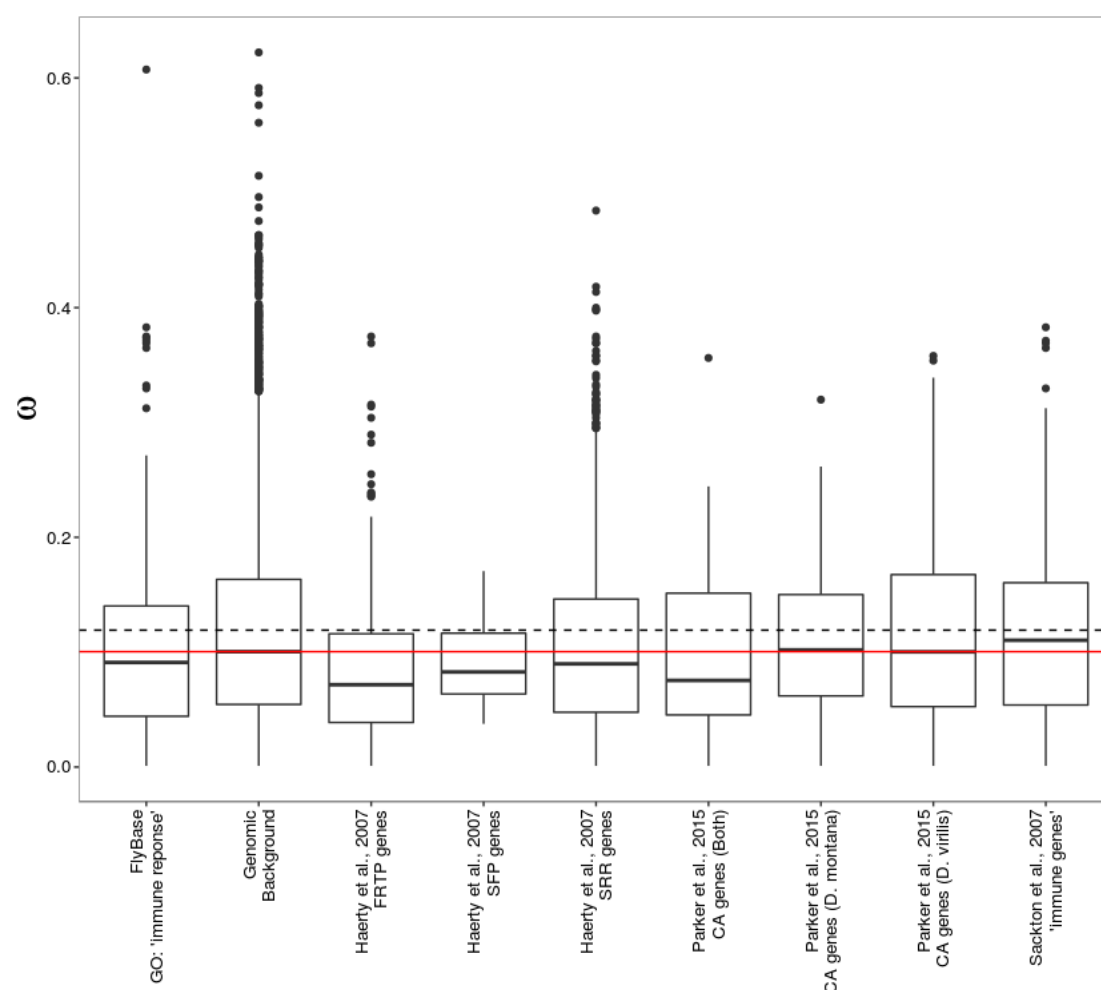
951

952

953

954

Figure 5 | Average values of ω between *D. montana* and *D. virilis* for candidate gene sets. FRTP = female reproductive tract SFP = seminal fluid proteins SRR = sex and reproduction related genes CA = cold acclimation genes. The red and dashed lines indicate the median and mean ω of the genomic background respectively.
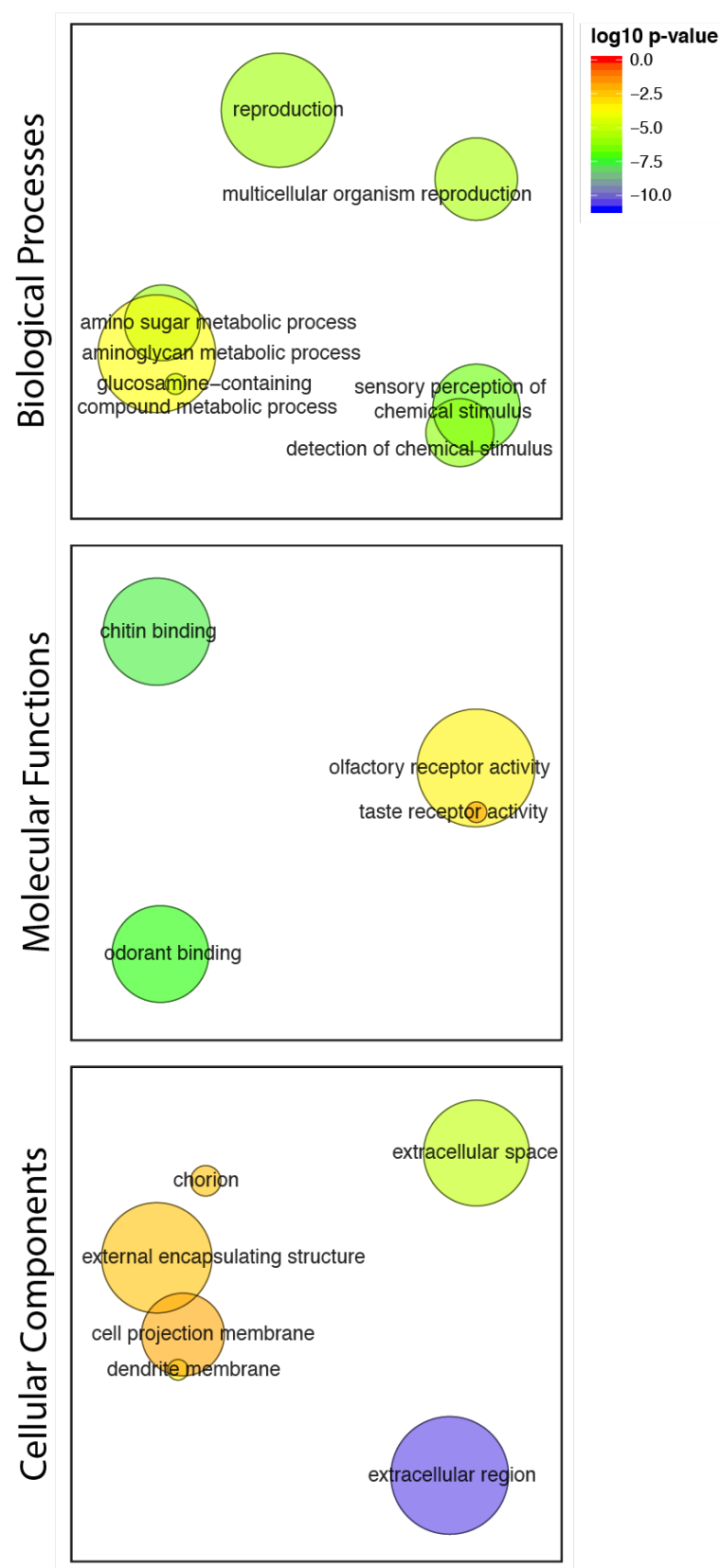
962

Figure 6 | Semantic clustering of significantly (FDR < 0.1) enriched GO-terms for genes showing high dN/dS between *D. montana* and *D. virilis*. Circle size corresponds to the number of genes annotated to the term in the reference database. Circle colour indicates $\log_{10}$ FDR of the GO term.
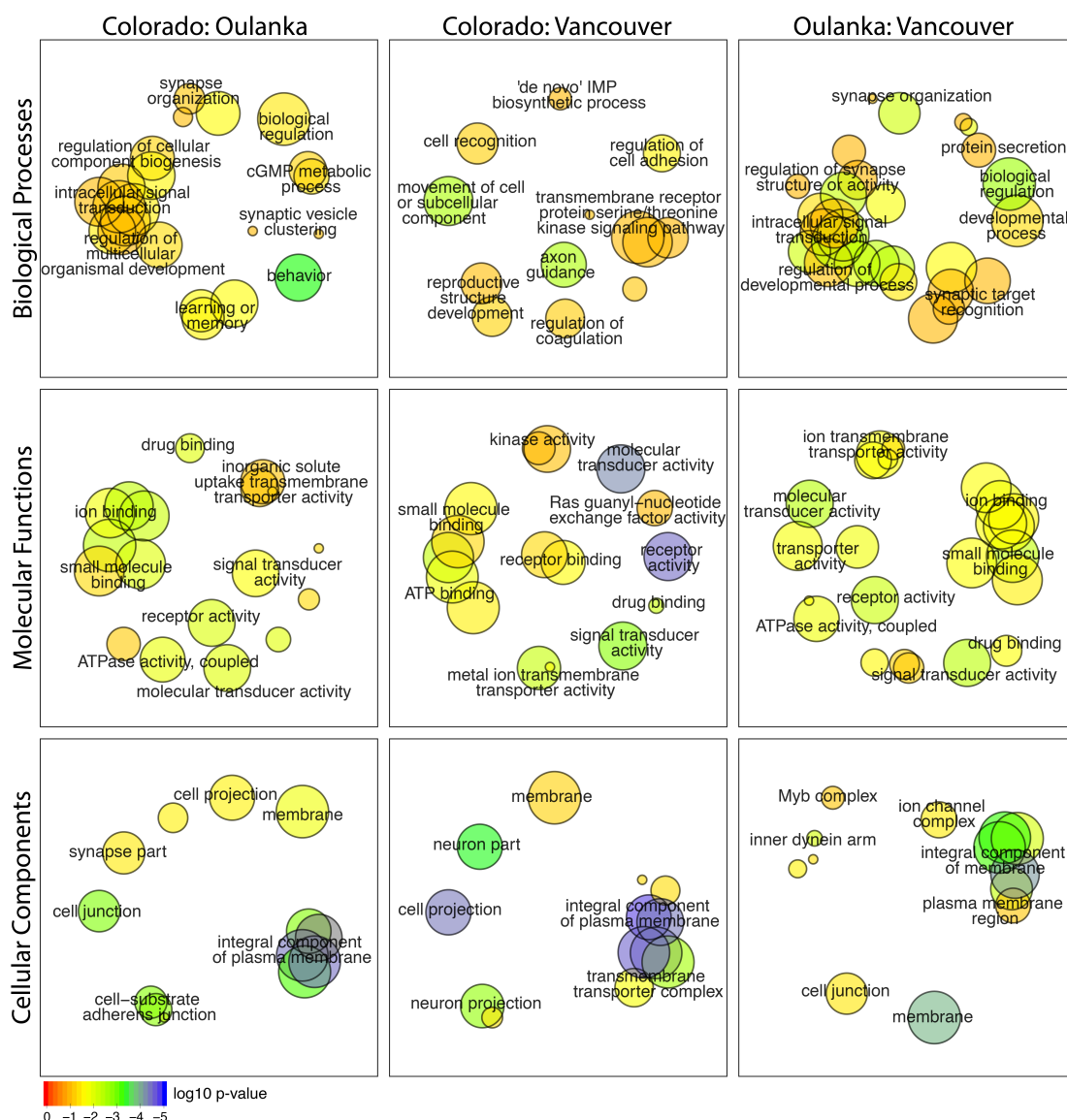
967

968

969

Figure 7 | Semantic clustering of significantly (FDR < 0.1) enriched GO-terms for genes showing significant divergence between populations of *D. montana*. Circle size corresponds to the number of genes annotated to the term in the reference database. Circle colour indicates $\log_{10}$ FDR of the GO term.
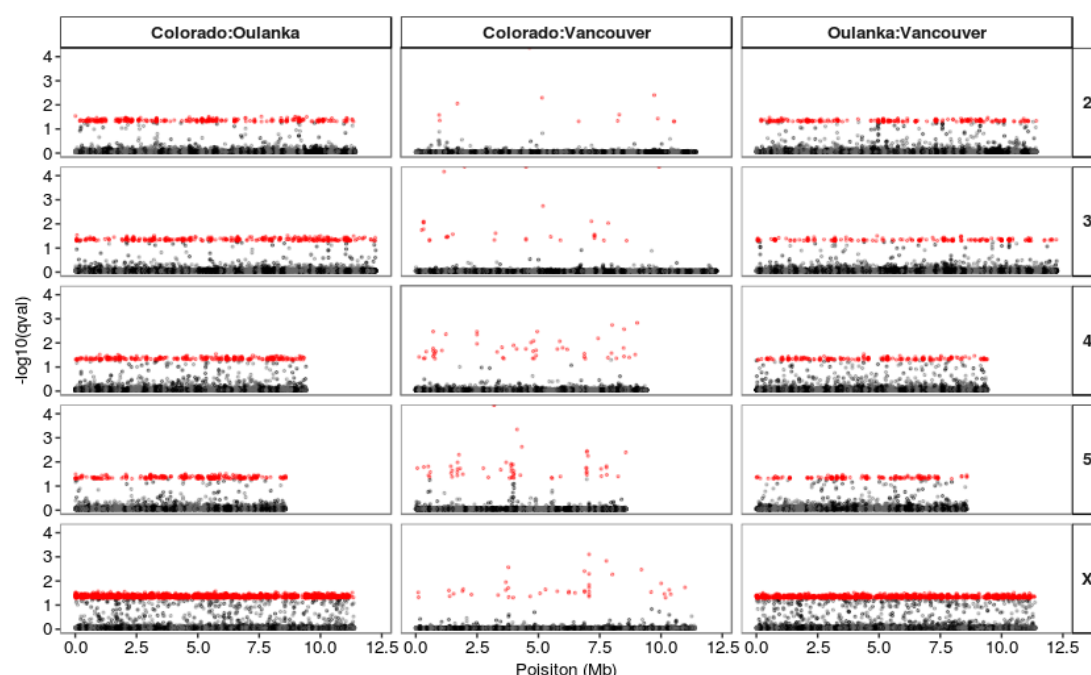
974

Figure 8 | Manhattan plot of q-values from the three pairwise BayeScan analyses for the SNPs on the mapped chromosomes. Red points denote SNPs which passed the 0.05 q-value FDR threshold. Alternating grey and black points denote different scaffolds that have been anchored to the chromosomes. The order of the mapped scaffolds is established but not their orientation.