Genomic signature of shifts in selection in a sub-alpine ant and its physiological adaptations

Cicconardi Francesco[1], Krapf Patrick[2], D'Annessa Ilda[3], Gamisch Alexander[2,4], Wagner Herbert C[2], Nguyen Andrew D[5], Economo Evan P[6], Mikheyev Alexander S[7], Guénard Benoit[8], Grabherr Reingard[9], Arthofer Wolfgang[2], di Marino Daniele[10], Steiner Florian M[2], Schlick-Steiner Birgit C[2]

[1] Department of Zoology, University of Cambridge, Cambridge, UK
[2] Department of Ecology, University of Innsbruck, Innsbruck Austria
[3] Istituto di Chimica del Riconoscimento Molecolare, CNR (ICRM-CNR), Milan, Italy
[4] Department of Biosciences, University of Salzburg, Salzburg, Austria
[5] Department of Entomology and Nematology, University of Florida, USA
[6] Biodiversity & Biocomplexity Unit, Okinawa Institute of Science & Technology, Onna, Japan
[7] Ecology and Evolution Unit, Okinawa Institute of Science & Technology, Onna, Japan
[8] School of Biological Sciences, The University of Hong Kong, Hong Kong, China
[9] Institute of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria
[10] Polytechnic University of Marche, Department of Life and Environmental Sciences, Ancona, Italy
Corresponding author: francicco@gmail.com

Abstract

Understanding how organisms adapt to extreme environments is fundamental and can provide insightful case studies for both evolutionary biology and climate-change biology. Here, we take advantage of the vast diversity of lifestyles in ants to identify genomic signatures of adaptation to extreme habitats such as high altitude. We hypothesised two parallel patterns would occur in a genome adapting to an extreme habitat: *i*) strong positive selection on genes related to adaptation and, *ii*) a relaxation of previous purifying selection. We tested this hypothesis by sequencing the high-elevation specialist *Tetramorium alpestre* and four related species. In support of our hypothesis, we recorded a strong shift of selective forces in *T. alpestre*. We further disentangled candidate molecular adaptations in both gene expression and protein-coding sequence that were identified by our genome wide analyses. In particular, we demonstrate that *T. alpestre* has *i*) a derived level of expression for *stv* and other heat-shock proteins in chill shock tests, and *ii*) enzymatic enhancement of Hex-T1, a rate-limiting regulatory enzyme that controls the entry of glucose into the glycolytic pathway. Together, our analyses highlight the adaptive molecular changes that support colonisation of high altitude environments.

Introduction

Adaptation of organisms to climate drives variation in diversification rates and species richness among clades. Their climatic niche affects a species' distribution in space and time [1], critically affecting both speciation and extinction. Therefore, understanding how organisms manage thermal adaptation is important in terms of both evolutionary biology and climate-change biology, considering the potential world-wide loss of ecological niches [2,3]. In this context, high elevations, characterized by a short growing season and low annual minimum and mean temperatures, with high daily fluctuation temperatures [4], are an important open-air laboratory to study speciation and adaptation to cold habitats; and compared to the great effort in gene-based studies on cold tolerance of model organisms, a lesser effort has been put towards understanding truly cold-tolerant animals at the genomic level [5,6]. This is even more evident when considering that among the approximately 600 sequenced insect genomes available today only four belong to species ecologically restricted to high altitudes or Antarctic habitats [7–10].

As yet, potential patterns of genomic signatures in organisms adapting to more extreme habitats, such as high elevations, have not been identified and, more generally, there is no theory that predicts the rates of genomic change for extreme habitats. Here, we hypothesise two parallel patterns to occur in a genome adapting to an extreme habitat. (1) Strong positive selection on genes related to adaptation, such as genes involved in metabolic pathways; and (2) Relaxation of previous selecting forces, as the conditions of the previous niche are lacking in the new niche. The latter should lead to a reduced number and/or a different set of genes under purifying selection. Specifically, in the case of high-elevation habitats, some heat-

67 shock proteins (HSPs), necessary for coping with extreme heat, should be under
68 relaxation, because heat resistance is not strongly selected for in the average alpine
69 species. A problem with the hypothesis of these two parallel patterns is that there is
70 no direct test available for the functional consequences of the loss or gain of a
71 specific gene. There are often many pleiotropic effects for genes that make
72 interpretations difficult, and HSPs, expressed upon exposure to stress or during
73 development and growth, are likely to fall into this category [11]. Therefore, we
74 expect strong positive selection in genes involved in energetic metabolism, and an
75 increase in the relaxation rates in other genes, without being able to specify which
76 these might be. Identifying such genes will help to set up hypotheses to test in the
77 future. A comparative approach offers the strongest method for testing our
78 hypothesis, by comparing genomes of species closely related to each other, but
79 divergent in terms of adaptation to different environments. However, this approach
80 can be limited by insufficient niche divergence within a group or the number of
81 genomes sequenced. Ants are emerging as a leading system for comparative
82 genomics due to the constantly increasing number of available genomes. This
83 resource offers the opportunity to improve the accuracy of orthology detection [12],
84 to scan for specific mutations, candidate genes, and patterns of acceleration and
85 relaxation in the ant genomes associated with adaptation to cold habitats.

86 Ants are key species in the Earth's terrestrial ecosystems. They comprise
87 more than 15,442 described species and display an impressive diversity of lifestyles
88 [13–15]. Ants are especially notable among insects for their ecological dominance as
89 predators, scavengers, and indirect herbivores. They compose at least one third of
90 the entire insect biomass [16] and colonise all kinds of habitats, including

91 thermobiologically challenging environments. The formicine ant *Melophorus bagoti*,
92 for instance, is active during the hottest periods of the summer day, when air
93 temperatures at ant height exceed 50 ˚C [17]. On the other hand, for example, the
94 myrmicine ant *Tetramorium alpestre* inhabits the montane and subalpine belt of the
95 Central and South European mountain systems, with the Alps as its main distribution
96 area [18,19]. This species lives mainly between 1300 and 2300 m above sea level
97 (a.s.l.), forages below the ground, and nests are established in cool grassland under
98 stones, in moss, rootage, and dead wood, especially subalpine and alpine grass mats
99 [15]. Since *T. alpestre* belongs to a species complex in which no relation between the
100 ecological niche and worker morphology has been detected, searching for adaptive
101 phenotypical differences outside of morphology, for example physiology, has already
102 been suggested [18]. *Tetramorium alpestre* is also relevant because, among the 79
103 Palearctic species of the worldwide distributed genus *Tetramorium* (Bolton 2016), it
104 is one of the only two species to form colonies with multiple queens (polygynous
105 colonies), showing intermediate states of aggression and a transition from
106 multicoloniality to supercoloniality [15,20]. *Tetramorium alpestre* is thus a system
107 well suited for studying both genomic adaptation to high elevation and socio-
108 behavioural evolution.

109 Here, we tested our hypothesis of both increased diversifying and reduced
110 purifying (relaxing) selection resulting from adaptation to an extreme niche using *T.*
111 *alpestre*. We newly sequenced its genome and those of four related *Tetramorium*
112 species with diverging ecological niches: *T. immigrans*, a species of the *T. caespitum*
113 complex to which *T. alpestre* also belongs to, in sympatry with *T. alpestre*, but
114 distinct in its altitudinal and ecological habitat; *T. parvispinum*, restricted to

115 mountain forest habitats of the Austral-Asian and Indo-Malayan subregions [21]; and
116 *T. bicarinatum* and *T. simillimum*, invasive generalist species which occur in warmer
117 habitats [22,23]. Our goals were to define the ecological niche of these ant species
118 based on environmental data, and subsequently to perform gene family
119 expansion/contraction and protein-coding scans for signatures of diversifying and
120 relaxing selection. We also annotated the five HSP subfamilies for 19 ant species to
121 test for possible shifts in selection acting on these gene families. Among the
122 signatures of diversifying selection detected, we performed two experiments in *T.*
123 *alpestre* and its relative, *T. immigrans*: we assessed (i) chill-shock triggered gene-
124 expression patterns of *starvin* (*stv*), a modulator the activity of Hsp70 chaperone
125 machinery during recovery from cold stress, and certain HSPs involved in recovering
126 from extreme cold, and (ii) the temperature dependence of the enzyme activity of
127 Hexokinase type 1 (Hex-T1), a rate-limiting and regulatory enzyme that controls the
128 entry of glucose into the glycolytic pathway, one of the most conserved and
129 essential hexokinase isoenzymes [24]. The genomic data presented here contribute
130 to a foundation for studying insect genome evolution, particularly also in the light of
131 climate change.

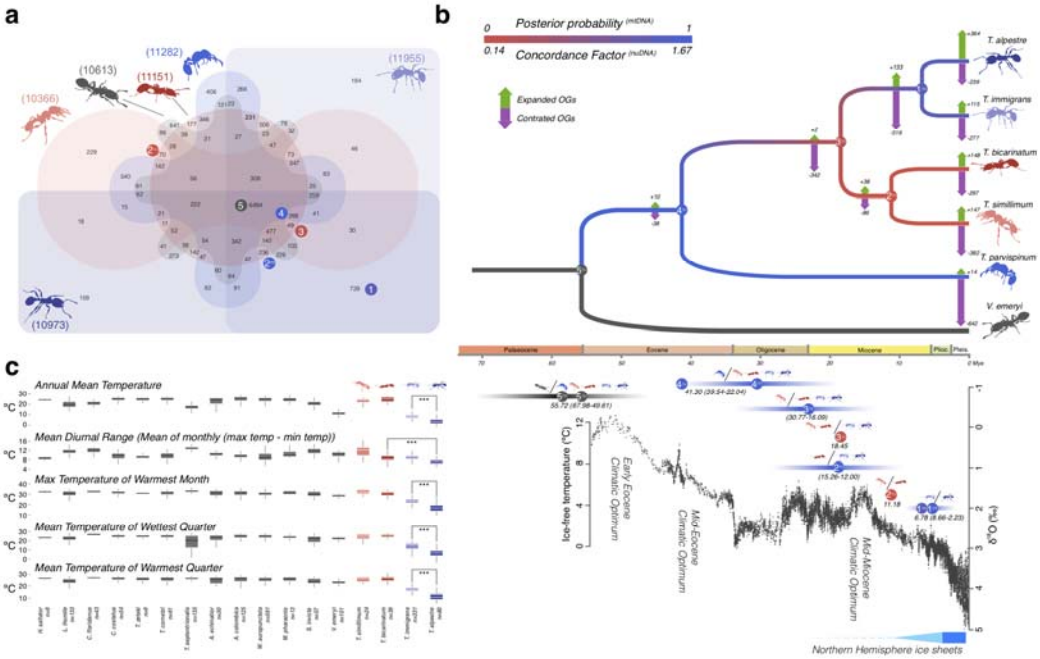


Figure 1. *a*) Venn diagram displaying overlap in orthologous genes in the five *Tetramorium* spp. + *V. emeryi* (Crematogastrini). Coloured circles with number represent phylogenetic splits in the nuDNA (N^{nu}) and mitochondrial phylogenies in *b*; *b*) The dated Crematogastrini nuclear phylogeny (above). Branch colour are based on the concordance factors of the two datasets (nuDNA and mtDNA). On each branch, the numbers of expanded (green arrows) or contracted (purple arrows) genes is shown, inferred from observed OGs sizes at terminal branches. In the bottom, the global $\delta^{18}O$ (‰) derived from analyses of two common and long-lived benthic taxa are given, *Cibicidoides* and *Nuttallides*, which reflect the global deep-sea oxygen and carbon isotope and thus the temperature (from Zachos et al., 2001). Bars represent the 95% confidence interval for the BI analysis, circles with numbers the phylogenetic splits in the nuDNA (N^{nu}) and mtDNA (N^{mt}) phylogenies. *c*) Boxplots of the five bioclimatic variables significantly different for *T. alpestre* and the other ant species.

Results

Climatic niches

To define the environment of *T. alpestre* and the other four *Tetramorium* species, a dataset contains 1835 localities representing 17 species was assembled, using georeferenced occurrence data compiled from the Global Ant Biodiversity Informatics database (see method section), with the number of localities per species ranging from 9 to 591 (Table S1). Of the 19 bioclimatic variables, four (bio1: Annual Mean Temperature, bio5: Max Temperature of Warmest Month, bio8: Mean Temperature of Wettest Quarter, bio10: Mean Temperature of Warmest Quarter)

153 differentiated between *T. alpestre* and other ants (adjusted *P*-values < 0.001) (Figure
154 1a). In detail, *T. immigrans* and *T. alpestre* occupy habitats that are colder in the
155 growing season than those of the remaining species, with *T. alpestre* revealing even
156 colder habitats compared with *T. immigrans*.
157

158 **Sequencing, assembly, and annotation of the *Tetramorium* genomes**

159 The combination of overlapping paired-end libraries with very high coverage,
160 standard short paired-end libraries, and mate pair libraries resulted in an overall
161 complete assembly of the five *de novo Tetramorium* genomes, at the level of both
162 contiguity and scaffolding. In particular, the *T. alpestre* genome contiguity (contig
163 N50) is the 3rd best among all ant genomes available (Table S2). The analysis of the *k*-
164 mer frequency distribution generated was unimodal (Figure S1; Supplementary
165 Material online), and an analysis of the distribution of GC content across the five *de*
166 *novo Tetramorium* genomes revealed similar distributions with a mean of 38%, that
167 is, slightly less than that found in *V. emeryi* (42%) (Figure S2). No significant bias was
168 recovered considering GC content and coverage (Figure S3). All five *de novo* draft
169 assemblies were ~241 Mb in length, comparable in size with the average of all ant
170 genomes available (~278 Mb) (Table S2). The total *T. alpestre* genome size was
171 245.72 Mb, including 18 Mb of gaps and unknown characters (N/X), slightly smaller
172 than the estimated genome size by flow-cytometry of 291.84 Mb +/- 1.76 Mb (n=12).
173 Repetitive elements made up similar proportions of each sequenced genome (*T.*
174 *alpestre* = 18%, *T. immigrans* = 20%, *T. bicarinatum* = 21%, *T. simillimum* = 20%,
175 *T. parvispinum* = 23%; Table S2), indicating a strong correlation between genome
176 size and total interspersed repeat content (Pearson ρ = 0.976, *P*-value < 0.005; Figure

177 S5). For the gene annotations in *T. alpestre*, a combinatorial approach of
178 unsupervised RNA-seq-based, homology-based, *ab initio*, and finally *de novo*
179 methods were used. By doing so, we approached the number of predicted protein-
180 coding genes in the phylogenetically closest annotated species (15,085 genes
181 predicted vs. 14,872 in the *V. emeryi* genome, Table S2). This improved the
182 homology-based annotation of the remaining four *Tetramorium* genome assemblies
183 that did not differ much in terms of estimated gene content, ranging from 16k to 15k
184 annotated loci (Table S2). BUSCO analyses estimated overall good representation,
185 with recovered genes in between the 97.6% and 99.9% (Table S2; Figure S6), and all
186 statistics on mRNA, coding regions (CDS), exon, and intron length gave highly similar
187 and overlapping distributions among all *Tetramorium* species and *D. melanogaster*
188 as a reference (Figure S7). Also, the transcript completeness gave good results with a
189 comparable distribution of percentage of alignment across species (Figure S8).
190 The orthology search analysis produced a total of 7195 OGs between *D.*
191 *melanogaster* and Hymenoptera; 3261 of these represented scOGs present in all 22
192 species analysed. Restricting the orthology analysis to *Tetramorium* spp. and *V.*
193 *emeryi* resulted in 6494 OGs; among *Tetramorium*, the highest fraction was
194 recovered between *T. alpestre* and *T. immigrans* with 729 OGs, followed by *T.*
195 *simillimum* and *T. parvispinum* (540 OGs) (Figure 1a). For all *Tetramorium* species.,
196 the fraction of orthologous genes with at least a match with one of the three
197 outgroups (*A. mellifera*, *N. vitripennis*, *D. melanogaster*) was on average 7286, with
198 the highest in *T. alpestre* (7879) and the lowest in *T. parvispinum* (6977), in line with
199 the average value found in ants (7831).
200

201 **Assembly and annotation of ant mitogenomes and their gene translocations**

202 We assembled complete mtDNA genome sequences from 24 ant species. Because of
203 the differences among SRA datasets, we used different numbers of reads to
204 assemble the complete sequences. The new mtDNA genomes were slightly longer
205 (on average, 17k vs. 16k; Table S2), probably due to the reconstruction of the D-loop
206 and ribosomal genes, which are sometimes harder to assemble. We slightly mis-
207 assembled a coding gene (*nad2*) in the *T. bicarinatum* genome and a few other tRNA
208 loci in other genomes (Table S2). The average GC-content for coding genes was 23.9
209 ± 3.0% and varied slightly among taxa with the lowest in *Formica fusca* (19.0%) and
210 the highest in *Leptomyrmex pallens* (32.8%) (Table S2). While almost all species had
211 the typical gene order and orientation, some rearrangements were observed. All
212 myrmicine ants were characterized by a translocation of *trnV* after *rrnS*; *Solenopsis*
213 spp. had a swapped position of *rrnS* and *trnN*; *Monomorium pharaonis* had at least
214 three rearrangements: the inversion of *nad6*, *cob*, and *trnS2*, the translocation of
215 *trnE* upstream to *trnA*, and the translocation of *trnR* upstream to *trnE*; *Tetramorium*
216 spp. underwent a swap of *trnR* and *trnN* (Figure S9).

217

218 **Phylogenetic analyses of *Tetramorium* spp., their diversification dating, and the**
219 **Miocene-Pliocene origin of the lineage leading to *T. alpestre***

220 The estimated divergence ages from both datasets (mtDNA, nuDNA) were congruent
221 for most nodes, except for recent nodes within the Attini (Figure 1b; Figure S10-13).
222 In this case, the penalized likelihood approach estimates were always more recent
223 than the estimations made with BI using BEAST. The dating of basal nodes like the
224 MRCAs (e.g., Crematogastrini, Myrmicinae, and Formicinae) were congruent with a

225 previous analysis based on more taxa but fewer genes [26]. The two methods also
226 converged regarding the date of the most recent common ancestor (MRCA) between
227 *T. alpestre* and *T. immigrans*. In that specific case, PL dated that split at 6.8 million
228 years ago (mya) within the confidence interval given by BI (Median: 5.15 mya; 95%
229 HPD: 2.66, 8.23).

230 The phylogenetic relationships were generally highly congruent and with
231 predominantly strong support (Figure 1b and Figures S10-13). In the ML phylogeny
232 of scOGs (nuDNA), all nodes were supported with a bootstrap value of 1. Across the
233 whole phylogeny, only a few nodes had weak signals, resulting in very few
234 topological disagreements and low coalescent units (CU). Within the clade of
235 Crematogastrini (*Tetramorium* spp. + *V. emeryi*), the relation of *T. bicarinatum* and *T.*
236 *simillimum* was uncertain: nuDNA phylogeny recovered them as sister species with
237 low CU (0.11), while mtDNA phylogeny placed *T. bicarinatum* as sister to the *T.*
238 *alpestre*-*T. immigrans* cluster, with a bootstrap value of 0.77 and a posterior
239 probability of 0.99.

240

241 **Contractions of OGs in the *T. immigrans* and *T. alpestre* lineages and expansion of**
242 **retrotransposon-related genes in *T. alpestre***

243 We examined the evolutionary dynamics of OGs looking at more dynamic ranges
244 corresponding with terminal branches. We observed an increased number of OG
245 deaths in the *T. parvispinum* lineage (– 642) and in the lineage leading to *T. alpestre*
246 and *T. immigrans* (– 516). Ortholog groups showed very little expansion in internal
247 branches of the *Tetramorium* radiation. Expanded OGs were common in terminal
248 branches, except for *T. parvispinum* with very few expanded OGs (+ 14), suggesting a

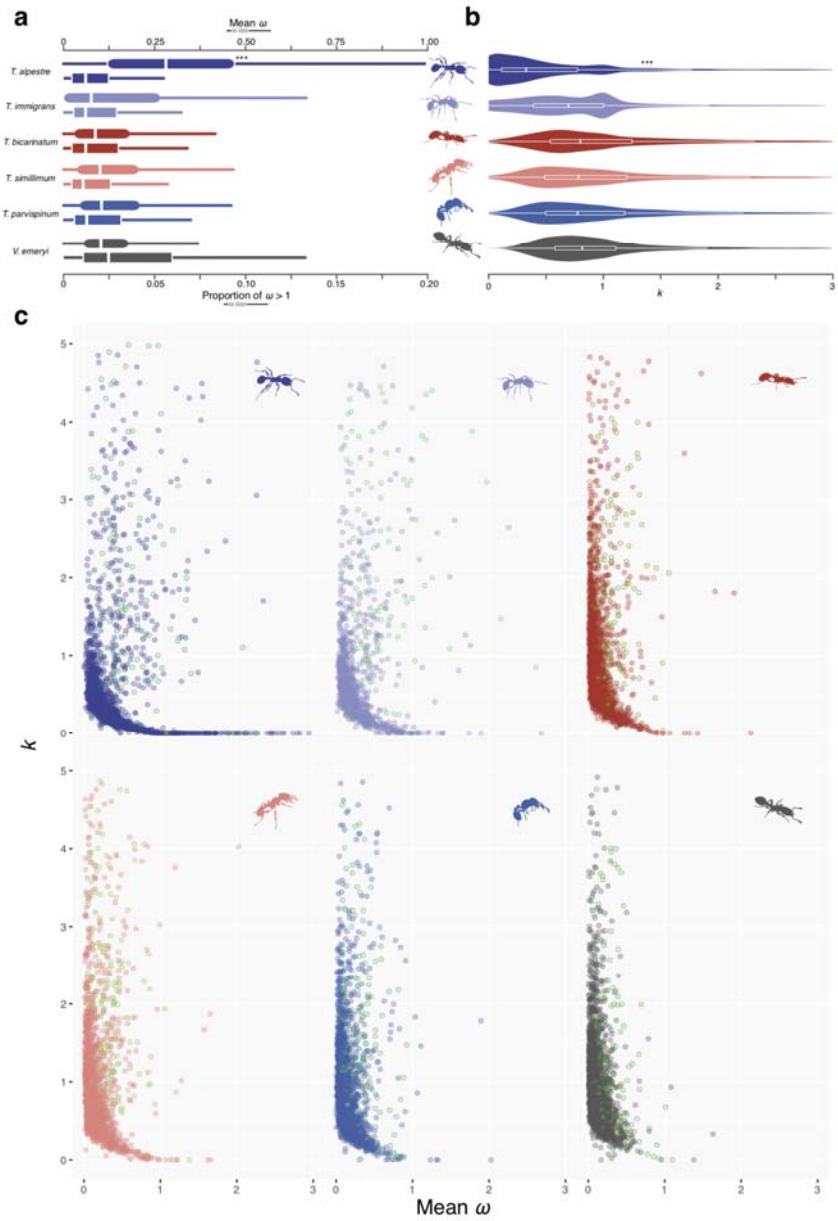
249 possible lack of annotation. *Tetramorium alpestre* showed the highest number of
250 expanded OGs (+ 364) (Figure 1b), with more than 370 of them (~ 40%) belonging to
251 genes with reverse transcriptase (PFAM: rve, rve_1, rve_2) or transposase and
252 retrotransposon activities (PFAM: Retrotrans_gag, Retrotran_gag_2). Among other
253 PFAM, haemolymph juvenile hormone-binding proteins (PFAM: JHBP; 20 genes), zinc
254 fingers (PFAM: zf-CCHC, zf-H2C2, zf-met, 32 genes), and sugar transporters (PFAM:
255 Sugar_tr, 4 genes) were found.
256

257 **Dual signature of evolutionary pressures in single-copy OGs in the *T. alpestre***
258 **genome**

259 We investigated the impact of evolutionary pressures on one-to-one orthology in the
260 five *Tetramorium* spp. and 14 other ant species and three outgroups (*Apis mellifera*,
261 *Nasonia vitripennis*, and *Drosophila melanogaster*) by computing both mean ω
262 (d_N/d_S) for terminal branches and relaxing selection (k) within the Crematogastrini.
263 The measure of the mean branch ω showed an overall high level of purifying
264 selection along the whole ant phylogeny, including for short branches [27,28], with
265 median values of 0.10. However, this distribution was significantly shifted in *T.*
266 *alpestre*, which showed a median value of 0.38 (Wilcoxon rank-sum test '*greater*', *P*-
267 value < $2.2e^{-16}$), whilst keeping the proportion of $\omega > 1$ in each branch unchanged
268 (Figure 2a, S13). The evaluation of k in the five *Tetramorium* spp. and *V. emeryi* had a
269 mean of median values of 0.99, that is, twice the value in *T. alpestre* (median k =
270 0.47; Wilcoxon rank-sum test '*less*', *P*-value < $2.2e^{-16}$) (Figure 2b). We investigated
271 the nature of the increased ω (episodic diversifying selection) and decreased k
272 (relaxed selection) in *T. alpestre* by combining the two measures. Two evolutionary

273 trajectories appeared to act simultaneously (two tails of the distribution, Figure 2c):

274 One trajectory seems to have led to episodic diversifying selection,



275

276 ***Figure2.*** *a*) Boxplots showing the distribution and median (vertical line) of mean ω (d_N/d_S) rates (square shapes)
277 in terminal branches of *Tetramorium* spp. and *V. emeryi* (Crematogastrini) in scOGs, and the proportion of genes
278 for which ω is higher than one. *b*) Boxplots overlapped by violin plots showing the distribution of *k* in
279 Crematogastrini. The distribution in *T. alpestre* is bimodal as in *T. immigrans*, but with a much more skewed
280 distribution towards zero values. In both sections, asterisks indicate the degree of significance between *T.*
281 *alpestre* and the other species (Wilcoxon rank-sum tests). *c*) Scatter plots for each Crematogastrini species
282 included, showing for each scOG branch both values of ω and *k*. The green dots are scOGs with uncorrected *P*-
283 values associated with diversifying selection less than 0.05. The distribution in *T. alpestre* shows the longest tail

corresponding to very low values of *k* and high values of *ω*. It is also notable a much wider and scattered distribution of scOGs with high *k* and *ω* compared with all other species.

promoting the fixation of nonsynonymous mutations with presumably advantageous fitness effects for genes with increased values of *ω* and *k*, also present in the other species. The other trajectory showed a relaxation of the overall purifying selection in genes with increased values of *ω* and decreased values of *k*, only mildly present in *T. immigrans*.

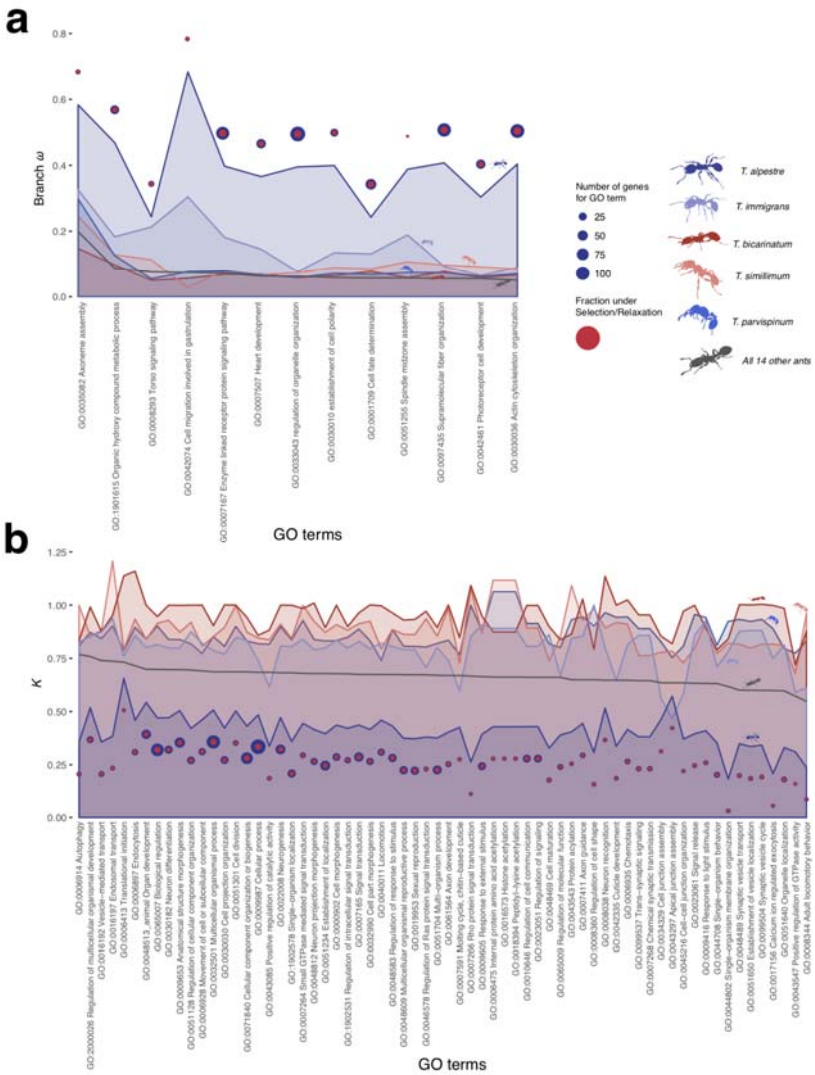


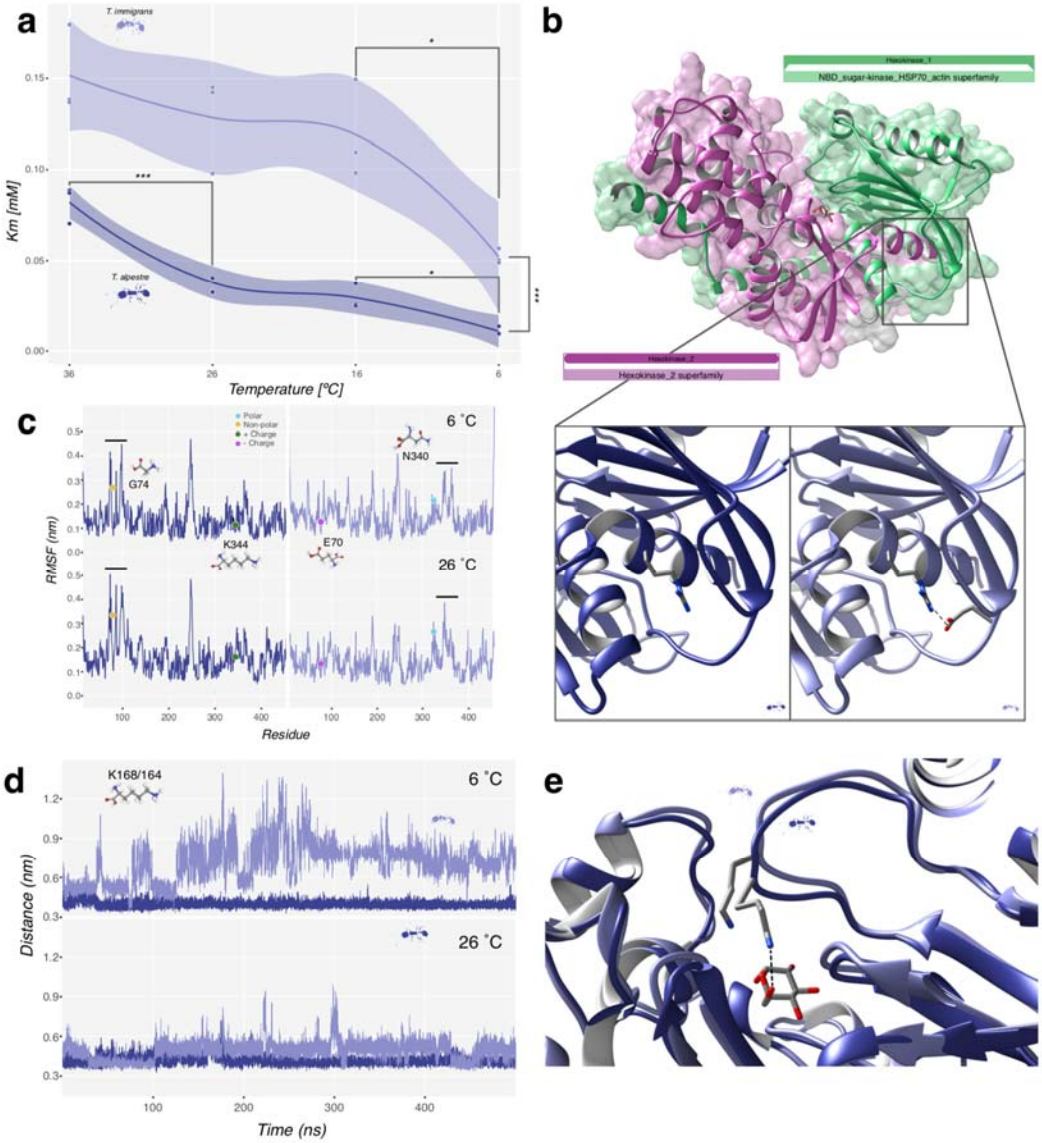
Figure 3. For each enriched Gene Ontology (GO) term category *a*) the median values of *ω*; and *b*) *k* for genes enriching the specific category. *Tetramorium* spp. are colour coded, the other 14 species, merged to obtain a background signal, are in gray. Blue circles are scOGs tested for each GO term, red circles are scOGs returned as significant. Circle sizes are proportional to the number of genes. To be noted are the overall extremely higher

values of ω and lower values of k in *T. alpestre* branches compared to the other species. Seldomly also *T. immigrans* shows few GO terms with high ω and low k values, in particular the Apical junction assembly (GO:0043297) shows and inverted trend between *T. alpestre* and *T. immigrans*.

After a stringent correction for multiple testing, we found 175 scOGs with a putative signature of diversifying selection (adjusted *P*-values < 0.005; Table S5), with 50 genes enriching 13 biological process terms (*P*-values < 0.005; Figure 3a, Table S6), nine related to cell development (e.g., cell migration involved in gastrulation, photoreceptor cell development, cell fate determination) and organization (e.g., regulation of organelle organization), two related to signalling pathways (enzyme linked receptor protein signalling pathway and torso signalling pathway), one to heart development, and one to organic hydroxy-compound metabolic processes. The most representative genes were *Pten* and *Gbeta13F*, involved in six biological processes, *DCTN1-p150* and *dsh*, involved in five of them. All key genes are involved in the proliferation and division of cells, especially neurological stem cells (*Gbeta13F*), respiratory system development (*Pten* and *dsh*), and the serine/threonine-protein kinase *polo*, present in four enriched Gene Ontology (GO) terms. In the catalytic domain of this protein (255 aa), we found five amino acid substitutions compared with *T. immigrans*, three of them corresponding to active sites ($TalpPolo_{G29V}$, $TalpPolo_{G30R}$, $TalpPolo_{S104R}$). We also tested for possible enrichment in KEGG pathways and found three significant pathways (*P*-values < 0.02; Table S7), all involved in the metabolism of sugars, specifically: glycolysis, pentose phosphate pathway and galactose metabolism. Within these pathways, five enzymes were found under diversifying selection: a glucose-6-phosphate (*CG9008*), the hexokinase type 1 (*Hex-T1*), a phosphofructokinase (*Pfk*), a transketolase (*CG8036*), and a phosphoglucosemutase (*pgm*).

322 132 scOGs had a signature of putative relaxation (*P*-values < 0.005; Table S8).
323 While the absolute number of scOGs under relaxation was slightly lower than the
324 number of genes under diversifying selection, the gene-set enrichment tests
325 revealed many more biological processes significantly enriched: 70 terms (*P*-values <
326 0.005; Figure 3b, Table S9) enriched by 118 genes. Fifteen terms were related to the
327 regulation of cellular development, signal transduction, and cell communication
328 (e.g., regulation of response to stimuli, regulation of cell communication, regulation
329 of Ras signal transduction), 12 terms related to neurogenesis, axon development,
330 and synaptic transmission (e.g., neuron differentiation, trans-synaptic signalling,
331 axon guidance), seven terms were related to cell morphogenesis and junction
332 organization, and many other related to response to stimuli (e.g., chemotaxis,
333 response to light) or related to adult development and morphogenesis (e.g.,
334 anatomical structure morphogenesis, molting cycle, chitin-based cuticle). The
335 enrichment of KEGG pathways revealed three pathways: mRNA surveillance
336 pathway, spliceosome, and protein processing in endoplasmic reticulum (*P*-values <
337 0.05; Table S10).
338
339 **The Hex-T1 activity, its model structure, and molecular dynamics simulations in *T.***
340 ***alpestre* and *T. immigrans***
341 As we describe in the previous section, the enrichment for KEGG pathways led to the
342 identification of three metabolic pathways, all involved in the metabolism of sugars,
343 with five genes under selection. In order to experimental validate these *in silico*
344 results, we cloned and expressed the Hex-T1 forms for both *T. alpestre* and *T.*
345 *immigrans* into *Drosophila melanogaster* cells lines, in order to compare their

346 activities. The purified enzyme of both ants was assayed in at 6, 16, 26, and 36 °C to
347 establish their efficiency under different thermic conditions and identify possible
348 metabolic adaptations to cold-environments. The assays showed that the Km
349 (substrate concentration necessary to reach maximum reaction speed) of Hex-T1 of
350 *T. alpestre* was lower than the enzyme of *T. immigrans* at all temperatures.



351
352 ***Figure 4.*** *a*) Scatter plot of Km score across the temperature gradient in the two species, *T. immigrans* on top and
353 *T. alpestre* on the bottom. Lines represent average values with confidence intervals. Significances: *** < 0.001;
354 ** < 0.001; * < 0.05. *b*) Overview of the homology-modeled structure of Hex-T1 in *T. alpestre* showing the two
355 Hexokinase domains, PFAM00349 (green) and PFAM03727 (pink). The lower panels show the location of TalpHex-
356 $T1_{G74}$ (left) and TimmHex-$T1_{E70}$ (right) as well as the arginine residue that forms a salt bridge with E70 (dashed
357 line). *c*) Per-Residue Root Mean Square Fluctuation (RMSF) calculated along 500 ns trajectories carried out at 6 °C
358 (upper panel) and 26 °C (lower panel) for TalpHex-T1 (left panel, blue line) and TimmHex-T1 (right panel, light

blue line). The location of the residues found mutated between the two sequences are indicated. *d*) Time evolution of the atomic distance between the centre of mass of glucose and the lateral chain nitrogen atom of TalpHex-T1$_{K168}$ (blue line) and TimmHex-T1$_{K164}$ (light blue line) in the binding pocket calculated along the 500 ns trajectories carried out at 6 °C (upper panel) and 26 °C (lower panel). *e*) Representative configurations extracted from the simulations at 26 °C showing the different orientation of TalpHex-T1$_{K168}$ and TimmHex-T1$_{K164}$, with the first one being the only one able to bind the glucose (dashed line).

(Wilcoxon rank-sum test 'less', *P*-value < 0.00004, Figure 4a), with values ranging from 0.08 mM to 0.01 mM in *T. alpestre* versus 0.15 mM to 0.05 mM in *T. immigrans*, from the highest to the lowest temperature. These results demonstrate that Hex-T1 of *T. alpestre* was always more efficient than *T. immigrans*, requiring less substrate to reach maximum activity. Although the number of replicates per condition was small, we performed statistical tests to attempt obtaining insight in the enzymatic activity differences within species at different temperatures and in their variance. Interestingly, it seems that for both species there is an improvement in the Km values shifting towards the lowest temperatures (6 °C *vs* 16 °C; two-way ANOVA followed by a Tukey's multiple comparison; adjusted *P*-values < 0.05), and a possible deterioration at higher temperatures in *T. alpestre*, from 0.04±0.01 mM (26 °C) to 0.08±0.01 mM (36 °C) (two-way ANOVA followed by a Tukey's multiple comparison; adjusted *P*-value < 0.0002). An F test (VAR.TEST() implemented in R) to compare the variances of the enzymatic activities (Km values) between the two species – by pooling the replicates at 16 and 26 °C – suggested that there is a possible significant difference in their variance (*P*-value = 0.015), with the confidence interval of Hex-T1 activity in *T. immigrans* twice wider as in *T. alpestre*.

The Hex-T1 enzyme is made of two structurally similar domains, Hexokinase 1 and 2 (PFAM00349 and PFAM03727), and an N-terminus that is highly variable both in terms of length and amino acidic composition (Figure 4b). The two domains are

386 relatively well conserved across all ants, other Hymenoptera and *D. melanogaster*.
387 Excluding the hypervariable N-terminus, and considering the two functional
388 domains, the amino acid sequences of *T. alpestre* and *T. immigrans* only differ by
389 two amino-acidic substitutions: a glutamic acid (E) replaced by a glycine (G) in
390 position 74 (TalpHex-T1$_{E74G}$) (Figure 4b inserts) and an asparagine (N) replaced by a
391 lysine (K) in position 344 (TalpHex-T1$_{N344K}$). While the second change probably does
392 not entail a significant functionality effect, the *TalpHex-T1$_{E74G}$* instead may be
393 significant. Because of the lack of the long chain and the negative charge of glutamic
394 acid, this substitution could give the protein a higher flexibility and therefore an
395 overall higher kinetics. We tested this hypothesis by simulating the protein dynamics
396 in the presence of the glucose substrate at different temperatures, i.e. 6 and 26 °C,
397 in the two species, with the intent of mirroring the condition in the in vitro assay. In
398 the simulations we can see a dual effect. For TalpHex-T1$_{N344K}$, located at the C-
399 terminal of a α-helix forming a helix-turn-helix motif (VSETE**K**DPKG), resulted in a
400 lower degree of Root Mean Square Fluctuation (RMSF) revealed a lower degree of
401 fluctuation in TalpHex-T1 (Figure 4c). This is possibly due to the alternation of
402 positive and negative charges creates a salt-bridges network that stabilizes the
403 whole motif (data not shown). In particular TalpHex-T1$_{K344}$ is involved in a stable
404 interaction with TalpHex-T1$_{E341}$ (data not shown), while TimmHex-T1$_{N340}$ is not
405 available for the formation of salt bridges and mainly establishes hydrogen bonds
406 with TimmHex-T1$_{S336}$ or TimmHex-T1$_{E337}$ at 6 or 26 °C, respectively. Although the *T.*
407 *alpestre* salt bridge brings a higher flexibility of the helix-turn-helix motive, it does
408 not interfere with the catalytic site, given that this motif is far away in the C-terminal
409 region of the protein. In contrast, at the other region (TalpHex-T1$_{E74G}$), the mutation-

410 caused amino acid change allows the loop to explore a broader conformational
411 space equipping it with a higher degree of flexibility already at 6 °C (Figure 4c).
412 Nevertheless, by increasing the temperature, TalpHex-T1 was subjected to an even
413 higher degree of fluctuation, once again confirming the intrinsic flexibility led by the
414 glycine. Notably, the increased flexibility is also transferred to residues relatively far
415 from the mutation, in particular in regions near the active site (168-169), causing a
416 different pattern of interaction with the substrate (Figure 4d, e). Indeed, TalpHex-T1
417 established a stronger interaction with glucose at both temperatures with respect to
418 TimmHex-T1 (Figure 4e).
419

420 **The heat-shock proteins in ants and their evolution in *T. alpestre***

421 Members of the five subfamilies of the heat shock proteins (HSPs: Hsp90s, Hsp70s,
422 Hsp60s, Hsp40s, and sHsps) were fully recovered and characterized in 19 ant species.
423 Overall, a strong purifying selection was observed within subfamilies (Figure 5a,
424 Table S11). In the Hsp90 gene family (*trap1*, *gp93*, and *Hsp83*) *Hsp83* was found
425 duplicated within the lineage of *W. auropunctata* (inparalog), as well as the *T.*
426 *parvispinum gp93*; duplication that probably occurred before emergence of the
427 lineage (outparalog) (Figure S14). Members of the Hsp70 family were recovered for
428 all species with 11 ortholog groups, with few duplications in *Hsc70-3* and *Hsc70-5*
429 (two inparalogs and two outparalogs, respectively). We confirmed the lack of Hsp70s
430 in ants [29] and a multiple duplication of *Hsc70-4* in Hymenoptera. One paralog
431 seemed to be lost in many species while the other two copies were still present in
432 almost all species, with the presence of other outparalogs in various species, mostly
433 in Attini (Figure S15). The chaperonin Cpn60/TCP-1 family (Hsp60) had nine OGs; six

434 inparalogs were found in *CCT7, CCT3* and four outparalogs in *Tcp1a*, *CCT8*. The heat
435 shock protein 40 kDa/Chaperone DnaJ was the largest subfamily, with 37 well-
436 defined OGs. Its turnover of duplication was relatively small, and only six OGs
437 showed some level of duplication. The small HSPs (sHsps) showed the expansion of
438 the lethal (2) essential for life (*l(2)efl*) gene paralog of *D. melanogaster*, with five
439 distinct OGs in all Hymenoptera. *L(2)efl3* seemed to be retained by only in 11
440 species, and with *l(2)efl4*, they are the two most divergent OGs, with ω equal to 0.23
441 and 0.33, respectively. The less divergent were *Hspb1* and *l(2)efl5*, for which ω was
442 0.11 and 0.07, respectively. Overall, the sHsps showed a significantly higher
443 distribution of ω (median = 0.22, *P*-value = 0.013), followed by Hsp40s, which also
444 showed a wide distribution of ω. In contrast, Hsp90s, Hsp70s, and Hsp60s seemed to
445 be very stable and under a very high purifying selection.

446 To understand other possible mechanisms of adaptation of *T. alpestre* to the
447 cold, we computed the mean values of ω and k for all branches leading to
448 *Tetramorium* spp. and *V. emeryi*, integrating 13 ant species, and scanned for
449 diversifying and relaxed selection specifically in *T. alpestre*. The first part of the
450 analysis showed a similar landscape as observed in scOGs. Among the five
451 *Tetramorium* species and *V. emeryi*, five species display a distribution of the mean-
452 branch ω ranging between 0.07 (*T. immigrans*) and 0.13 (*T. parvispinum*); *T. alpestre*
453 instead showed significant 2.7-fold higher value of the median mean-branch ω (ω =
454 0.36, Wilcoxon rank-sum test '*greater*', adjusted *P*-values < 0.004, Figure 5b). The
455 distribution of k among HSPs in the six species showed median values of k all close to
456 the expected (k = 1), between 0.90 in *V. emeryi* and 1.05 in *T. simillimum*, again with
457 the exception of *T. alpestre*, where the median k = 0.63, 1.4-fold lower than the

lowest median *k* (Wilcoxon rank-sum test '*less*', adjusted *P*-values < 0.05, Figure 5c). The scan for diversifying positive and relaxed selection in branches of the HSPs of *T. alpestre*

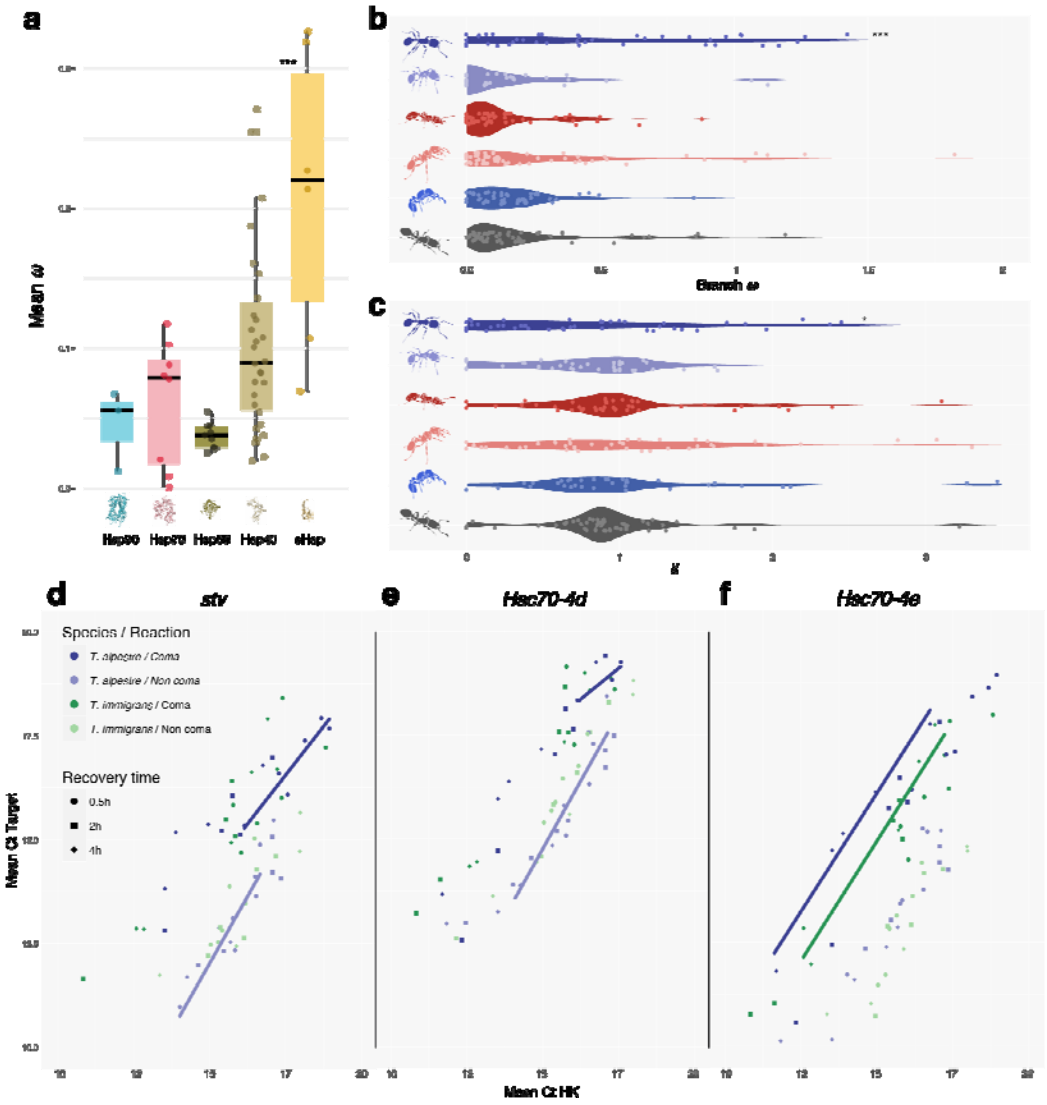


Figure 5. *a*) Boxplots showing the distribution and median (horizontal black line) of mean ω (d_N/d_S) rates for each of the HSP subfamily. *b*) Violin plot showing the distribution of the mean ω values; and *c*) *k* in the terminal branches of the Crematogastrini (from up to bottom: *T. alpestre*, *T. immigrans*, *T. bicarinatum*, *T. simillimum*, *T. parvispinum*, *V. emeryi*). *d-f*) Scatter plot of target gene (*stv*, *Hsc70-4d*, *Hsc70-4e*) vs. housekeeping gene (HK) concentrations; concentrations given as total number of cycles minus cycle threshold (Ct) values, that is, higher values represent higher concentrations. Straight lines are linear regressions for each target gene against HK using the different recovery times separately; based on one-way analysis of covariance followed by correction for multiple comparison, just lines for those treatment / recovery-time combinations are shown for which significances for just a single species arose, that is, *T. alpestre* coma vs. non-coma after 0.5 h (*d*, *e*; no difference in *T. immigrans*) or for which the two species differed, that is, *T. alpestre* coma vs *T. immigrans* coma after 4.0 h (*f*).

474 showed no gene under putative diversifying positive selection, while six loci were
475 identified to be under putative relaxing selection: one Hsp90 (*Trap1*), two Hsp60
476 (*CCT4*, *CCT5*), and three Hsp40s (*Sec63*, *DnajC8*, *DnajC11*) ($k < 0.14$, adjusted *P*-values
477 < 0.005). Using these genes in prediction interaction networks, we found that the
478 two Hsp60s interacted with other 21 genes, giving 17 significant enriched functions
479 (FDR < 0.05), mainly related to protein folding and microtubule, spindle organization;
480 the three Hsp40s interacted with 21 genes, giving two enriched functions (FDR <
481 0.05), related to response to heat and response to temperature stimulus (Figure
482 S19).
483
484 **Chill shock effect, recovery times, and expression patterns of *stv* and HSPs in *T.***
485 ***alpestre* and *T. immigrans***
486 Among the genes with a putative signature of diversifying selection, we also found
487 *starvin* (*stv*), a BAG-family member which seems to be implicated in Hsp70 ATPase
488 activity during cold recovery in *D. melanogaster* [30]. Therefore, we explored if the
489 modulation of its expression may play a role in how *T. alpestre* differently withstands
490 the cold in comparison to the related species *T. immigrans.* We thus performed a
491 pilot test measuring the different expression patterns of *stv*, two paralogs of *D.*
492 *melanogaster Hsc70-4* (*Hsc70-4d* and *Hsc70-4e*), a member of the Hsp90 subfamily
493 (*Hsp83*), and sHsp *l(2)efl4* in the two species.
494 We measured the transcription activity of three recovery times (0.5 h, 2.0 h,
495 and 4.0 h) and compared coma and non-coma individuals, within and between
496 species, and found a significant difference in the expression patterns of *stv*, *Hsc70-*
497 *4d*, *Hsc70-4e*, and *Hsp83* were observed, but not of *l(2)efl4* (one-way ANCOVA

498 followed by Bonferroni-Holm correction). In particular, looking at the different
499 recovery times, *stv* and *Hsc70-4d* responded quickly in *T. alpestre* but not in *T.*
500 *immigrans,* being significantly expressed only in workers of *T. alpestre* after 0.5 h
501 recovery, but not in *T. immigrans* (Figure 5d-e). Comparing the expression levels
502 between species, we found higher mRNA levels for *Hsc70-4e* in *T. alpestre* after 4 h
503 recovery (Figure 5f).
504
505
506 ***Discussion***
507 *Tetramorium alpestre* represents a critical example of evolutionary adaptation to
508 cold environments such as alpine habitats (Figure 1c). By comparing its genome with
509 other 18 ant species, we shed new light on how *T. aspestre* evolved to colonize sub-
510 alpine habitats and proposed it as a model for studies on arthropods adaptation to
511 cold. We found significant concordance between the nuclear and the mitochondrial
512 dated phylogenies, concerning the dating of the speciation of *T. alpestre*, which
513 seems to have diverged from its closest relative, *T. immigrans*, between 2 and 9 mya.
514 This period overlaps with the beginning of a new climatic zonation of the European
515 continent during the Middle and earliest Late Miocene [25]. At the end of this
516 period, a rapid uplift (~5 mya) fundamentally changed the paleogeographic and
517 topographic setting of central and southern Europe and triggered Alpine glaciation
518 [31]. Interestingly, both paleoclimatic and paleogeographic changes co-occurred
519 with the origin of the lineage leading to *T. alpestre* (Figure 1b). In fact, while these
520 glaciations dramatically reshaped global biodiversity patterns, eliminating terrestrial

521 biota from many mid- to high-latitude areas, they may have been the main driver of
522 diversification on highlands by reducing gene flow among populations [32].

523 The present results support our prediction that two parallel patterns will
524 occur in the genome of species adapting to extreme habitats. The genome of *T.*
525 *alpestre* appears to be under the influence of two strong evolutionary trajectories:
526 on the one hand, a diversifying selective force, two-fold higher compared with all
527 other ants, and on the other hand, a relaxation of the overall purifying selection
528 present in other ants, identified by the skewed distribution of *k* towards zero.
529 Specifically, the diversifying selective force affects more than a hundred genes in
530 many key biological processes, such as those related to development, cell migration,
531 and gastrulation. Genes under the category of biological processes are also under
532 selection in the genome of the Antarctic midge *Belgica antarctica* [8]. In fact, early
533 stages are the most fragile ones, having the highest rate of lethal phenotypes during
534 heat shock and UV irradiation, particularly around gastrulation [33]. The same may
535 pertain to genes related to cellular-level organization (e.g., spindle midzone
536 assembly, supramolecular fiber organization, actin cytoskeleton organization), where
537 environmental stress may jeopardise a well working cell division machinery, as SNPs
538 associated to long-term cold-adaptation plasticity were found to be present in genes
539 related to cytoskeletal and membrane structural components in *D. melanogaster*
540 [34]. Among 25 genes involved in these biological processes, *polo*, a polo-like kinase
541 which plays a central role as regulator of cell division and is required for several
542 events of mitosis and cytokinesis [35], shows three amino acid substitutions in
543 positions corresponding to active binding sites, and thus could be a good candidate
544 for further studies.

545 The two evolutionary trajectories were predicted in our hypothesis, and while
546 diversifying selection is somewhat expected in species adapting to a new
547 environmental niche, the magnitude of relaxed selection found is more surprising
548 and never recorded. Although both ω and k distributions in *T. alpestre* are highly
549 skewed, the absolute number of genes under diversifying and relaxed selection are
550 not markedly greater than other species in similar studies [27,28,36]. This cannot be
551 directly ascribed to the type *I* error rate because the same rate should be equally
552 randomly present in all the other short or long branches of the phylogeny.
553 Nevertheless, *T. alpestre*, as part of a species complex [18,19], may have more
554 recently diverged from more closely related species not yet adapted to the alpine
555 habitat. Magnitude of this relaxed selection is evidenced by the 70 enriched
556 biological processes, and can be seen as the consequence of a shift and/or
557 decreased magnitude in the purifying selection. The direct implication of this
558 relaxation is not clear, especially its effect on genes and their biological processes,
559 such as the regulation of cellular development, signal transduction, cell
560 communication, neurogenesis, axon development, and many others. If the
561 physiology of cold adaptation utilises different metabolic and developmental
562 strategies to minimise and optimise energy consumption, a new balance needs to be
563 reached. The two patterns here observed (diversifying/relaxing selection) might
564 represent this new shift. Interestingly, *T. immigrans*, which inhabit intermediate
565 environmental conditions (Figure 1c), shows also an intermediate pattern of k, with
566 an intriguing bimodal distribution (Figure 2b). Therefore, we can speculate that the
567 highly skewed distributions point to ongoing adaptation to a colder climate, and a
568 relaxation of selective forces present in warmer habitats. To our knowledge, this is

569 the first time that this effect is found in an organism, most likely due to the
570 underrepresentation of genomic studies on alpine and cold adapted species, and,
571 importantly, the lack of application of the RELAXED test in a genome-wise manner.

572 Examining the effect of those evolutionary trajectories, energetic metabolism
573 seems to be significantly involved in the sub-alpine ant's adaptation. Several clues
574 are pointing in that direction, such as the more specialized diet underlined by the
575 presence of commensal underground aphids in *T. alpestre* nests but not in those of
576 *T. immigrans* (unpublished data), the expansion of OGs related to sugar transporters,
577 and five enzymes under diversifying selection related to metabolism of sugars. We
578 deeply explored and validated one of these enzymes, Hex-T1, a key regulator and
579 rate-limiting enzyme for energy (sugars) metabolism and reactive oxygen species
580 (ROS) activity in insects [37], by contrasting its enzymatic activity between *T. alpestre*
581 and *T. immigrans.* We showed that the mutated form in *T. alpestre* is indeed more
582 efficient, with less substrate needed to reach the highest enzymatic activity, with the
583 protein modeling simulations demonstrating that the single amino acid mutation
584 enhances the catalytic efficiency of the enzyme due to the formation of a more
585 functional network of hydrogen bonds with its substrate. In fact, the path of the
586 hydrogen bonds found in TalpHex-T1 resembled the one depicted by the X-ray
587 solved structure of Hex-T1 from [38], used as a template to obtain the three-
588 dimensional coordinates of both TimmHex-T1 and TalpHex-T1 (see Methods).
589 Mulichak and colleagues (2000) identified their model residues Lys621 and Asp657
590 as crucial for the catalytic activity of the protein; the corresponding residues
591 TalpHex-T1$_{K168}$ and TalpHex-T1$_{D204}$ are indeed interacting, together with other
592 residues, with glucose in TalpHex-T1 at both 6 and 26 °C. On the other hand, while

593 the interaction with TalpHex-T1$_{D204}$ is conserved, the one with the lysine is lost in
594 TimmHex-T1 and a looser protein-sugar network is formed, explaining the higher
595 catalytic efficiency of TalpHex-T1 versus that of TimmHex-T1. We can indeed infer
596 that the higher flexibility conferred by TalpHex-T1$_{G74}$ increases the shift of the
597 residue TalpHex-T1$_{K168}$, belonging to the catalytic site (Figure 4d, e), allowing it to
598 move closer and establish a stronger interaction with glucose, possibly contributing
599 to lower the Km of the catalytic reaction. In eco-evolutionary terms all these findings
600 hint at a more intimate relation between *T. alpestre* and the aphids. This relationship
601 between ants and aphids is one of the most studied examples of mutualistic
602 relationships in the animal kingdom. Aphids produce honeydew, and in return ants
603 offer protection from predators, parasitoids, fungal infection and adverse conditions.
604 Honeydew contains monosaccharides (glucose and fructose), disaccharides (maltose,
605 sucrose), and Melezitose, a trisaccharide, that is one of the main sugar synthesized
606 by aphids, and which has been found to attract ants and maintain the ant-aphid
607 mutualism [39]. Small amounts of amino-acids, proteins, and lipids are also present
608 [40]. Usually, this interaction is above ground level, on plant leaves and branches,
609 but there are cases in which ants are specialized to farm subterranean aphids, just
610 like in *Lasius flavus* [41]. *Tetramorium alpestre* may have evolved the same behavior
611 through a coevolution of physiological factors such as palatability, fluid intake rate
612 and digestibility of sugar molecules, improving the ants' survival by enhancing
613 energy intake, a limiting factor in high-elevation ecosystems.

614 The complete annotation and characterization of the five HSP subfamilies in
615 19 ant species show good overall conservation, and strong purifying selection was
616 observed in most subfamilies. The exception are sHsps, which seem to be the most

617 diversified HSP subfamily, with the highest and the widest range of ω. It is also the
618 only subfamily that took a different evolutionary path compared with Diptera,
619 having very few orthologs in common with them. Given that sHsps are virtually
620 ubiquitous molecular chaperones that can prevent the irreversible aggregation of
621 denaturing proteins [42], they may have played a central role in the adaptation of
622 ants to strongly different climates and environments. Although we observe a higher
623 distribution of ω in *T. alpestre*, no HSPs show a significant sign of diversifying
624 selection. Rather, six loci show relaxed purifying selection (Hsp90: *Trap1*;
625 Hsp60:*CCT4*, *CCT5*; Hsp40: *Sec63*, *DnajC8*, *DnajC11*). The impact of the TNF receptor-
626 associated protein (Trap1) on cellular bioenergetics of *T. alpestre* may contribute to
627 its adaptation to cold environments by differently modulating cellular metabolism.
628 Given that it functions as a negative regulator of mitochondrial respiration, able to
629 modulate the balance between oxidative phosphorylation and aerobic glycolysis,
630 possibly another form of metabolic adaptation. It has been shown that reduced or
631 absent *Trap1* expression deregulated mitochondrial respiration leading to a high
632 energy state. This state is characterized by accumulation of ATP, tricarboxylic acid
633 cycle intermediates and reactive oxygen species, as well as an increase in fatty acid
634 oxidation [43]. *Trap1* overexpression is associated with increased expression of
635 genes associated with cell proliferation [44]. Since the two Hsp60s are associated to
636 protein folding and cytoplasmic microtubule organization, biological processes also
637 enriched by genes under diversifying selection, this suggests profound changes in
638 the cellular physiology of *T. alpestre*. It is also plausible that adapting to a cold
639 environment will depress selection on genes directly related to severe heat shock. As
640 we have shown, *T. alpestre*'s habitat is characterised by temperatures that rarely

641 exceed 20 °C (median = 16.5 °C; 99% confidence interval = 17.8 °C), although rocky
642 habitats may produce warmer microhabitats. This could mean that this species
643 experienced relaxed selection for HSPs associated with the denaturation to protein
644 folding by heat shock, as suggested by the relaxation of purifying selection on *Cct5*
645 and the three Hsp40s associated with responses to heat and temperature stimulus.

646 Although we provide a range of evidence that adaptation to cold
647 environments can be related to changes in protein-coding sequences, it has long
648 been postulated that phenotypic divergence between closely related species, such as
649 *T. alpestre* and *T. immigrans*, is primarily driven by quantitative and spatiotemporal
650 changes in gene expression, mediated by alterations in regulatory elements (e.g.:
651 Danko et al., 2018; Prescott et al., 2015). We tested this hypothesis by looking at the
652 expression patterns of *stv* and four other HSPs (*Hsc70-4d*, *Hsc70-4e*, *Hsp83*, *l(2)efl4*),
653 between *T. alpestre* and *T. immigrans*. *Stv* is a member of Bcl-2-associated
654 athanogene (BAG)-family proteins, which interact with Hsc70s and Hsp70s, and can
655 modulate, either positively or negatively, the functions of these chaperones [47]. It is
656 implicated in the recovery from chill shock in *D. melanogaster* during cold recovery
657 [30,48]. The concerted synthesis of *D. melanogaster stv* and Hsp70s suggests
658 cooperation to offset cold injury, possibly by preserving the folding/degradation and
659 by regulating apoptosis [30]. In our experiment, an acute cold stress (-6 °C) was used
660 to knock down individuals. By looking at the differential expression of these genes in
661 coma and non-coma individuals within and between species, we were not only able
662 to show that *stv* may interact with Hsc70s, but also that *T. alpestre* has a quicker
663 response in these genes compared with, *T. immigrans,* the related species not cold
664 adapted. Given these results, which are in line with the selection signature we

665 identified, it is probable that *stv* may play a similar role in the physiology of ants. As
666 we have shown in this study, ants completely lack Hsp70s (Figure S14) and retained
667 only the heat shock cognates (Hsc70s) with mainly two conserved paralogs (*Hsc70-*
668 *4d* and *Hsc70-4e*), which may have evolved to function as the Dipteran Hsp70s.
669 Although we found an interesting and promising correlation between chill shock, *stv*,
670 and Hsc70s, it will be important to continue exploring other expression patterns in
671 ant physiology to assess whether *stv* has tissue-specific expression patterns, as was
672 found in *D. melanogaster* [49], and to study its expression across different species
673 with different ecological adaptations.

674 One of the most enduring problems of evolutionary biology is explaining how
675 complex adaptive traits originate [50]. Although it is widely assumed that new traits
676 arise solely from selection on genetic variation, many researchers ask whether
677 phenotypic plasticity precedes and facilitates adaptation [51]. This alternative
678 scenario, the plasticity-first hypothesis, states that under novel conditions,
679 environmentally induced variation can be refined by selection and, depending on
680 whether plasticity is favoured, become developmentally canalized through genetic
681 assimilation. Under this scenario, environmentally induced phenotypic change can
682 precede and promote the evolutionary origins of a complex adaptive trait.

683 Four criteria have been proposed for verifying this hypothesis [51]: 1) the
684 focal trait needs to be environmentally induced in the ancestral-proxy lineages; 2)
685 cryptic genetic variation will be uncovered when ancestral-proxy lineages experience
686 the derived environment; 3) the focal trait will exhibit evidence of having undergone
687 an evolutionary change in its regulation, in its form, or in both in the derived
688 lineages; and 4) the focal trait will exhibit evidence of having undergone adaptive

689 refinement in the derived lineages. Though it is methodically difficult to prove each
690 point, our findings on the evolution and adaptation of *T. alpestre* to cold appear
691 consistent with plasticity-first evolution. In more detail, *T. immigrans*, the ancestral-
692 proxy lineage, shows considerable plasticity under variable temperature conditions
693 compared to other ant species (Criterion 1, Figure 1c), showing an ability to recover
694 from severe chill-shocks almost as good as the derived lineage (*T. alpestre*), and,
695 unexpectedly, better efficiency of Hex-t1 at lower rather than higher temperatures
696 (Figure 4a). We suggest that the derived lineage, *T. alpestre*, underwent genetic
697 assimilation of the cold response (Criterion 3), followed by evolutionary refinement
698 via changes in regulation, resulting, for example, in *stv* differential expression
699 (Criterion 4). Consistent with evolutionary refinement, the derived lineage has
700 evolved more extreme traits such as an overall improved activity of Hex-T1 with
701 reduced variability.

702 Our comparative genomic analysis shows how natural selection could trigger
703 complicated patterns of change in genomes, especially protein-coding sequences,
704 both in terms of diversifying selection and relaxation of purifying selection. Many of
705 these changes are in genes that may be associated with aspects of development,
706 either directly or through the associated complex changes in ecology and natural
707 history. Furthermore, this work characterized some candidate genes activities, both
708 *in vitro* and *in silico*, confirming our hypothesis and opening for the identification of
709 other genes possibly playing important role in adaptation to new ecological niches.
710 Overall, this study represents the first systematic attempt to provide a framework
711 for the genomic analysis of adaptation to extreme environments and underlines the

712 importance of studying organisms with different ecological niches in understanding
713 the genetic basis of ecological adaptation.
714
715
716 ## *Methods*
717 **Species-locality and climatic data**
718 To define the environment of *T. alpestre* and the other four *Tetramorium* species,
719 georeferenced occurrence data were compiled and complemented with those for
720 the native ranges of other ant species with available genomic data. Occurrence
721 records for 17 species (Table S2) were compiled from the Global Ant Biodiversity
722 Informatics database (GABI, Guénard et al., 2017, accessible via www.antmaps.org,
723 Janicki et al., 2016). The species determinations in the records were evaluated by
724 specialists in ant taxonomy, and erroneous data were excluded. Slightly incorrect
725 coordinates (i.e. in the ocean within 8.5 km of land) were assigned to the nearest
726 land point. Additionally, georeferenced occurrence data of two species (*T. alpestre,*
727 *T. immigrans*) were compiled from the latest taxonomic revision [18]. The
728 occurrence data of each species were spatially rarefied at a 5 km level using the R
729 package SPTHIN v. 0.1.0 [53], to counter spatial autocorrelation and to provide some
730 protection against sampling bias issues [54]. The Bioclim variables bio1-bio19 from
731 the WORLDCLIM database v. 1.4 [55] were extracted for each unique species locality at
732 a 2.5 arc min resolution (4.65 x 4.65 = 21.6225 km^2 at the equator) using ARCGIS v.
733 10.4 (ESRI, Redland, CA). A Wilcoxon-Mann-Whitney rank sum test as implemented
734 in the R package WILCOX.TEST (http://www.r-project.org) was adopted to test for
735 differences between *T. alpestre* and the other species individually for each of the

736 Bioclim variables. The Bonferroni-Holm sequential rejection procedure (Abdi 2010)
737 was used to control for multiple testing.
738

739 **Biological sample processing, genome size estimation, and genome sequencing**

740 Six colonies of *T. alpestre* ants were sampled, two for flow cytometry (nest 18811:
741 Penser Joch, Italy, 46.83379° N, 11.44652° E, 21 June 2016; nest 18813: Kühtai,
742 Austria, 47.27923, 11.07590° E, 21 June 2016), three for RNA extraction (nest 18586:
743 Zirmbachalm, Austria, 47.21920° N, 11.07530° E, 30 July 2015; nest 18590:
744 Zirmbachalm, 47.21910° N, 11.07620° E, 30 July 2015; nest 18594: Penser Joch,
745 46.82940°N, 11.43830° E, 2 August 2015), and one for DNA extraction (nest 18592:
746 Penser Joch, 46.83380° N, 11.44610° E, 2 August 2015). The ants were brought alive
747 to the laboratory (for flow cytometry) or immediately submerged in RNAlater (for
748 RNA extraction) or 96% ethanol (for DNA extraction) and stored at -70 °C (RNA) or -
749 20 °C (DNA).

750 Flow cytometry was used to estimate relative genome size from single *T.*
751 *alpestre* worker heads, using strain ISO-1 of *Drosophila melanogaster* as internal
752 standard (obtained from Bloomington Drosophila Stock Center, Indiana University,
753 Bloomington, USA; nuclear DNA content: 0.35 pg, Gregory 2018). Six individuals each
754 were measured from *T. alpestre* nests 18811 and 18813. In more detail, single *T.*
755 *alpestre* heads and, separately, heads of 10 ISO-1 females were chopped in 500 µL of
756 ice-cold Otto 1 Buffer (0.1 mol/L citric acid, 0.5% Tween20 (Merck KGaA, Darmstadt,
757 Germany)). The suspensions were filtered using 42-µm nylon meshes and incubated
758 for 5 min. Then, 1 mL of staining solution (DAPI (4 µL/mL) and 2-mercaptoethanol (2
759 µL/mL) in Otto 2 buffer (0.4 mol/L $Na_2HPO_4 \cdot 12$ H2O)) was added to each suspension.

760 For each measurement, 200 µL suspension of a single *T. alpestre* worker and 10 µL of
761 the ISO-1 standard were used. Fluorescence intensity of 3000 cell nuclei per sample
762 was measured with a Partec CyFlow space flow cytometer (Partec GmbH, Münster,
763 Germany). Gating and peak analysis were done automatically using the software
764 Partec Flo Max.

765 For the DNA and RNA extractions, the individuals were thawed on ice and the
766 gasters removed. DNA was extracted using the QIAamp DNA Micro Kit (Qiagen). The
767 extraction followed the protocol of the manufacturer, except that 20 µl RNase A
768 solution (Sigma-Aldrich) was added to the vials and incubated for 5 minutes at room
769 temperature. DNA concentration and quality were checked via Nanodrop (Thermo
770 Scientific) and gel electrophoresis. Two adult males were used to generate the
771 overlapping-read and the paired-end (PE) libraries, while for the mate-pair (MP)
772 libraries, four adult females were extracted of nest 18592. RNA was extracted using
773 the RNeasy Mini Kit (Qiagen) following the instructions of the manufacturer. Six
774 developmental stages were extracted as pools of individuals (nest 18586: worker
775 larvae – 30 individuals, worker pupae – 40, worker adult – 50; nest 18594: gyne
776 pupae – 2; nest 18590: gyne adult – 7, male adult – 7). RNA concentration and
777 quality were checked via Nanodrop (Thermo Scientific) and gel electrophoresis.

778 All *T. alpestre* DNA-seq and RNA-seq Illumina libraries were prepared by
779 IGATech (http://igatechnology.com/). Specifically, four libraries were used for
780 genome sequencing of *T. alpestre*: one overlapping-reads library (insert size ~400 bp;
781 paired-end (PE) sequencing with 2×250 bp read length), one regular PE library (insert
782 size ~400 bp, 2×100 bp read length), and two mate pair libraries (insert size ~4 kb,
783 2×100 bp read length) (see Table S2). For RNA-seq, six strand-specific libraries (insert

784 size ~500 bp, 2×200 bp read length) were constructed from three stages of workers
785 (larvae, pupae, adult), two stages of gynes (larvae and adult), and adult males. All
786 libraries were sequenced on an Illumina HiSeq 2500 platform in standard (DNA) and
787 RAPID (RNA) mode. For each *T. immigrans*, *T. bicarinatum*, *T. simillimum*, and *T.*
788 *parvispinum*, one TrueSeq library (insert size ~500 bp, 2×100 bp read length) was
789 prepared and sequenced on an Illumina HiSeq 4000 machine at OIST sequencing
790 center.
791
792 **Genomic and mitogenomic assembly and gene prediction**
793 Genomic sequencing reads of *T. alpestre* were filtered to remove low-quality reads
794 and PCR duplicates, corrected, and assembled using ALLPATHS-LG v. 52488 [56].
795 Contigs were first constructed based on overlapping reads and then scaffolded using
796 paired-end and mate-pair information from all DNA libraries. Given the high
797 coverage of the overlapping reads, based on the estimated genome size, multiple
798 trials were performed using different coverages to establish the best combination of
799 read coverage. The genomes of the other four related species were assembled using
800 SPADES v. 3.10.0 [57] with enabled read error correction algorithm, a *k*mer range of
801 21, 33, 55, 77, 99, and mismatch correction. Scaffolds belonging to contaminants
802 were identified and removed using BLOBOLOGY v. 2013-09-12 [58].
803 A bioinformatic pipeline that included combinatorial approaches of
804 homology, *ab initio*, and *de novo* methods were implemented to predict gene
805 models. REPEATMASKER v. 4.0.6 [59] was used to generate repeat hints. For the
806 homology-based exon hints, protein data sets of *D. melanogaster*, *Apis mellifera*,
807 *Nasonia vitripennis*, and 12 ant species (see Table S1) were aligned to the *T. alpestre*

808 genome using TBLASTN (*e*-value cut-off: $1e^{-5}$) and EXONERATE v. 2.2.0 [60]. As an *ab*
809 *initio* procedure, the merged RNA-seq libraries were mapped using STAR v. 020201
810 [61] (settings: ALIGNENDSTYPE Local; ALIGNINTRONMIN 30; ALIGNINTRONMAX 450,000;
811 OUTSJFILTERINTRONMAXVSREADN 80 100 500 1000 2000 5000 20,000;
812 ALIGNSJOVERHANGMIN 10; ALIGNTRANSCRIPTSPERREADNMAX 100,000) to generate the
813 splice-site position-specific scoring matrix (PSSM) and CUFFLINKS v. v2.2.1 [62] to
814 generate UTR and intron hints. This information was then integrated with GENEMARK-
815 ES v. 4 [63] to train the analysis and use the outcome under the BRAKER1 procedure
816 [63] and AUGUSTUS v. 3.2.1. [64]. To avoid missing genes due to unassembled
817 regions of the genome, previously unmapped reads were *de novo* assembled using
818 TRINITY v. 2.2.0 [65] and annotated with FRAMA [66]. The three results where then
819 merged, and Benchmarking Universal Single Copy Orthologs (BUSCO v. 1.2) [67] was
820 adopted to evaluate the annotation completeness together with integrity and
821 completeness evaluation using DELTABLAST [68]. A visual inspection of the annotation
822 and gene integrity was also performed on more than 250 loci.

823 All available complete ant mitochondrial genomes (mtDNA) were
824 downloaded from GENBANK [69] (Table S1). Illumina raw SRA data of other ant
825 species were downloaded and processed following Cicconardi et al., (2017b). In
826 brief, reads were quality filtered and assembled using IDBA-UD v. 1.1.1 [70] and
827 SPADES v. 3.10.0 [57], and annotated using MITOS v. 2 web server [71].

828

829 **Phylogenetic analysis**

830 Nuclear single-copy ortholog genes (scOGs) and complete mitochondrial genomes
831 were used to compute the ant species tree. Once scOGs were determined (see

832 below), single-locus trees and a species tree of concatenated scOGs were estimated
833 using Maximum Likelihood (ML) search as implemented in FastTree v. 2.1.8 SSE3
834 [72]. Gene trees were summarized with MP-EST [73] using the STRAW web-server
835 [74].

836 For the mtDNA phylogeny, only protein-coding genes were used. Each gene
837 was aligned using MACSE v. 1.01b [75], and all alignments were concatenated. The
838 best partitions and models of evolution were identified with PartitionFinder v.
839 1.1.1_Mac [76]. MtDNA trees were searched with ML and Bayesian Inference (BI)
840 algorithms. For ML, GARLI v. 2.01.1067 (http://code.google.com/p/garli) was used
841 performing 20 + 5 runs from random starting trees. The runs were continued until no
842 further improvement in log-likelihood was found, followed by 1000 bootstrap
843 pseudoreplicates. The results were summarized using SumTrees v. 3.3.1
844 (http://bit.ly/DendroPy), using *A. mellifera*, *N. vitripennis*, and *D. melanogaster* as
845 outgroups. For BI, BEAST v. 2 [77] was adopted to obtain the dated mtDNA tree and
846 to estimate the divergence of the most recent common ancestors (MRCAs)
847 (template: BEAST [78]). Fossil data were taken mainly from Moreau & Bell (2013)
848 and PaleoBio DB (paleobiodb.org) (see Table S4 for a detailed list of references) and
849 used as minimum calibration points constraining nodes in the topology as listed in
850 Table S4. Lognormal prior distribution was implemented with an offset
851 corresponding to the minimum fossil age, log(mean) of 1.0, and log(SD) of 1.0. Ten
852 independent runs of 8 x 10^7 generations were sampled every 100,000th generation.
853 Each partition was modelled with an uncorrelated relaxed clock with a Yule process,
854 species tree prior, and its best substitution model. Evolutionary models not
855 implemented in BEAUTi v2 [77] were manually edited in the xml file. Tracer v. 1.6

856 (http://beast.bio.ed.ac.uk/Tracer) was used to evaluate convergence and the
857 parameters' effective sampling size (ESS) and a burn-in was manually implemented
858 and roughly 10% from each run were removed. LogCombiner [77] and TreeAnnotator
859 [77] were used to summarize the results in a single consensus tree. The non-ant
860 species were excluded from the analysis, and no outgroup was adopted. Divergence
861 dates for the maximum-likelihood nuclear DNA (nuDNA) tree (only ants) were also
862 inferred using the penalized likelihood approach [79] implemented in the R package
863 APE v. 5.1 [80] by calibrating four nodes: the root, the *Tetramorium*, the
864 *Solenopsis*+*Monomorium*, and the Attini branches (Table S4).
865
866 **Functional annotation and orthologous-group dynamic evolution**
867 Genes from 19 ant and three outgroup species (Table S3) were clustered into
868 ortholog groups (OGs) with Hieranoid v. 2 [81], which implements InParanoid v. 8
869 [82] using a guide tree topology based on Moreau & Bell (2013) and BLASTp (*e*-value
870 cutoff of 10^{-5}) to improve the reciprocal hit accuracy [83], while homology
871 relationships were searched with deltaBLAST [68]. The functional annotation of each
872 putative protein coding sequence was performed by identifying both the protein
873 domain architecture and Gene Ontology (GO) terms. For each sequence, first
874 HMMER v. 3.1b2 (hmmscan) [84] was used to predict Pfam-A v. 31.0 domains, then
875 Domain Annotation by a Multi-objective Approach (DAMA) v. 2 [85] was applied to
876 identify architectures combining scores of domain matches, previously observed
877 multi-domain co-occurrence, and domain overlapping. Annotation of GO terms was
878 performed as implemented in the CATH assignments for large sequence datasets
879 [86]. Briefly, each input sequence was scanned against the library of CATH functional

880 families (FUNFAMS) HMMs [87] using HMMER3 [84] to assign FUNFAMS to regions on
881 the query sequence (with conditional *E*-value < 0.005). Then the GO annotations for
882 a matching FUNFAM were transferred to the query sequence with its confidence
883 scores, calculated by considering the GO term frequency among the annotated
884 sequences. Finally, a non-redundant set of GO annotations was retained, making up
885 the GO annotations for the query protein sequence [87].

886 To identify orthologous groups specifically expanded in the *T. alpestre*
887 genome, BADIRATE v. 1.35 [88] was performed twice, once for the nuDNA and one for
888 the mtDNA ultrametric tree topologies, reconstructing ancestral family sizes, gain,
889 death, and innovation (GDI), applying stochastic models and allowing estimation of
890 the family turnover rate (λ) by ML. Using this approach, we observed a correlation
891 between the turnover rate (λ) and branch lengths (Spearman correlation: ρ = -0.52,
892 *P*-value = 0), and a bimodal distribution of λ. We interpreted the distribution closer
893 to 0 as background noise and therefore considered only expanded and contracted
894 OGs with λ within the second distribution. The boundary between the two
895 distributions was defined by the valley values between the two, which were
896 computed using OPTIMIZE and APPROXFUN functions implemented in the STATS v. 3.4.0
897 package in R.

898

899 **Heat-shock protein family analysis**

900 From the functional annotation analyses of all Hymenoptera, all sequences bearing a
901 valid protein domain matching HSPs were extracted and aligned using the CLUSTALW
902 v.1.2.1 EBI web server [89] (settings: MBED true; MBEDITERATION true; ITERATIONS 5;
903 GTITERATIONS 5; HMMITERATIONS 5). A phylogenetic tree was computed using ML as

904 implemented in FASTTREE v. 2.1.8 SSE3 [72] and manually checked to identify and
905 provisionally annotate the five main subfamily proteins (Hsp90, Hsp70, Hsp60,
906 Hsp40, and sHsp) and their OG. Sequences falling outside clusters and with low
907 bootstrap values were manually checked with the DELTABLAST [68] web-server and
908 with the CONSERVED DOMAIN DATABASE (CDD) web-server [90] to identify and remove
909 bacterial contaminants and spurious/incomplete sequences. Then, nucleotide
910 sequences belonging to each subfamily were revers translated into amino acids,
911 aligned using CLUSTALW v.1.2.1, and the obtained protein alignment was used to
912 derive the nucleotide one. FASTTREE was used again to compute phylogenetic trees
913 for the five HSP subfamilies.
914
915 **Selection on single-copy orthologous groups and gene families**
916 All selected scOGs and HSP-OGs were scanned to evaluate selection signature on
917 coding regions of *T. alpestre*. This was done by computing the mean ω and the
918 relaxation of each branch of the phylogeny of *Tetramorium* spp. + *Vollenhovia*
919 *emeryi*. More specifically, adopting a pipeline similar to that in Cicconardi et al.
920 (2017a, 2017b), the signatures of diversifying selection were searched in codon-
921 based aligning groups of one-to-one orthologous ant genes with MACSE v. 1.01b
922 [75], filtering with GBLOCKS v. 0.91b (Castresana 2000) under a "relaxed" condition
923 [10,28,91] and using the aBSREL algorithm as implemented in the HYPHY batch
924 language [92] using a batch script (BRANCHSITEREL) in HYPHY
925 (http://github.com/veg/hyphy).
926 In parallel, the same scOGs and HSP-OGs were scanned using the RELAX test
927 [93] to search for putative signals of relaxation. In brief, RELAX tests the hypothesis

928 of evolutionary-rate relaxation in selected branches of a phylogenetic tree compared
929 with reference branches. A *k* value is computed to evaluate whether the selective
930 strength ω shifts towards neutrality. The rate of d_N/d_S (ω) can relax ($k < 1$), stay
931 stable ($k = 1$), or intensify ($k > 1$). For both scOGs and HSP-OGs, and for both the
932 ABSREL and RELAX tests, the Bonferroni-Holm sequential rejection procedure (Abdi
933 2010) was used to control the false discovery rate with a very stringent cutoff value
934 of 0.005 for the adjusted *P*-values applied [94]. Hsp60s and Hsp40s under relaxed
935 selection were analyzed using the GENEMANIA prediction server [95] to predict their
936 functions by creating an interaction network with genes by including protein and
937 genetic interactions, pathways, co-expression, co-localization, and protein domain
938 similarity.

939 Finally, the GOSTATS package for R [96] (settings: ANNOTATION org.Dm.eg.db;
940 CONDITIONAL TRUE; TESTDIRECTION over) was used to check for GO terms of biological
941 processes and KEGG pathway enrichments of scOGs under selection and relaxation
942 using the whole set of tested scOGs as background. *P*-values < 0.005 and < 0.05 were
943 implemented for GO terms and KEGG pathways, respectively.

944

945 **Synthesis of Hex-t1 gene and its enzyme activity assays**

946 Following identification as under diversifying selection in *T. alpestre* (see results
947 section "Dual signature of evolutionary pressures in single-copy OGs in the *T.*
948 *alpestre* genome"), synthesis of genes and enzymes was attempted for
949 phosphofructokinase, transketolase, phosphoglucosemutase, and Hex-t1 was
950 successful just for the Hex-t1. In more detail, the Hex-t1 genes of *Tetramorium*
951 *alpestre* (1398 bp) and *T. immigrans* (1386 bp) were synthesized by ThermoFisher

952 lifetechnologies (Carlsbad, CA, USA) and Eurofins Genomics (Ebersberg, Germany),
953 respectively. Both constructs carried a 5' EcoRI and a 3' HindIII recognition site for
954 subsequent cloning as well as a C-terminal His-tag. The genes were excised from
955 their carrier backbones by enzymatic digestion. The fragments were purified by
956 agarose gel excision and sub-cloned into the pACEBac1 expression vector
957 (MULTIBACTM, Geneva Biotech, Geneva, Switzerland). Correct insertion was verified
958 by Sanger sequencing, and transformation into *D. melanogaster* cells and protein
959 expression followed the standard protocols of the expression vector manufacturer.
960 After cell harvesting, the presence of the recombinant protein in the cell lysates was
961 verified by a western blot targeting the His-tag. The relative amounts of protein
962 were quantified from the western blot using ImageJ (https://imagej.net/Welcome).
963 The cell lysates were centrifuged at 4000 rpm at 4 °C for 5 min, and the supernatant
964 was directly used as protein source.

965 Enzyme activity assays followed a modified protocol of Crabtree &
966 Newsholme (1972). Briefly, 1000 µl assay medium containing 75 mM Tris pH=7.5
967 (Merck Millipore, Burlington, MA, USA), 7.5 mM $MgCl_2$ (Merck), 0.8 mM EDTA
968 (Sigma-Aldrich, St. Louis, MO, USA), 1.5 mM KCl (Merck), 4.0 mM mercaptoethanol
969 (Sigma), 0.4 mM NADP+ (Abcam, Cambridge, UK), 2.5 mM ATP (Abcam), and 10 mM
970 creatine phosphate (Abcam) were placed in a disposable cuvette (Brand, Germany)
971 in a SPECORD 210 PLUS UV/VIS spectrophotometer (Analytic Jena AG, Germany)
972 with an attached 1157P programmable thermostate (VWR, Radnor, PA, USA).
973 Variable amounts of D-glucose (VWR) and synthetic protein were added, and the
974 cuvette was allowed to reach the assay temperature. Then, the reaction was started

975 by adding 15 U creatine phosphokinase (Sigma) and 5 U glucose 6-phospahte
976 dehydrogenase (Sigma).

977 The Hex-t1 activity was measured in triplicate as the rate of reduction of
978 $NADP^+$ causing an extinction at 340 nm. Each measurement lasted for 60 s, and every
979 second, an extinction value was recorded. For the determination of the Michaelis-
980 Menten constant (Km-value) of both enzymes, assays were conducted along a
981 substrate gradient from 0.00167 to 24.30000 mM D-glucose, achieving saturated
982 extinction curves. Assay temperatures ranged from 6 °C to 36 °C with 10 K
983 increments. The software PRISM v. 8 (GraphPad Software, San Diego, CA, USA) was
984 used to calculate the Km-values.

985

986 **Hex-t1 structure modelling and molecular dynamics simulations**

987 The three-dimensional structures of *T. alpestre* and *T. immigrans* Hex-t1 proteins are
988 not available. To study the effect of the amino acid changes on the proteins'
989 function/dynamics, we modeled the structures of these two proteins via homology
990 modelling. The X-ray structure of the Hex-t1 isoform from *Schistosoma mansoni* [38],
991 solved at 2.6 Å resolution was used as a template and the models were generated
992 using the software MODELLER v. 9.21 [97]. Once we obtained the models, we built the
993 simulative systems to run molecular dynamics simulations. The two proteins
994 TalpHex-t1 and TimmHex-t1, in complex with the glucose substrate, were immersed
995 in a triclinic box filled with TIP3P water molecules [98] and rendered electroneutral
996 by the addition of cloride counterions. The topology of the two systems, consisting
997 of ~ 80.000 atoms each, was built using the AMBER14 force field [99], further
998 converted in GROMACS v. 4.6 [100] format using ACPYPE [101]. Simulations were run

999 on a GPU cluster using GROMACS v. 4.6 with the following protocol: 1) 25000 steps of
1000 steepest descent followed by 25000 steps of coniugate gradient minimization; 2)
1001 5*100 ps of equilibration runs in the NVT environment starting at 50 K and
1002 performed by increasing the temperature of 50 K after each run untill a final value of
1003 250 K; 3) 5*100 ps of equilibration runs in the NPT environment starting at 50 K and
1004 performed by increasing the temperature of 50 K after each run untill a final value of
1005 250 K; 4) the equilibrated systems were simulated for 500 ns at 279 or 300 K, i.e. at 6
1006 or 26 °C, for a total of four simulations and 2 µs of sampling. Electrostatic
1007 interactions were considered by means of the Particle Mesh Ewald method (PME)
1008 with a cut off of 1.2 nm for the real space and Van der Waals interactions [102].
1009 Bond lengths and angles were constrained via the LINCS algorithm [103]. The
1010 temperatures were kept constant at 279 or 300 °K by using the velocity rescale with
1011 a coupling constant of 0.1 ps, and the pressure was kept at 1 bar using the
1012 Parrinello-Rahman barostat with a coupling constant of 1.0 ps during sampling [104].
1013 The four trajectories were collected and comparative analyses performed with the
1014 GROMACS suite or with in house written code.
1015

1016 **Chill coma assay and quantitative qRT-PCR assay**

1017 The cold hardiness and expression of selected genes were compared between *T.*
1018 *alpestre* and *T. immigrans* using chill coma assays. From four colonies additional to
1019 those used for other experiments in this study, two per species, about 500 workers
1020 per colony were collected in August 2017; *T. alpestre*: Jaufenpass (18980: 46.83791°
1021 N, 11.29768° E) and Penser Joch (18978: 46.81402° N, 11.44198° E); *T. immigrans*:
1022 Vienna (18977: 48.12448° N, 16.43522° E.; 18983: 48.31341° N, 16.42529° E).

1023 Workers were collected alive using aspirators, transported to the laboratory in
1024 Innsbruck, and kept in polypropylene boxes (18.5 cm x 11 cm) with chambers of
1025 various sizes. The walls of the boxes were Fluon-coated (GP1, De Monchy
1026 International BV, Rotterdam, Netherlands) to prevent workers from escaping. Food
1027 (sugar-honey-water and deep-frozen *Drosophila* flies) and tap water were provided
1028 *ad libitum* three times a week. Workers were kept in a climate chamber (MIR-254,
1029 Panasonic, Etten Leur, Netherlands) at a 12L:12D photoperiod at constant 18 °C. The
1030 temperature of 18 °C was the average of the monthly mean temperatures
1031 experienced at 200 m a.s.l. (*T. immigrans*, approx. 23.7 °C) and at 2000 m a.s.l. (*T.*
1032 *alpestre*, approx. 12.3 °C) in July 2017. Workers were acclimatized to 18 °C for at
1033 least three weeks before the chill coma assays.

1034 After the acclimation period, from each of the four colonies, 90 workers were
1035 randomly chosen and divided into 18 equivalent pools, each were used for the chill
1036 coma and for the control assays. For the chill coma assay, worker pools were
1037 transferred into empty 5-ml glass vials, sealed, and immersed for 6.5 hours in a
1038 water:ethane-1,2-diol mix (1:1) bath set at -6 °C; in pilot experiments, this
1039 temperature had been determined as the highest temperature at which the ants had
1040 fallen into chill coma after 6.5 hours exposition; the temperature was identical for
1041 both species. Temperature was monitored using an electronic thermometer (TFX
1042 430, ebro Electronic GmbH; Ingolstadt, Germany) with an accuracy of 0.05 °C
1043 inserted into an additional, empty vial treated in the same way as the vials with ants.
1044 After chill coma, the vials were transferred to a climate chamber at 18 °C, and the
1045 workers were allowed to recover for 0.5, 2.0, and 4.0 hours. For each recovery time,
1046 three replicate pools were used. After recovery, the workers were transferred to

1047 RNAse-free reaction tubes and killed in liquid nitrogen. Tubes were stored at -70 °C.
1048 For the control assay, worker pools were transferred to empty glass vials and placed
1049 in the climate chamber at 18 °C. After 6.5 hours plus respective recovery time, the
1050 workers were killed and stored as described above.

1051 For each recovery time and replicate, complete worker pools were used for
1052 the molecular analyses. RNA was extracted using the NucleoSpin® RNA Kit
1053 (Macherey-Nagel, Düren, Germany) following the instructions of the manufacturer.
1054 First-strand cDNA was synthesized using 200 U RevertAid reverse transcriptase (all
1055 reagents by Thermo Fisher Scientific, Waltham, USA), 40 U RiboLock RNase inhibitor,
1056 5 µM random hexamer primers, 100 µM dNTPs, and 2 µl RNA extract in a total
1057 volume of 40 µl. The mixture was incubated for 5 min at 25 °C, 60 min at 42 °C, and 5
1058 min at 70 °C on a UnoCycler 1200 (VWR, Radnor, USA). Quantitative PCR was
1059 conducted on a Rotorgene Q (Qiagen) PCR system. Each reaction contained 1×
1060 Rotor-Gene SYBR Green PCR Mastermix (Qiagen), 0.2 µM of primers, specifically
1061 designed for target genes (Table S12), and 1 µl cDNA in a total volume of 10 µl. All
1062 qRT-PCR reactions were performed as triplicates. Cycling conditions were 95 °C for 5
1063 min, followed by 40 cycles of 94 °C for 15 s, 58 °C for 10 s, and 72 °C for 15 s.
1064 Fluorescence was acquired at the end of each elongation step. PCR was followed by
1065 a melting curve analysis from 60 to 95 °C with 0.1 °C increments held for 5 s before
1066 fluorescence acquisition. Cycle threshold (Ct) values were subtracted from total
1067 number of cycles. On these values, linear-regression analyses were based separately
1068 for each target gene, treatment, recovery time, and species (in each regression
1069 analysis, two populations and three replicate values were used, i.e., n = 6) using the
1070 housekeeping gene as independent variable and the target gene as dependent

1071 variable. For each target gene / recovery-time combination one-way analysis of
1072 covariance (ANCOVA) followed by Bonferroni-Holm correction was then used to
1073 identify significant differences between treatments within species and between
1074 species within treatment.
1075
1076
1077 ***Data access***
1078 The sequence data from this study have been submitted to the National Center for
1079 Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject
1080 numbers PRJNA532334, PRJNA533534, PRJNA533535, PRJNA533536 and
1081 PRJNA533537 (https://www.ncbi.nlm.nih.gov/bioproject/). The mtDNA genome
1082 assemblies have been submitted to the NCBI (GenBank accessions: MK861047 -
1083 MK861070). All accession codes of deposited and retrieved data are provided in
1084 Table S1.
1085
1086
1087 ***Acknowledgements***
1088 Philipp Andesner and Stefan Gross for valuable work in the laboratory. Research was
1089 supported by the Austrian Science Fund (FWF, P23409, P30861). The computational
1090 results presented here have been achieved in part by using the HPC infrastructure of
1091 the University of Innsbruck (LEO), the Vienna Scientific Cluster (VSC), the
1092 collaborative system of the Universities Innsbruck and Linz (MACH), and the Swiss
1093 National Supercomputing Centre (CSCS) hosted in Lugano.

1094

1095

1096 ***References***

1097 1. Soberón J. Grinnellian and Eltonian niches and geographic distributions of

1098 species. Ecology Letters. 2007. pp. 1115–1123. doi:10.1111/j.1461-

1099 0248.2007.01107.x

1100 2. Lamprecht A, Semenchuk PR, Steinbauer K, Winkler M, Pauli H. Climate

1101 change leads to accelerated transformation of high-elevation vegetation in

1102 the central Alps. New Phytol. 2018; 447–459. doi:10.1111/nph.15290

1103 3. Rogora M, Frate L, Carranza ML, Freppaz M, Stanisci A, Bertani I, et al.

1104 Assessment of climate change effects on mountain ecosystems through a

1105 cross-site analysis in the Alps and Apennines. Sci Total Environ. Elsevier B.V.;

1106 2018;624: 1429–1442. doi:10.1016/j.scitotenv.2017.12.155

1107 4. Körner C, Paulsen J, Spehn EM. A definition of mountains and their bioclimatic

1108 belts for global comparisons of biodiversity data. Alp Bot. 2011;121: 73–78.

1109 doi:10.1007/s00035-011-0094-4

1110 5. Clark MS, Worland MR. How insects survive the cold: molecular

1111 mechanisms—a review. J Comp Physiol B. 2008;178: 917–933.

1112 doi:10.1007/s00360-008-0286-4

1113 6. Parker DJ, Wiberg RAW, Trivedi U, Tyukmaeva VI, Gharbi K, Butlin RK, et al.

1114 Inter and Intraspecific Genomic Divergence in Drosophila montana Shows

1115 Evidence for Cold Adaptation. 2018;10: 2086–2101. doi:10.1093/gbe/evy147

1116 7. Keeling CI, Yuen MM, Liao NY, Roderick Docking T, Chan SK, Taylor G a, et al.

1117 Draft genome of the mountain pine beetle, Dendroctonus ponderosae

1118 Hopkins, a major forest pest. Genome Biol. BioMed Central Ltd; 2013;14: R27.
1119 doi:10.1186/gb-2013-14-3-r27
1120 8. Kelley JL, Peyton JT, Fiston-Lavier A-S, Teets NM, Yee M-C, Johnston JS, et al.
1121 Compact genome of the Antarctic midge is likely an adaptation to an extreme
1122 environment. Nat Commun. Nature Publishing Group; 2014;5: 4611.
1123 doi:10.1038/ncomms5611
1124 9. Macdonald HC, Cunha L, Bruford MW, Florestas E, Macdonald HC, Sir T, et al.
1125 Development of genomic resources for four potential environmental
1126 bioindicator species: *Isoperla grammatica*, *Amphinemura sulcicollis*, *Oniscus*
1127 *asellus* and *Baetis rhodani*. bioRxivorg. 2016; doi:10.1101/046227
1128 10. Cicconardi F, Di Marino D, Olimpieri PP, Arthofer W, Schlick-Steiner BC, Steiner
1129 FM. Chemosensory adaptations of the mountain fly Drosophila nigrosparsa
1130 (Insecta: Diptera) through genomics' and structural biology's lenses. Sci Rep.
1131 Nature Publishing Group; 2017;7: 43770. doi:10.1038/srep43770
1132 11. King AM, MacRae TH. Insect Heat Shock Proteins During Stress and Diapause.
1133 Annu Rev Entomol. 2015;60: 59–75. doi:10.1146/annurev-ento-011613-
1134 162107
1135 12. Nygaard S, Hu H, Li C, Schiøtt M, Chen Z, Yang Z, et al. Reciprocal genomic
1136 evolution in the ant-fungus agricultural symbiosis. Nat Commun. 2016; 1–9.
1137 doi:10.1038/ncomms12233
1138 13. Hölldobler B, Wilson EO. The Ants. Harvard University Press. 1990.
1139 14. Bolton B. An online catalog of the ants of the world. Available from
1140 http://antcat.org [Internet]. 2018.
1141 15. Seifert B. The ants of Central and North Europe. Tauer, editor. Lutra Verlags-

1142 und Vertriebsgesellschaft; 2018.

1143 16. Wilson EO. Success and Dominance in Ecosystems: The Case of Social Insects.
1144 Excellence in ecology. 1990.

1145 17. Christian KA, Morton SR. Extreme thermophilia in a central Australian ant,
1146 Melophorus bagoti. Physiol Zool. 1992;65: 885–905.

1147 18. Wagner HC, Arthofer W, Seifert B, Muster C, Steiner FM, Schlick-Steiner BC.
1148 Light at the end of the tunnel: Integrative taxonomy delimits cryptic species in
1149 the Tetramorium caespitum complex (Hymenoptera: Formicidae).
1150 Myrmecological News. 2017;25: 95–129.

1151 19. Steiner FM, Seifert B, Moder K, Schlick-Steiner BC. A multisource solution for a
1152 complex problem in biodiversity research: Description of the cryptic ant
1153 species Tetramorium alpestre sp.n. (Hymenoptera: Formicidae). Zool Anzeiger
1154 - A J Comp Zool. 2010;249: 223–254. doi:10.1016/j.jcz.2010.09.003

1155 20. Krapf P, Russo L, Arthofer W, Möst M, Steiner FM, Schlick-Steiner BC. An
1156 Alpine ant's behavioural polymorphism: monogyny with and without internest
1157 aggression in Tetramorium alpestre. Ethol Ecol Evol. 2018;30: 220–234.
1158 doi:10.1080/03949370.2017.1343868

1159 21. Liu C, Guénard B, Garcia FH, Yamane S, Blanchard B, Yang DR, et al. New
1160 records of ant species from Yunnan, China. Zookeys. 2015;78: 17–78.
1161 doi:10.3897/zookeys.477.8775

1162 22. Bertelsmeier C, Ollier S, Liebhold A, Keller L. Recent human history governs
1163 global ant invasion dynamics. Nat Ecol Evol. 2017;1. doi:10.1038/s41559-017-
1164 0184

1165 23. Guénard B, Weiser MD, Gómez K, Narula N, Economo EP. The Global Ant

1166 Biodiversity Informatics (GABI) database: Synthesizing data on the geographic
1167 distribution of ant species (Hymenoptera: Formicidae). Myrmecological News.
1168 2017. pp. 83–89.
1169 24. Jayakumar PC, Shouche YS, Patole MS. Functional analysis of *Drosophila*
1170 *melanogaster* hexokinase Hex-A locus: multiple Initiator-like elements
1171 enhance DPE containing promoter activity. 2007;16: 3–13.
1172 doi:10.1111/j.1365-2583.2006.00664.x
1173 25. Zachos J, Pagani M, Sloan L, Thomas E, Billups K. Trends, Rythms, and
1174 Aberration in Global Climate 65 Ma to Present. Science (80-). 2001;292: 686–
1175 693. doi:10.1126/science.1059412
1176 26. Moreau CS, Bell CD. Testing The Museum Versus Cradle Tropical Biological
1177 Diversity Hypothesis: Phylogeny, Diversification, And Ancestral Biogeographic
1178 Range Evolution Of The Ants. Evolution (N Y). 2013;67: 2240–2257.
1179 doi:10.1111/evo.12105
1180 27. Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M, Keller L. Patterns
1181 of positive selection in seven ant genomes. Mol Biol Evol. 2014;31: 1661–
1182 1685. doi:10.1093/molbev/msu141
1183 28. Cicconardi F, Marcatili P, Arthofer W, Schlick-Steiner BC, Steiner FM. Positive
1184 diversifying selection is a pervasive adaptive force throughout the Drosophila
1185 radiation. Mol Phylogenet Evol. 2017;112: 230–243.
1186 doi:10.1016/j.ympev.2017.04.023
1187 29. Nguyen AD, Gotelli NJ, Cahan SH. The evolution of heat shock protein
1188 sequences, cis-regulatory elements, and expression profiles in the eusocial
1189 Hymenoptera. BMC Evol Biol. BMC Evolutionary Biology; 2016;16: 15.

1190 doi:10.1186/s12862-015-0573-0

1191 30. Colinet H, Hoffmann A. Gene and protein expression of *Drosophila* Starvin
1192 during cold stress and recovery from chill coma. Insect Biochem Mol Biol.
1193 Elsevier Ltd; 2010;40: 425–428. doi:10.1016/j.ibmb.2010.03.002

1194 31. Kuhlemann J. Paleogeographic and paleotopographic evolution of the Swiss
1195 and Eastern Alps since the Oligocene. Glob Planet Change. 2007;58: 224–236.
1196 doi:10.1016/j.gloplacha.2007.03.007

1197 32. Wallis GP, Waters JM, Upton P, Craw D. Transverse Alpine Speciation Driven
1198 by Glaciation. Trends Ecol Evol. Elsevier Ltd; 2016;31: 916–926.
1199 doi:10.1016/j.tree.2016.08.009

1200 33. Uchida Y, Uesaka M, Yamamoto T, Takeda H, Irie N. Embryonic lethality is not
1201 sufficient to explain hourglass-like conservation of vertebrate embryos.
1202 Evodevo. BioMed Central; 2018;9: 1–11. doi:10.1186/s13227-018-0095-0

1203 34. Gerken AR, Eller OC, Hahn D a, Morgan TJ. Constraints, independence, and
1204 evolution of thermal plasticity: Probing genetic architecture of long- and
1205 short-term thermal acclimation. Proc Natl Acad Sci U S A. 2015;112: 4399–
1206 404. doi:10.1073/pnas.1503456112

1207 35. Archambault V, Lépine G, Kachaner D. Understanding the Polo Kinase
1208 machine. Oncogene. 2015;34: 4799–4807. doi:10.1038/onc.2014.451

1209 36. Harpur BA, Kent CF, Molodtsova D, Lebon JMD, Alqarni AS, Owayss AA, et al.
1210 Population genomics of the honey bee reveals strong signatures of positive
1211 selection on worker traits. Proc Natl Acad Sci. 2014;111: 2614–2619.
1212 doi:10.1073/pnas.1315506111

1213 37. Xian-Wu L, Wei-Hua X. Hexokinase is a key regulator of energy metabolism

1214 and ROS activity in insect lifespan extension. Aging (Albany NY). 2016;8: 245–
1215 258. Available:
1216 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4789580/pdf/aging-08-
1217 245.pdf
1218 38. Mulichak AM, Wilson JE, Padmanabhan K, Garavito RM. The structure of
1219 mammalian hexokinase-1. Nat Struct Biol. 2002; doi:10.1038/811
1220 39. Fischer MK, Shingleton AW. Host plant and ants influence the honeydew
1221 sugar. Funct Ecol. 2001;15: 544–550.
1222 40. Pringle EG, Novo A, Ableson I, Barbehenn R V, Vannette RL. Plant-derived
1223 differences in the composition of aphid honeydew and their effects on
1224 colonies of aphid-tending ants. 2014; doi:10.1002/ece3.1277
1225 41. Ivens ABF, Kronauer DJC, Pen I, Weissing FJ, Boomsma JJ. Ants farm
1226 subterranean aphids mostly in single clone groups - An example of prudent
1227 husbandry for carbohydrates and proteins? BMC Evol Biol. 2012;12.
1228 doi:10.1186/1471-2148-12-106
1229 42. Haslbeck M, Vierling E. A first line of stress defense: Small heat shock proteins
1230 and their function in protein homeostasis. J Mol Biol. Elsevier Ltd; 2015;427:
1231 1537–1548. doi:10.1016/j.jmb.2015.02.002
1232 43. Yoshida S, Tsutsumi S, Muhlebach G, Sourbier C, Lee M-J, Lee S, et al.
1233 Molecular chaperone TRAP1 regulates a metabolic switch between
1234 mitochondrial respiration and aerobic glycolysis. Proc Natl Acad Sci. 2013;110:
1235 E1604–E1612. doi:10.1073/pnas.1220659110
1236 44. Liu D, Hu J, Agorreta J, Cesario A, Zhang Y, Harris AL, et al. Tumor necrosis
1237 factor receptor-associated protein 1(TRAP1) regulates genes involved in cell

1238 cycle and metastases. Cancer Lett. 2010;296: 194–205.
1239 doi:10.1016/j.canlet.2010.04.017
1240 45. Prescott SL, Srinivasan R, Marchetto MC, Grishina I, Narvaiza I, Selleri L, et al.
1241 Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp
1242 Neural Crest. Cell. 2015;163: 68–84. doi:10.1016/j.cell.2015.08.036
1243 46. Danko CG, Choate LA, Marks BA, Rice EJ, Wang Z, Chu T, et al. Dynamic
1244 evolution of regulatory element ensembles in primate CD4+ T cells. Nat Ecol
1245 Evol. Springer US; 2018;2. doi:10.1038/s41559-017-0447-5
1246 47. Doong H, Vrailas A, Kohn EC. What's in the "BAG"? - A functional domain
1247 analysis of the BAG-family proteins. Cancer Lett. 2002;188: 25–32.
1248 doi:10.1016/S0304-3835(02)00456-1
1249 48. Colinet H, Hoffmann AA. Comparing phenotypic effects and molecular
1250 correlates of developmental, gradual and rapid cold acclimation responses in
1251 Drosophila melanogaster. Funct Ecol. 2012;26: 84–93. doi:10.1111/j.1365-
1252 2435.2011.01898.x
1253 49. Coulson M, Robert S, Saint R. Drosophila starvin encodes a tissue-specific
1254 BAG-domain protein required for larval food uptake. Genetics. 2005;171:
1255 1799–1812. doi:10.1534/genetics.105.043265
1256 50. Wagner GP, Lynch VJ. Evolutionary novelties. Curr Biol. 2010;20: 48–52.
1257 Available: https://pdf.sciencedirectassets.com/272099/1-s2.0-
1258 S0960982210X00021/1-s2.0-S0960982209019459/main.pdf?x-amz-security-
1259 token=AgoJb3JpZ2luX2VjECIaCXVzLWVhc3QtMSJHMEUCIDsUh2jRKJL1EeOiMr
1260 NglWa%2BljWFLis7tEu%2FF65JtK%2FrAiEA3Ge%2FEhYUyff41n0NLuT%2F6f3O
1261 MYi9qUIAiE

1262 51. Levis NA, Pfennig DW. Evaluating ‘ Plasticity-First ’ Evolution in Nature: Key
1263 Criteria and Empirical Approaches. Trends Ecol Evol. Elsevier Ltd; 2016;31:
1264 563–574. doi:10.1016/j.tree.2016.03.012
1265 52. Janicki J, Narula N, Ziegler M, Guénard B, Economo EP. Visualizing and
1266 interacting with large-volume biodiversity data using client-server web-
1267 mapping applications: The design and implementation of antmaps.org. Ecol
1268 Inform. 2016;32: 185–193. doi:10.1016/j.ecoinf.2016.02.006
1269 53. Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. spThin:
1270 An R package for spatial thinning of species occurrence records for use in
1271 ecological niche models. Ecography (Cop). 2015;38: 541–545.
1272 doi:10.1111/ecog.01132
1273 54. Merow C, Smith MJ, Silander JA. A practical guide to MaxEnt for modeling
1274 species’ distributions: What it does, and why inputs and settings matter.
1275 Ecography (Cop). 2013;36: 1058–1069. doi:10.1111/j.1600-0587.2013.07872.x
1276 55. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution
1277 interpolated climate surfaces for global land areas. Int J Climatol. 2005;25:
1278 1965–1978. doi:10.1002/joc.1276
1279 56. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al.
1280 High-quality draft assemblies of mammalian genomes from massively parallel
1281 sequence data. Proc Natl Acad Sci U S A. 2011;108: 1513–1518.
1282 doi:10.1073/pnas.1017351108
1283 57. Bankevich A, Nurk S, Antipov D, Gurevich A a., Dvorkin M, Kulikov AS, et al.
1284 SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell
1285 Sequencing. J Comput Biol. 2012;19: 455–477. doi:10.1089/cmb.2012.0021

1286 58. Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. Blobology: exploring
1287 raw genome data for contaminants, symbionts and parasites using taxon-
1288 annotated GC-coverage plots. Front Genet. 2013;4: 237.
1289 doi:10.3389/fgene.2013.00237
1290 59. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015 . [Internet].
1291 http://www.repeatmasker.org. 2013. Available: http://repeatmasker.org
1292 60. Slater GSC, Birney E. Automated generation of heuristics for biological
1293 sequence comparison. BMC Bioinformatics. 2005;6: 1–11. doi:10.1186/1471-
1294 2105-6-31
1295 61. Dobin A, Davis C a, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR:
1296 ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21.
1297 doi:10.1093/bioinformatics/bts635
1298 62. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential
1299 gene and transcript expression analysis of RNA-seq experiments with TopHat
1300 and Cufflinks. Nat Protoc. Nature Publishing Group; 2012;7: 562–578.
1301 doi:10.1038/nprot.2012.016
1302 63. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1:
1303 Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and
1304 AUGUSTUS. Bioinformatics. 2015;32: 767–769.
1305 doi:10.1093/bioinformatics/btv661
1306 64. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in
1307 eukaryotes with a generalized hidden Markov model that uses hints from
1308 external sources. BMC Bioinformatics. 2006;7: 62. doi:10.1186/1471-2105-7-
1309 62

1310 65. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson D a, Amit I, et al. Full-
1311 length transcriptome assembly from RNA-Seq data without a reference
1312 genome. Nat Biotechnol. 2011;29: 644–52. doi:10.1038/nbt.1883
1313 66. Bens M, Sahm A, Groth M, Jahn N, Morhart M, Holtze S, et al. FRAMA□: from
1314 RNA-seq data to annotated mRNA assemblies. BMC Genomics. BMC
1315 Genomics; 2016; 1–12. doi:10.1186/s12864-015-2349-8
1316 67. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM.
1317 BUSCO: Assessing genome assembly and annotation completeness with
1318 single-copy orthologs. Bioinformatics. 2015;31: 3210–3212.
1319 doi:10.1093/bioinformatics/btv351
1320 68. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL.
1321 Domain enhanced lookup time accelerated BLAST. Biol Direct. 2012;7: 12.
1322 doi:10.1186/1745-6150-7-12
1323 69. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.
1324 GenBank. Nucleic Acids Res. 2014;43: D30–D35. doi:10.1093/nar/gku1216
1325 70. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a *de novo* assembler for
1326 single-cell and metagenomic sequencing data with highly uneven depth.
1327 Bioinformatics. 2012;28: 1420–1428. doi:10.1093/bioinformatics/bts174
1328 71. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al.
1329 MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol
1330 Phylogenet Evol. Elsevier Inc.; 2013;69: 313–319.
1331 doi:10.1016/j.ympev.2012.08.023
1332 72. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood
1333 trees for large alignments. PLoS One. 2010;5: e9490.

1334 doi:10.1371/journal.pone.0009490

1335 73. Liu L, Yu L, Edwards S V. A maximum pseudo-likelihood approach for
1336 estimating species trees under the coalescent model. BMC Evol Biol. 2010;10:
1337 302. doi:10.1186/1471-2148-10-302

1338 74. Shaw TI, Ruan Z, Glenn TC, Liu L. STRAW: Species TRee Analysis Web server.
1339 Nucleic Acids Res. 2013;41: 238–241. doi:10.1093/nar/gkt377

1340 75. Ranwez V, Harispe S, Delsuc F, Douzery EJP. MACSE: Multiple alignment of
1341 coding SEquences accounting for frameshifts and stop codons. PLoS One.
1342 2011;6: e22594. doi:10.1371/journal.pone.0022594

1343 76. Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. Selecting optimal
1344 partitioning schemes for phylogenomic datasets. BMC Evol Biol. 2014;14: 82.
1345 doi:10.1186/1471-2148-14-82

1346 77. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a
1347 software platform for Bayesian evolutionary analysis. PLoS Comput Biol.
1348 2014;10: e1003537. doi:10.1371/journal.pcbi.1003537

1349 78. Heled J, Drummond AJ. Bayesian inference of species trees from multilocus
1350 data. Mol Biol Evol. 2010;27: 570–80. doi:10.1093/molbev/msp274

1351 79. Sanderson MJ. Estimating absolute rates of molecular evolution and
1352 divergence times: a penalized likelihood approach. Mol Biol Evol. 2002;19:
1353 101–109. doi:10.1093/oxfordjournals.molbev.a003974

1354 80. Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution
1355 in R language. Bioinformatics. 2004;20: 289–290.
1356 doi:10.1093/bioinformatics/btg412

1357 81. Kaduk M, Sonnhammer E. Improved orthology inference with Hieranoid 2.

1358 Bioinformatics. 2017;5: btw774. doi:10.1093/bioinformatics/btw774

1359 82. Sonnhammer ELL, Östlund G. InParanoid 8: Orthology analysis between 273
1360 proteomes, mostly eukaryotic. Nucleic Acids Res. 2015;43: D234–D239.
1361 doi:10.1093/nar/gku1203

1362 83. Edgar RC. Search and clustering orders of magnitude faster than BLAST.
1363 Bioinformatics. 2010;26: 2460–2461. doi:10.1093/bioinformatics/btq461

1364 84. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7.
1365 doi:10.1371/journal.pcbi.1002195

1366 85. Bernardes JS, Vieira FRJ, Zaverucha G, Carbone A. A multi-objective
1367 optimization approach accurately resolves protein domain architectures.
1368 Bioinformatics. 2015;32: 345–353. doi:10.1093/bioinformatics/btv582

1369 86. Das S, Sillitoe I, Lee D, Lees JG, Dawson NL, Ward J, et al. CATH FunFHMMer
1370 web server: Protein functional annotations using functional family
1371 assignments. Nucleic Acids Res. 2015;43: W148–W153.
1372 doi:10.1093/nar/gkv488

1373 87. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, et al. CATH:
1374 Comprehensive structural and functional annotations for genome sequences.
1375 Nucleic Acids Res. 2015;43: D376–D381. doi:10.1093/nar/gku947

1376 88. Librado P, Vieira FG, Rozas J. BadiRate: Estimating family turnover rates by
1377 likelihood-based methods. Bioinformatics. 2012;28: 279–281.
1378 doi:10.1093/bioinformatics/btr623

1379 89. Larkin M a., Blackshields G, Brown NP, Chenna R, Mcgettigan P a., McWilliam
1380 H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23: 2947–
1381 2948. doi:10.1093/bioinformatics/btm404

1382 90. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE:
1383 Functional classification of proteins via subfamily domain architectures.
1384 Nucleic Acids Res. 2017;45: D200–D203. doi:10.1093/nar/gkw1129
1385 91. Parker J, Tsagkogeorga G, Cotton J a, Liu Y, Provero P, Stupka E, et al. Genome-
1386 wide signatures of convergent evolution in echolocating mammals. Nature.
1387 Nature Publishing Group; 2013;502: 1–9. doi:10.1038/nature12511
1388 92. Kosakovsky Pond SL, Frost SDW, Muse S V. HyPhy: Hypothesis testing using
1389 phylogenies. Bioinformatics. 2005;21: 676–679.
1390 doi:10.1093/bioinformatics/bti079
1391 93. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX:
1392 Detecting relaxed selection in a phylogenetic framework. Mol Biol Evol.
1393 2014;32: 1–13. doi:10.1093/molbev/msu400
1394 94. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R,
1395 et al. Redefine statistical significance. Nat Hum Behav. 2017;
1396 doi:10.1038/s41562-017-0189-z
1397 95. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al.
1398 The GeneMANIA prediction server: biological network integration for gene
1399 prioritization and predicting gene function. Weaver J, Anija K, Daspit T,
1400 editors. Nucleic Acids Res. Oxford University Press; 2010;38: W214–W220.
1401 doi:10.1093/nar/gkq537
1402 96. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term
1403 association. Bioinformatics. 2007;23: 257–258.
1404 doi:10.1093/bioinformatics/btl567
1405 97. Webb B, Sali A. Protein structure modeling with MODELLER. Methods Mol

Biol. 2014;1137: 1–15. doi:10.1007/978-1-4939-0366-5_1

98. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; doi:10.1063/1.445869

99. Case DAA, Babin V, Berryman JTT, Betz RMM, Cai Q, Cerutti DSS, et al. AMBER14. University of California, San Francisco. 2014. doi:10.1007/s13398-014-0173-7.2

100. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput. 2008; doi:10.1021/ct700301q

101. Sousa Da Silva AW, Vranken WF. ACPYPE - AnteChamber PYthon Parser interfacE. BMC Res Notes. 2012; doi:10.1186/1756-0500-5-367

102. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys. 1993; doi:10.1063/1.464397

103. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A Linear Constraint Solver for molecular simulations. J Comput Chem. 1997; doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

104. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 1981; doi:10.1063/1.328693