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

Adaptation to climate can drive variation in diversification rates and species richness. Understanding how organisms adapt to extreme environments can therefore 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 influences a species' occurrence in space and time (Soberón 2007) and thus is critically important for both speciation (e.g., ecological speciation through climatic niche divergence) and extinction (e.g., due to climate change). 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 (Lamprecht et al. 2018; Rogora et al. 2018). In this context, high elevations, characterized by a short growing season and low annual minimum and mean temperatures, with high daily fluctuation temperatures (Körner et al. 2011), 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 (Clark and Worland 2008; Parker et al. 2018). 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 (Keeling et al. 2013; Kelley et al. 2014; Macdonald et al. 2016; Cicconardi et al. 2017a).

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. On the one hand, strong positive selection on genes related to adaptation, such as genes involved in metabolic pathways, and on the other hand, a

67 relaxation of previous selecting forces, as the conditions of the previous niche are
68 lacking in the new niche. The latter should lead to a reduced number and/or a
69 different set of genes under purifying selection. Specifically, in the case of high-
70 elevation habitats, some heat-shock proteins (HSPs), necessary for coping with
71 extreme heat, should be under relaxation, because heat resistance is not strongly
72 selected for in the average alpine species. A problem with the hypothesis of these two
73 parallel patterns is that there is no test available for the functional consequences of
74 the loss or gain of a specific gene. There are often many pleiotropic effects for genes
75 that make interpretations difficult, and HSPs, expressed upon exposure to stress or
76 during development and growth, are likely to fall into this category. Therefore, we
77 expect signatures of strong positive selection in metabolic genes, and an increase in
78 the relaxation rates in other genes, without being able to specify which these might
79 be. Identifying such genes will help to set up hypotheses to test in the future. A
80 comparative approach offers the strongest method for testing our hypothesis, by
81 comparing genomes of species closely related to each other, but divergent in terms of
82 adaptation to different environments. However, this approach can be limited by
83 insufficient niche divergence within a group or the number of genomes sequenced.
84 Ants are emerging as a leading system for comparative genomics due to the constantly
85 increasing number of available genomes. This resource offers the opportunity to
86 improve the accuracy of orthology detection (Nygaard et al. 2016), to scan for specific
87 mutations, candidate genes, and patterns of acceleration and relaxation in the ant
88 genomes associated with adaptation to cold habitats.

89 Ants are key species in the Earth's terrestrial ecosystems. They comprise more
90 than 15,442 described species and display an impressive diversity of lifestyles

91 (Hölldobler and Wilson 1990; Bolton 2018; Seifert 2018). Ants are especially notable
92 among insects for their ecological dominance as predators, scavengers, and indirect
93 herbivores. They compose at least one third of the entire insect biomass (Wilson 1990)
94 and colonise all kinds of habitats, including thermobiologically challenging
95 environments. The formicine ant *Melophorus bagoti*, for instance, is active during the
96 hottest periods of the summer day, when air temperatures at ant height exceed 50 ˚C
97 (Christian and Morton 1992). On the other hand, for example, the myrmicine ant
98 *Tetramorium alpestre* inhabits the montane and subalpine belt of the Central and
99 South European mountain systems, with the Alps as its main distribution area (Wagner
100 et al. 2017; Steiner et al. 2010). This species lives mainly between 1300 and 2300 m
101 above sea level (a.s.l.), forages below the ground, and nests are established in cool
102 grassland under stones, in moss, rootage, and dead wood, especially subalpine and
103 alpine grass mats (Seifert 2018). Since *T. alpestre* belongs to a species complex in
104 which no relation between the ecological niche and worker morphology has been
105 detected, searching for adaptive phenotypical differences outside of morphology, for
106 example physiology, has already been suggested (Wagner et al. 2017). This species is
107 also relevant because, among the 79 Palearctic species of the worldwide distributed
108 genus *Tetramorium* (Bolton 2016), it is one of only two species to form colonies with
109 multiple queens (polygynous colonies), showing intermediate states of aggression and
110 a transition from multicoloniality to supercoloniality (Krapf et al. 2018; Seifert 2018).
111 *Tetramorium alpestre* is thus a system well suited for studying both genomic
112 adaptation to high elevation and socio-behavioural evolution.

113 Here, we tested our hypothesis of both increased diversifying and reduced
114 purifying (relaxing) selection resulting from adaptation to an extreme niche using *T.*

115 *alpestre*. We newly sequenced its genome and those of four related *Tetramorium*
116 species with diverging ecological niches: *T. immigrans*, a species of the *T. caespitum*
117 complex to which *T. alpestre* also belongs to, in sympatry with *T. alpestre*, but distinct
118 in its altitudinal and ecological habitat; *T. parvispinum*, restricted to mountain forest
119 habitats of the Austral-Asian and Indo-Malayan subregions (Liu et al. 2015); and *T.*
120 *bicarinatum* and *T. simillimum*, invasive generalist species which occur in warmer
121 habitats (Bertelsmeier et al. 2017; Guénard et al. 2017). Our goals were to define the
122 ecological niche of these ant species based on environmental data, and subsequently
123 to perform gene family expansion/contraction and protein-coding scans for signatures
124 of diversifying and relaxing selection. We also annotated the five HSP subfamilies for
125 19 ant species to test for possible shifts in selection acting on these gene families.
126 Among the signatures of diversifying selection detected, we performed two
127 experiments in *T. alpestre* and its relative, *T. immigrans*: we assessed (i) chill-shock
128 triggered gene-expression patterns of *starvin* (*stv*), a Hsp70 modulator, and certain
129 HSPs involved in recovering from extreme cold, and (ii) the temperature dependence
130 of the enzyme activity of Hexokinase type 1 (Hex-T1), a rate-limiting and regulatory
131 enzyme that controls the entry of glucose into the glycolytic pathway, one of the most
132 conserved and essential hexokinase isoenzymes (Jayakumar et al. 2007). The genomic
133 data presented here contribute to a foundation for studying insect genome evolution,
134 particularly also in the light of climate change.

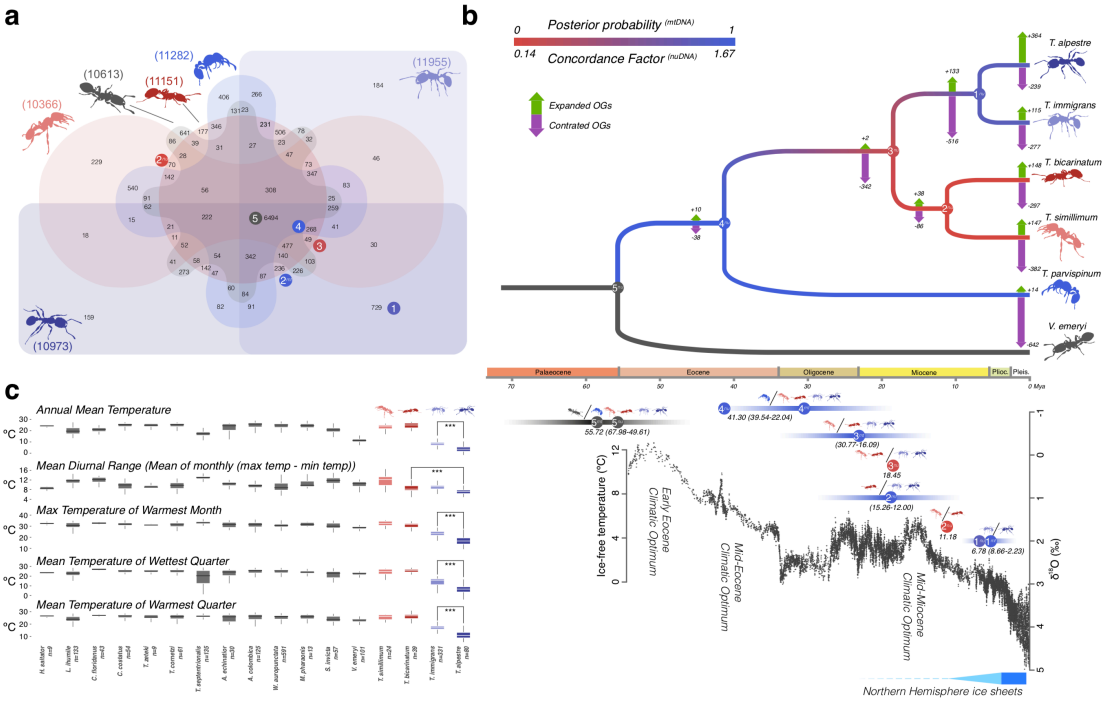

Figure 1. *a*) Venn diagram displaying overlap in orthologous genes in the five *Tetramorium* spp. + *V. emeryi* (Crematogastrini). Coloured circles with number represent phylogenetic splits in the nuDNA (N^{nu}) and mitochondrial phylogenies in *b*; *b*) The dated Crematogastrini nuclear phylogeny (above). 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

A dataset contains 1835 localities representing 17 species was assembled, 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) differentiated between *T. alpestre* and other ants (adjusted *P*-values < 0.001) (Figure 1a). In detail, *T. immigrans* and *T. alpestre* occupy habitats that

155 are colder in the growing season than those of the remaining species, with *T. alpestre*
156 revealing even 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, standard
160 short paired-end libraries, and mate pair libraries resulted in an overall complete
161 assembly of the five *de novo Tetramorium* genomes, at the level of both contiguity
162 and scaffolding. In particular, the *T. alpestre* genome contiguity (contig N50) is the 3rd
163 best among all ant genomes available (Table S2). The analysis of the *k*-mer frequency
164 distribution generated was unimodal (Figure S1; Supplementary Material online), and
165 an analysis of the distribution of GC content across the five *de novo Tetramorium*
166 genomes revealed similar distributions with a mean of 38%, that is, slightly less than
167 that found in *V. emeryi* (42%) (Figure S2). No significant bias was recovered
168 considering GC content and coverage (Figure S3). All five *de novo* draft assemblies
169 were ~241 Mb in length, comparable in size with the average of all ant genomes
170 available (~278 Mb) (Table S2). The total *T. alpestre* genome size was 245.72 Mb,
171 including 18 Mb of gaps and unknown characters (N/X), slightly smaller than the
172 estimated genome size by flow-cytometry of 291.84 Mb +/- 1.76 Mb (n=12). Repetitive
173 elements made up similar proportions of each sequenced genome (*T. alpestre* = 18%,
174 *T. immigrans* = 20%, *T. bicarinatum* = 21%, *T. simillimum* = 20%, *T. parvispinum* = 23%;
175 Table S2), indicating a strong correlation between genome size and total interspersed
176 repeat content (Pearson ρ = 0.976, *P*-value < 0.005; Figure S5). For the gene
177 annotations in *T. alpestre*, a combinatorial approach of unsupervised RNA-seq-based,
178 homology-based, *ab initio*, and finally *de novo* methods were used. By doing so, we

179 approached the number of predicted protein-coding genes in the phylogenetically
180 closest annotated species (15,085 genes predicted vs. 14,872 in the *V. emeryi* genome,
181 Table S2). This improved the homology-based annotation of the remaining four
182 *Tetramorium* genome assemblies that did not differ much in terms of estimated gene
183 content, ranging from 16k to 15k annotated loci (Table S2). BUSCO analyses estimated
184 overall good representation, with recovered genes in between the 97.6% and 99.9%
185 (Table S2; Figure S6), and all statistics on mRNA, coding regions (CDS), exon, and intron
186 length gave highly similar and overlapping distributions among all *Tetramorium*
187 species and *D. melanogaster* as a reference (Figure S7). Also, the transcript
188 completeness gave good results with a comparable distribution of percentage of
189 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. emeryi*
193 resulted in 6494 OGs; among *Tetramorium*, the highest fraction was recovered
194 between *T. alpestre* and *T. immigrans* with 729 OGs, followed by *T. simillimum* and *T.*
195 *parvispinum* (540 OGs) (Figure 1a). For all *Tetramorium* species., the fraction of
196 orthologous genes with at least a match with one of the three outgroups (*A. mellifera*,
197 *N. vitripennis*, *D. melanogaster*) was on average 7286, with the highest in *T.*
198 *alpestre* (7879) and the lowest in *T. parvispinum* (6977), in line with the average value
199 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 assemble
204 the complete sequences. The new mtDNA genomes were slightly longer (on average,
205 17k vs. 16k; Table S2), probably due to the reconstruction of the D-loop and ribosomal
206 genes, which are sometimes harder to assemble. We slightly mis-assembled a coding
207 gene (*nad2*) in the *T. bicarinatum* genome and a few other tRNA loci in other genomes
208 (Table S2). The average GC-content for coding genes was 23.9 ± 3.0% and varied
209 slightly among taxa with the lowest in *Formica fusca* (19.0%) and the highest in
210 *Leptomyrmex pallens* (32.8%) (Table S2). While almost all species had the typical gene
211 order and orientation, some rearrangements were observed. All myrmicine ants were
212 characterized by a translocation of *trnV* after *rrnS*; *Solenopsis* spp. had a swapped
213 position of *rrnS* and *trnN*; *Monomorium pharaonis* had at least three rearrangements:
214 the inversion of *nad6*, *cob*, and *trnS2*, the translocation of *trnE* upstream to *trnA*, and
215 the translocation of *trnR* upstream to *trnE*; *Tetramorium* spp. underwent a swap of
216 *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 (Moreau and Bell 2013). The

226 two methods also converged regarding the date of the most recent common ancestor
227 (MRCA) between *T. alpestre* and *T. immigrans*. In that specific case, PL dated that split
228 at 6.8 million years ago (mya) within the confidence interval given by BI (Median: 5.15
229 mya; 95% 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 of
232 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 topological
234 disagreements and low coalescent units (CU). Within the clade of Crematogastrini
235 (*Tetramorium* spp. + *V. emeryi*), the relation of *T. bicarinatum* and *T. simillimum* was
236 uncertain: nuDNA phylogeny recovered them as sister species with low CU (0.11),
237 while mtDNA phylogeny placed *T. bicarinatum* as sister to the *T. alpestre*-*T. immigrans*
238 cluster, with a bootstrap value of 0.77 and a posterior probability of 0.99.

239

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

242 We examined the evolutionary dynamics of OGs looking at more dynamic ranges
243 corresponding with terminal branches. We observed an increased number of OG
244 deaths in the *T. parvispinum* lineage (– 642) and in the lineage leading to *T. alpestre*
245 and *T. immigrans* (– 516). Ortholog groups showed very little expansion in internal
246 branches of the *Tetramorium* radiation. Expanded OGs were common in terminal
247 branches, except for *T. parvispinum* with very few expanded OGs (+ 14), suggesting a
248 possible lack of annotation. *Tetramorium alpestre* showed the highest number of
249 expanded OGs (+ 364) (Figure 1b), with more than 370 of them (~ 40%) belonging to

250 genes with reverse transcriptase (PFAM: rve, rve_1, rve_2) or transposase and
251 retrotransposon activities (PFAM: Retrotrans_gag, Retrotran_gag_2). Among other
252 PFAM, haemolymph juvenile hormone-binding proteins (PFAM: JHBP; 20 genes), zinc
253 fingers (PFAM: zf-CCHC, zf-H2C2, zf-met, 32 genes), and sugar transporters (PFAM:
254 Sugar_tr, 4 genes) were found.
255

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

258 We investigated the impact of evolutionary pressures on one-to-one orthology in the
259 five *Tetramorium* spp. and 14 other ant species and three outgroups (*Apis mellifera*,
260 *Nasonia vitripennis*, and *Drosophila melanogaster*) by computing both mean ω (d_N/d_S)
261 for terminal branches and relaxing selection (k) within the Crematogastrini.

262 The measure of the mean branch ω showed an overall high level of purifying
263 selection along the whole ant phylogeny, including for short branches (Roux et al.
264 2014; Cicconardi et al. 2017b), with median values of 0.10. However, this distribution
265 was significantly shifted in *T. alpestre*, which showed a median value of 0.38 (Wilcoxon
266 rank-sum test '*greater*', *P*-value < $2.2e^{-16}$), whilst keeping the proportion of $\omega > 1$ in
267 each branch unchanged (Figure 2a, S13). The evaluation of k in the five *Tetramorium*
268 spp. and *V. emeryi* had a mean of median values of 0.99, that is, twice the value in *T.*
269 *alpestre* (median $k = 0.47$; Wilcoxon rank-sum test '*less*', *P*-value < $2.2e^{-16}$) (Figure 2b).
270 We investigated the nature of the increased ω (episodic diversifying selection) and
271 decreased k (relaxed selection) in *T. alpestre* by combining the two measures. Two

evolutionary trajectories appeared to act simultaneously (two tails of the distribution, Figure 2c): One trajectory seems to have led to episodic diversifying selection,

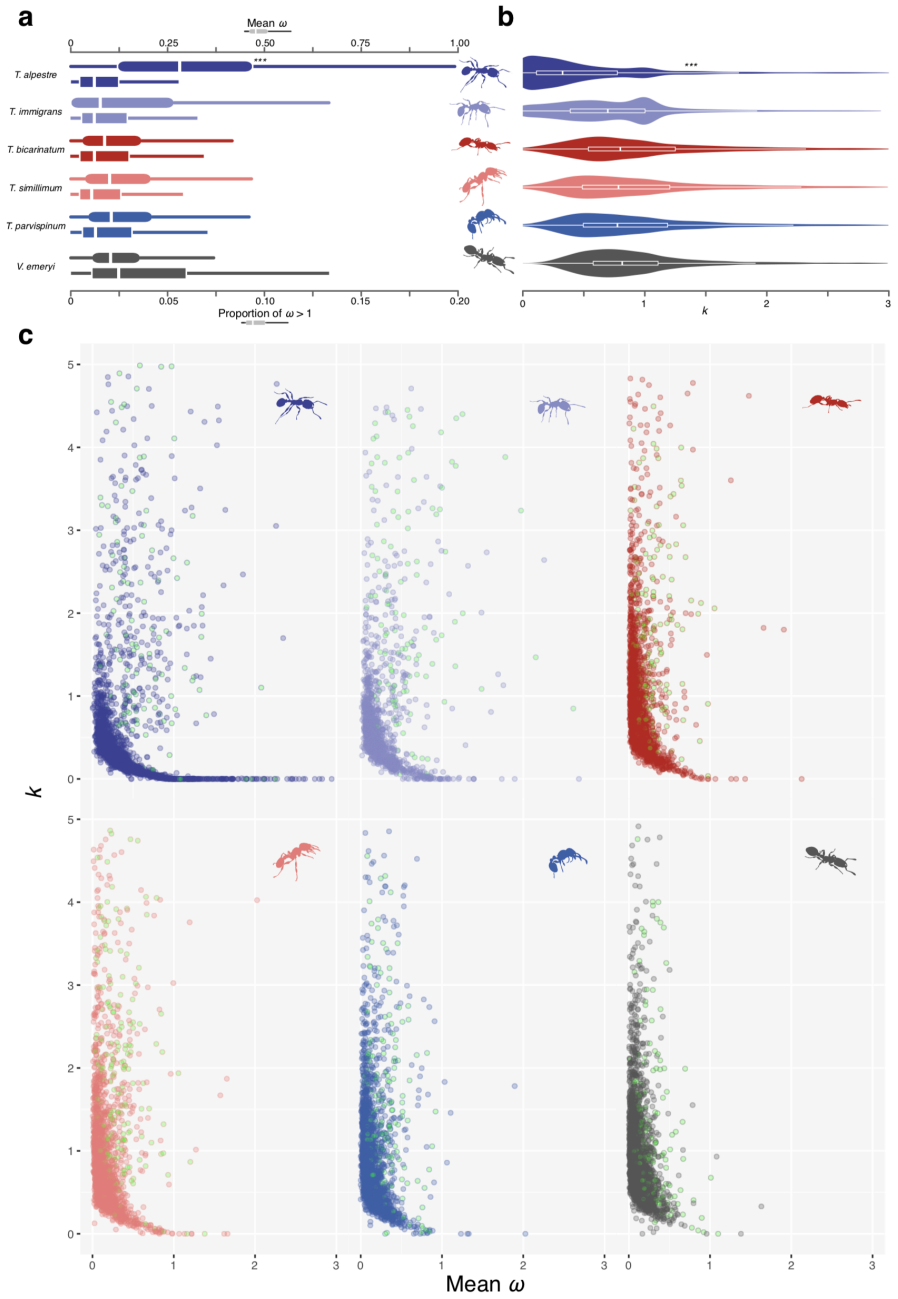

Figure2. *a*) Boxplots showing the distribution and median (vertical line) of mean ω (d_N/d_S) rates (square shapes) in terminal branches of *Tetramorium* spp. and *V. emeryi* (Crematogastrini) in scOGs, and the proportion of genes for which ω is higher than one. *b*) Boxplots overlapped by violin plots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

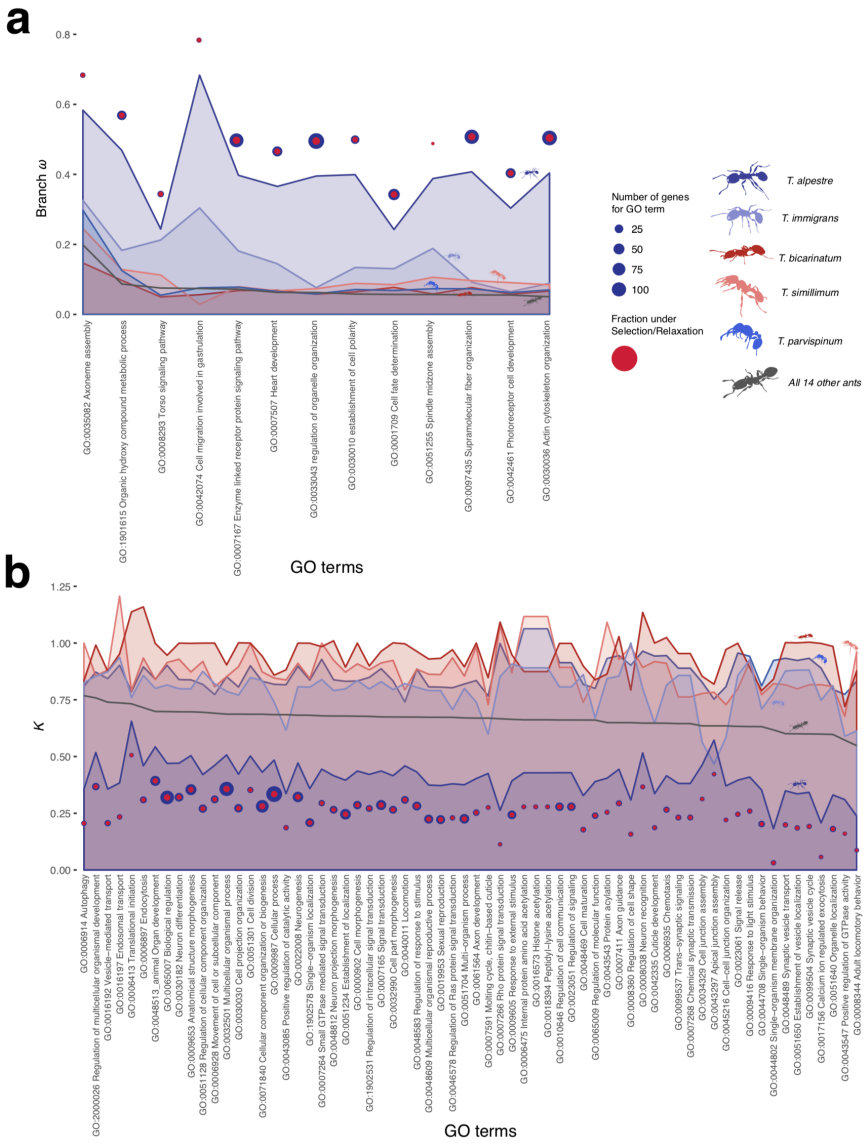

Figure 3. For each enriched Gene Ontology (GO) term category *a*) the median values of *ω*; and *b*) *k* for genes enriching the specific category. *Tetramorium* spp. are colour coded, the other 14 species, merged to obtain a background signal, are in gray. Blue circles are scOGs tested for each GO term, 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 metabolic processes of galactose metabolism and the glycolysis pathway, the pentose phosphate pathway, and galactose metabolism, with five enzymes under diversifying selection: a glucose-6-phosphate (*CG9008*), the hexokinase type 1 (*Hex-T1*), a phosphofructokinase (*Pfk*), a transketolase (*CG8036*), and a phosphoglucosemutase (*pgm*).

321 132 scOGs had a signature of putative relaxation (*P*-values < 0.005; Table S8).
322 While the absolute number of scOGs under relaxation was slightly lower than the
323 number of genes under diversifying selection, the gene-set enrichment tests revealed
324 many more biological processes significantly enriched: 70 terms (*P*-values < 0.005;
325 Figure 3b, Table S9) enriched by 118 genes. Fifteen terms were related to the
326 regulation of cellular development, signal transduction, and cell communication (e.g.,
327 regulation of response to stimuli, regulation of cell communication, regulation of Ras
328 signal transduction), 12 terms related to neurogenesis, axon development, and
329 synaptic transmission (e.g., neuron differentiation, trans-synaptic signalling, axon
330 guidance), seven terms were related to cell morphogenesis and junction organization,
331 and many other related to response to stimuli (e.g., chemotaxis, response to light) or
332 related to adult development and morphogenesis (e.g., anatomical structure
333 morphogenesis, molting cycle, chitin-based cuticle). The enrichment of KEGG
334 pathways revealed three pathways: mRNA surveillance pathway, spliceosome, and
335 protein processing in endoplasmic reticulum (*P*-values < 0.05; Table S10).
336
337 **The Hex-T1 activity, its model structure, and molecular dynamics simulations in *T.***
338 ***alpestre* and *T. immigrans***
339 As we describe in the previous section, the enrichment for KEGG pathways led to the
340 identification of three metabolic pathways, all involved in the metabolism of sugars,
341 with five genes under selection. We therefore aimed at the experimental validation of
342 their functionalities comparing their activity in *T. alpestre* and the most closely related
343 species in this dataset, *T. immigrans*. We tried to assay the four enzymes, but
344 unfortunately, we were able to correctly synthesise only Hex-T1. The assay was

designed to measure the efficiency of Hex-T1 in both species and in a temperature gradient, at 6, 16, 26, and 36 °C. The assays showed that the Km value was always

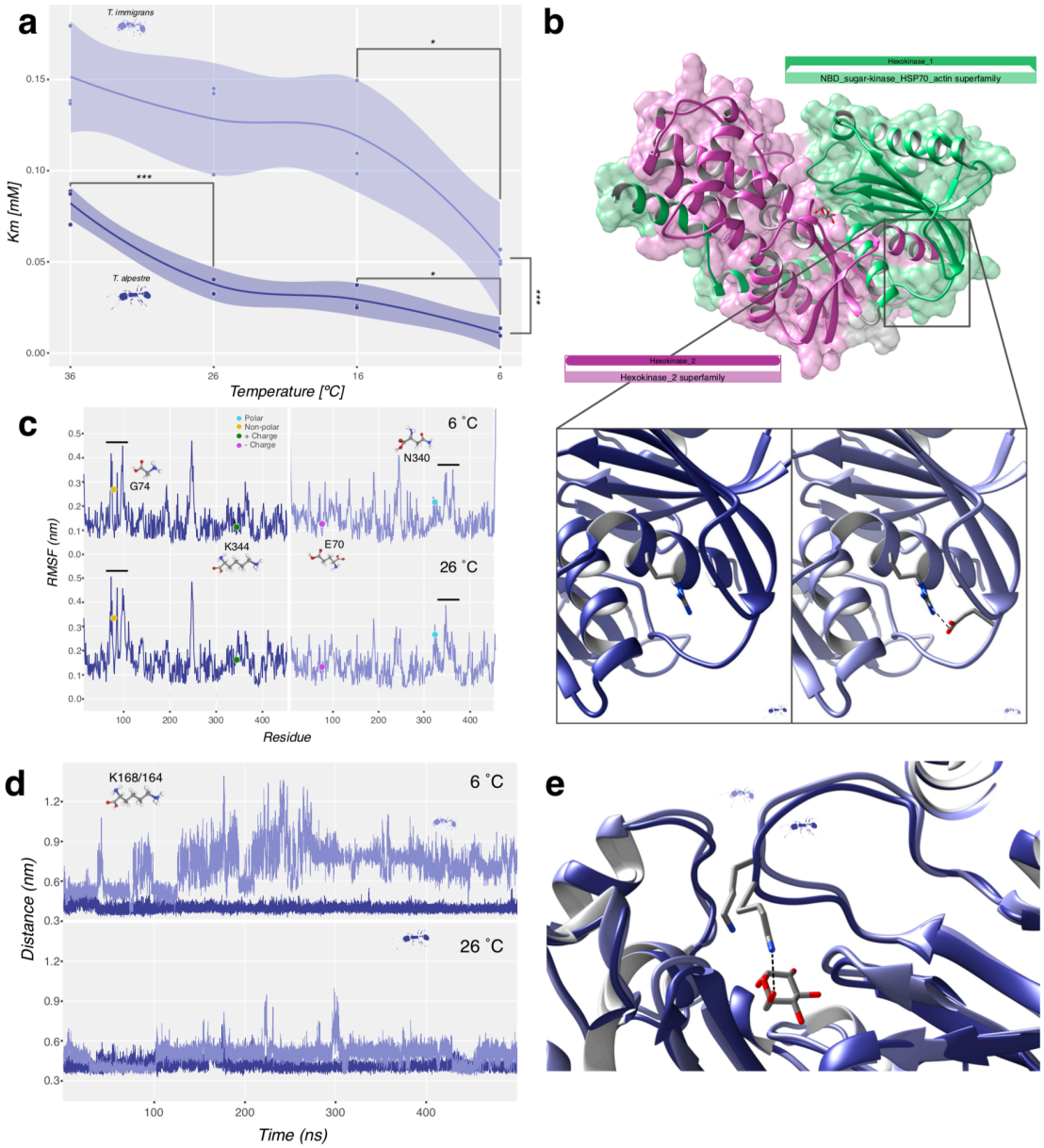

Figure 4. *a*) Scatter plot of Km score across the temperature gradient in the two species, *T. immigrans* on top and *T. alpestre* on the bottom. Lines represent average values with confidence intervals. Significances: *** < 0.001; ** < 0.001; * < 0.05. *b*) Overview of the homology-modeled structure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

361

362 lower in *T. alpestre* than in *T. immigrans* (Wilcoxon rank-sum test 'less', *P*-value <
363 0.00004, Figure 4a), with values ranging from 0.08 mM to 0.01 mM in *T. alpestre*
364 versus 0.15 mM to 0.05 mM in *T. immigrans*, from the highest to the lowest
365 temperature; reflecting that less substrate concentration was needed for the *T.*
366 *alpestre* enzyme to reach half of its maximal reaction speed, that is, Hex-T1 of *T.*
367 *alpestre* was always more efficient than of *T. immigrans*. Although the number of
368 replicates per condition was small, we performed statistical tests to attempt obtaining
369 insight in the enzymatic activity differences within species at different temperatures
370 and in their variance. Interestingly, it seems that for both species there is an
371 improvement in the Km values shifting towards the lowest temperatures (6 °C *vs* 16
372 °C; two-way ANOVA followed by a Tukey's multiple comparison; adjusted *P*-values <
373 0.05), and a possible deterioration at higher temperatures in *T. alpestre*, from
374 0.04±0.01 mM (26 °C) to 0.08±0.01 mM (36 °C) (two-way ANOVA followed by a Tukey's
375 multiple comparison; adjusted *P*-value < 0.0002). An F test (VAR.TEST() implemented in
376 R) to compare the variances of the enzymatic activities (Km values) between the two
377 species – by pooling the replicates at 16 and 26 °C – suggested that there is a possible
378 significant difference in their variance (*P*-value = 0.015), with the confidence interval
379 of Hex-T1 activity in *T. immigrans* twice wider as in *T. alpestre*.

380 The Hex-T1 enzyme is made of two structurally similar domains, Hexokinase 1
381 and 2 (PFAM00349 and PFAM03727), and an N-terminus that is highly variable both in
382 terms of length and amino acidic composition (Figure 4b). The two domains are
383 relatively well conserved across all ants, other Hymenoptera and *D. melanogaster*.
384 Excluding the hypervariable N-terminus, and considering the two functional domains,

385 the amino acid sequences of *T. alpestre* and *T. immigrans* only differ by two amino-
386 acidic changes: a glutamic acid (E) replaced by a glycine (G) in position 74 (TalpHex-
387 $T1_{E74G}$) (Figure 4b inserts) and an asparagine (N) replaced by a lysine (K) in position
388 344 (TalpHex-$T1_{N344K}$). While the second change probably does not entail a significant
389 functionality effect, the *TalpHex-$T1_{E74G}$* instead may be significant. Because of the lack
390 of the long chain and the negative charge, this modification could give the protein a
391 higher flexibility and therefore an overall higher kinetics. We tested this hypothesis by
392 simulating the protein dynamics in the presence of the glucose substrate at different
393 temperatures, i.e. 6 and 26 °C, with the intent of mirroring the condition in the in vitro
394 assay. In the simulations we can see a dual effect. For TalpHex-$T1_{N344K}$, located at the
395 C-terminal of an α-helix forming a helix-turn-helix motif (VSETE**K**DPKG), the Root
396 Mean Square Fluctuation (RMSF) revealed a lower degree of fluctuation in TalpHex-
397 T1 (Figure 4c). This is possibly due to the alternation of positive and negative charges
398 creates a salt-bridges network that stabilizes the whole motif (data not shown). In
399 particular TalpHex-$T1_{K344}$ is involved in a stable interaction with TalpHex-$T1_{E341}$ (data
400 not shown), while TimmHex-$T1_{N340}$ is not available for the formation of salt bridges
401 and mainly establishes hydrogen bonds with TimmHex-$T1_{S336}$ or TimmHex-$T1_{E337}$ at 6
402 or 26 °C, respectively. Although the *T. alpestre* salt bridge brings a higher flexibility of
403 the helix-turn-helix motive, it does not interfere with the catalytic site, given that this
404 motif is far away in the C-terminal region of the protein.

405 In contrast, at the other region (TalpHex-$T1_{E74G}$), the mutation-caused amino
406 acid change allows the loop to explore a broader conformational space equipping it
407 with a higher degree of flexibility already at 6 °C (Figure 4c). Nevertheless, by
408 increasing the temperature, TalpHex-T1 was subjected to an even higher degree of

409 fluctuation, once again confirming the intrinsic flexibility led by the glycine. Notably,
410 the increased flexibility is also transferred to residues relatively far from the mutation,
411 in particular in regions near the active site (168-169), causing a different pattern of
412 interaction with the substrate (Figure 4d, e). Indeed, TalpHex-T1 established a
413 stronger interaction with glucose at both temperatures with respect to TimmHex-T1
414 (Figure 4e).
415
416 **The heat-shock proteins in ants and their evolution in *T. alpestre***
417 Members of the five subfamilies of the heat shock proteins (HSPs: Hsp90s, Hsp70s,
418 Hsp60s, Hsp40s, and sHsps) were fully recovered and characterized in 19 ant species.
419 Overall, a strong purifying selection was observed within subfamilies (Figure 5a, Table
420 S11). In the Hsp90 gene family (*trap1*, *gp93*, and *Hsp83*) *Hsp83* was found duplicated
421 within the lineage of *W. auropunctata* (inparalog), as well as the *T. parvispinum gp93*;
422 duplication that probably occurred before emergence of the lineage (outparalog)
423 (Figure S14). Members of the Hsp70 family were recovered for all species with 11
424 ortholog groups, with few duplications in *Hsc70-3* and *Hsc70-5* (two inparalogs and
425 two outparalogs, respectively). We confirmed the lack of Hsp70s in ants (Nguyen et
426 al. 2016) and a multiple duplication of *Hsc70-4* in Hymenoptera. One paralog seemed
427 to be lost in many species while the other two copies were still present in almost all
428 species, with the presence of other outparalogs in various species, mostly in Attini
429 (Figure S15). The chaperonin Cpn60/TCP-1 family (Hsp60) had nine OGs; six inparalogs
430 were found in *CCT7, CCT3* and four outparalogs in *Tcp1a*, *CCT8*. The heat shock protein
431 40 kDa/Chaperone DnaJ was the largest subfamily, with 37 well-defined OGs. Its
432 turnover of duplication was relatively small, and only six OGs showed some level of

433 duplication. The small HSPs (sHsps) showed the expansion of the lethal (2) essential
434 for life (*l(2)efl*) gene paralog of *D. melanogaster*, with five distinct OGs in all
435 Hymenoptera. *L(2)efl3* seemed to be retained by only in 11 species, and with *l(2)efl4*,
436 they are the two most divergent OGs, with ω equal to 0.23 and 0.33, respectively. The
437 less divergent were *Hspb1* and *l(2)efl5*, for which ω was 0.11 and 0.07, respectively.
438 Overall, the sHsps showed a significantly higher distribution of ω (median = 0.22, *P*-
439 value = 0.013), followed by Hsp40s, which also showed a wide distribution of ω. In
440 contrast, Hsp90s, Hsp70s, and Hsp60s seemed to be very stable and under a very high
441 purifying selection.

442 To understand other possible mechanisms of adaptation of *T. alpestre* to the
443 cold, we computed the mean values of ω and k for all branches leading to
444 *Tetramorium* spp. and *V. emeryi*, integrating 13 ant species, and scanned for
445 diversifying and relaxed selection specifically in *T. alpestre*. The first part of the
446 analysis showed a similar landscape as observed in scOGs. Among the five
447 *Tetramorium* species and *V. emeryi*, five species display a distribution of the mean-
448 branch ω ranging between 0.07 (*T. immigrans*) and 0.13 (*T. parvispinum*); *T. alpestre*
449 instead showed significant 2.7-fold higher value of the median mean-branch ω (ω =
450 0.36, Wilcoxon rank-sum test '*greater*', adjusted *P*-values < 0.004, Figure 5b). The
451 distribution of k among HSPs in the six species showed median values of k all close to
452 the expected (k = 1), between 0.90 in *V. emeryi* and 1.05 in *T. simillimum*, again with
453 the exception of *T. alpestre*, where the median k = 0.63, 1.4-fold lower than the lowest
454 median k (Wilcoxon rank-sum test '*less*', adjusted *P*-values < 0.05, Figure 5c). The scan
455 for diversifying positive and relaxed selection in branches of the HSPs of *T. alpestre*

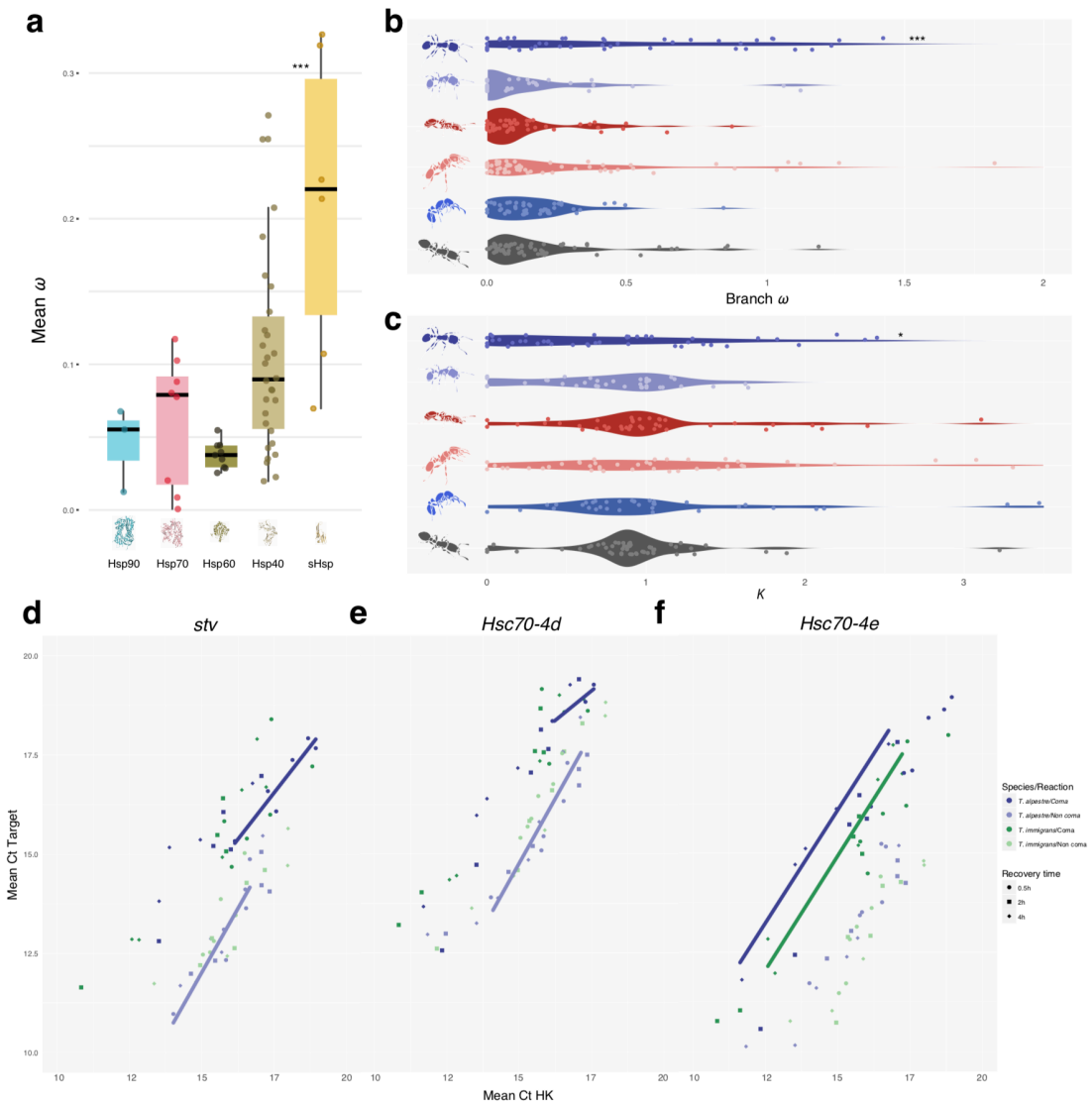

Figure 5. *a*) Boxplots showing the distribution and median (horizontal black line) of mean ω (d_N/d_S) rates for each of the HSP subfamily. *b*) Violin plot showing the distribution of the mean ω values; and *c*) *k* in the terminal branches of the Crematogastrini (from up to bottom: *T. alpestre*, *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*).

showed no gene under putative diversifying positive selection, while six loci were identified to be under putative relaxing selection: one Hsp90 (*Trap1*), two Hsp60 (*CCT4*, *CCT5*), and three Hsp40s (*Sec63*, *DnajC8*, *DnajC11*) ($k < 0.14$, adjusted *P*-values < 0.005). Using these genes in prediction interaction networks, we found that the two

472 Hsp60s interacted with other 21 genes, giving 17 significant enriched functions (FDR
473 < 0.05), mainly related to protein folding and microtubule, spindle organization; the
474 three Hsp40s interacted with 21 genes, giving two enriched functions (FDR < 0.05),
475 related to response to heat and response to temperature stimulus (Figure S19).
476
477 **Chill shock effect, recovery times, and expression patterns of *stv* and HSPs in *T.***
478 ***alpestre* and *T. immigrans***
479 Among the genes with a putative signature of diversifying selection, we also found
480 *starvin* (*stv*), a BAG-family member which seems to be implicated in Hsp70 ATPase
481 activity during cold recovery in *D. melanogaster* (Colinet and Hoffmann 2010).
482 Therefore, we explored if the modulation of its expression may play a role in how *T.*
483 *alpestre* differently withstands the cold compared to the related species *T. immigrans.*
484 We thus performed a pilot test measuring the different expression patterns of *stv*, two
485 paralogs of *D. melanogaster Hsc70-4* (*Hsc70-4d* and *Hsc70-4e*), a member of the
486 Hsp90 subfamily (*Hsp83*), and sHsp *l(2)efl4* in the two species.
487 We measured the transcription activity of three recovery times (0.5 h, 2.0 h,
488 and 4.0 h) and compared coma and non-coma individuals and found a significant
489 difference in the expression patterns of *stv*, *Hsc70-4d*, *Hsc70-4e*, and *Hsp83*, but not
490 of *l(2)efl4* (one-way ANCOVA followed by Bonferroni-Holm correction). In particular,
491 looking at the different recovery times, *stv* and *Hsc70-4d* responded quickly in *T.*
492 *alpestre* but not in *T. immigrans,* being significantly expressed only in workers of *T.*
493 *alpestre* after 0.5 h recovery, but not in *T. immigrans* (Figure 5d-e). Comparing the
494 expression levels between species, we found higher mRNA levels for *Hsc70-4e* in *T.*
495 *alpestre* after 4 h recovery (Figure 5f).

496

497

498 ## *Discussion*

499 *Tetramorium alpestre* represents a critical example of evolutionary adaptation to cold
500 environments such as alpine habitats, characterized by low temperatures during the
501 growing season (Figure 1c). By analyzing its genome in the context of 18 other ant
502 species, we shed new light on the evolution and adaptation of this ant, which can
503 serve as a potential model for arthropod adaptation to cold habitats. We found good
504 concordance between the nuclear and the mitochondrial dated phylogenies,
505 concerning the dating of the speciation of *T. alpestre*, which seems to have diverged
506 from its closest relative, *T. immigrans*, between 2 and 9 mya. This period overlaps with
507 the beginning of a new climatic zonation of the European continent during the Middle
508 and earliest Late Miocene (Zachos et al. 2001). At the end of this period, a rapid uplift
509 (~5 mya) fundamentally changed the paleogeographic and topographic setting of
510 central and southern Europe and triggered Alpine glaciation (Kuhlemann 2007).
511 Interestingly, both paleoclimatic and paleogeographic changes co-occurred with the
512 origin of the lineage leading to *T. alpestre* (Figure 1b). In fact, while these glaciations
513 dramatically reshaped global biodiversity patterns, eliminating terrestrial biota from
514 many mid- to high-latitude areas, they may have been the main driver of
515 diversification on highlands by reducing gene flow among populations (Wallis et al.
516 2016).

517 The present results support our prediction that two parallel patterns will occur
518 in the genome of species adapting to extreme habitats. The genome of *T. alpestre*
519 appears to be under the influence of two strong evolutionary trajectories: on the one

520 hand, a diversifying selective force, two-fold higher compared with all other ants, and
521 on the other hand, a relaxation of the overall purifying selection present in other ants,
522 identified by the skewed distribution of *k* towards zero. Specifically, the diversifying
523 selective force affects more than a hundred genes in many key biological processes,
524 such as those related to development, cell migration, and gastrulation. Genes under
525 the category of biological processes are also under selection in the genome of the
526 Antarctic midge *Belgica antarctica* (Kelley et al. 2014). In fact, early stages are the
527 most fragile ones, having the highest rate of lethal phenotypes during heat shock and
528 UV irradiation, particularly around gastrulation (Uchida et al. 2018). The same may
529 pertain to genes related to cellular-level organization (e.g., spindle midzone assembly,
530 supramolecular fiber organization, actin cytoskeleton organization), where
531 environmental stress may jeopardise a well working cell division machinery, as SNPs
532 associated to long-term cold-adaptation plasticity were found to be present in genes
533 related to cytoskeletal and membrane structural components in *D. melanogaster*
534 (Gerken et al. 2015). Among 25 genes involved in these biological processes, *polo*, a
535 polo-like kinase which plays a central role as regulator of cell division and is required
536 for several events of mitosis and cytokinesis (Archambault et al. 2015), shows three
537 amino acid substitutions in positions corresponding to active binding sites, and thus
538 could be a good candidate for further studies.

539 The two evolutionary trajectories were predicted in our hypothesis, and while
540 diversifying selection is somewhat expected in species adapting to a new
541 environmental niche, the magnitude of relaxed selection found is more surprising and
542 never recorded. Although both ω and *k* distributions in *T. alpestre* are highly skewed,
543 the absolute number of genes under diversifying and relaxed selection are not

544 markedly greater than other species in similar studies (Cicconardi et al. 2017b; Roux
545 et al. 2014; Harpur et al. 2014). This cannot be directly ascribed to the type *I* error rate
546 because the same rate should be equally randomly present in all the other short or
547 long branches of the phylogeny. Nevertheless, *T. alpestre*, as part of a species complex
548 (Steiner et al. 2010; Wagner et al. 2017), may have more recently diverged from more
549 closely related species not yet adapted to the alpine habitat. Magnitude of this relaxed
550 selection is evidenced by the 70 enriched biological processes, and can be possibly
551 seen as the consequence of a shift and/or decreased magnitude in the purifying
552 selection. The direct implication of this relaxation is not clear, especially its effect on
553 genes and their biological processes, such as the regulation of cellular development,
554 signal transduction, cell communication, neurogenesis, axon development, and many
555 others. If the physiology of cold adaptation utilises different metabolic and
556 developmental strategies to minimise and optimise energy consumption, a new
557 balance needs to be reached. The two patterns observed in present study
558 (diversifying/relaxing selection) might represent this new shift. Interestingly, *T.*
559 *immigrans*, which inhabit intermediate environmental conditions (Figure 1c), shows
560 also an intermediate pattern of *k*, with an intriguing bimodal distribution (Figure 2b).
561 Therefore, we can speculate that the highly skewed distributions point to ongoing
562 adaptation to a colder climate, and a relaxation of selective forces present in warmer
563 habitats. To our knowledge, this is the first time that this effect is found in an
564 organism, most likely due to the underrepresentation of genomic studies on alpine
565 and cold adapted species, and, importantly, the lack of application of the RELAXED
566 test in a genome-wise manner.

567 Examining the effect of those evolutionary trajectories, metabolism seems to
568 be significantly involved in the sub-alpine ant`s adaptation. Several clues are pointing
569 in that direction, such as the presence of underground aphids in *T. alpestre* nests but
570 not in those of *T. immigrans* (unpublished data), the expansion of OGs related to sugar
571 transporters, and five enzymes under diversifying selection related to metabolism and
572 phosphorylation of sugars. We deeply explored and validated one of these enzymes,
573 Hex-T1, a key regulator and rate-limiting enzyme for energy (sugars) metabolism and
574 reactive oxygen species (ROS) activity in insects (Xian-Wu and Wei-Hua 2016), by
575 contrasting its enzymatic activity between *T. alpestre* and *T. immigrans.* We showed
576 that the mutated form in *T. alpestre* is indeed overall more efficient, with less
577 substrate needed to reach the highest reaction velocity of the enzyme, possibly due
578 to a single amino acid mutation that enhance effect on the catalytic efficiency of the
579 enzyme due to the formation of a more functional network of hydrogen bonds with
580 its substrate. In fact, the path of the hydrogen bonds found in TalpHex-T1 resembled
581 the one depicted by the X-ray solved structure of Hex-T1 from (Mulichak et al. 2002),
582 used as a template to obtain the three-dimensional coordinates of both TimmHex-T1
583 and TalpHex-T1 (see Methods). Mulichak and colleagues identified their model
584 residues Lys621 and Asp657 as crucial for the catalytic activity of the protein; the
585 corresponding residues TalpHex-T1$_{K168}$ and TalpHex-T1$_{D204}$ are indeed interacting,
586 together with other residues, with glucose in TalpHex-T1 at both 6 and 26 °C. On the
587 other hand, while the interaction with TalpHex-T1$_{D204}$ is conserved, the one with the
588 lysine is lost in TimmHex-T1 and a looser protein-sugar network is formed, explaining
589 the higher catalytic efficiency of TalpHex-T1 versus that of TimmHex-T1. We can
590 indeed infer that the higher flexibility conferred by TalpHex-T1$_{G74}$ increases the shift

591 of the residue TalpHex-T1$_{K168}$, belonging to the catalytic site (Figure 4d, e), allowing it
592 to move closer and establish a stronger interaction with glucose, possibly contributing
593 to lower the Km of the catalytic reaction. In eco-evolutionary terms all these findings
594 hint at a more intimate relation between *T. alpestre* and the aphids. This relationship
595 of ants and aphids is one of the most studied mutualistic relationships in the animal
596 kingdom. Aphids produce honeydew, and in return ants offer protection from
597 predators, parasitoids, fungal infection and adverse conditions. Honeydew contains
598 monosaccharides (glucose and fructose), disaccharides (maltose, sucrose), and
599 Melezitose, a trisaccharide, that is one of the main sugar synthesized by aphids, and
600 which has been found to attract ants and maintain the ant-aphid mutualism (Fischer
601 and Shingleton 2001). Small amounts of amino-acids, proteins, and lipids are also
602 present (Pringle et al. 2014). Usually, this interaction is above ground level, on plant
603 leaves and branches, but there are cases in which ants are specialized to farm
604 subterranean aphids, just like in *Lasius flavus* (Ivens et al. 2012). *Tetramorium alpestre*
605 may have evolved the same behavior through a coevolution of physiological factors
606 such as palatability, fluid intake rate and digestibility of sugar molecules, improving
607 the ants' survival by enhancing energy intake, a limiting factor in high-elevation
608 ecosystems.

609 The complete annotation and characterization of the five HSP subfamilies in 19
610 ant species show good overall conservation, and strong purifying selection was
611 observed in most subfamilies. The exception are sHsps, which seem to be the most
612 diversified HSP subfamily, with the highest and the widest range of ω. It is also the
613 only subfamily that took a different evolutionary path compared with Diptera, having
614 very few orthologs in common with them. Given that sHsps are virtually ubiquitous

615 molecular chaperones that can prevent the irreversible aggregation of denaturing
616 proteins (Haslbeck and Vierling 2015), they may have played a central role in the
617 adaptation of ants to strongly different climates and environments. Although we
618 observe a higher distribution of ω in *T. alpestre*, no HSPs show a significant sign of
619 diversifying selection. Rather, six loci show relaxed purifying selection (Hsp90: *Trap1*;
620 Hsp60:*CCT4*, *CCT5*; Hsp40: *Sec63*, *DnajC8*, *DnajC11*). The impact of the TNF receptor-
621 associated protein (Trap1) on cellular bioenergetics of *T. alpestre* may contribute to
622 its adaptation to cold environments by differently modulating cellular metabolism.
623 Given that it functions as a negative regulator of mitochondrial respiration, able to
624 modulate the balance between oxidative phosphorylation and aerobic glycolysis,
625 possibly another form of metabolic adaptation. It has been shown that reduced or
626 absent *Trap1* expression leads to deregulated mitochondrial respiration, that is, a high
627 energy state characterized by elevated ATP and reactive oxygen species, an increase
628 in mitochondrial respiration and fatty acid oxidation, and a cellular accumulation of
629 tricarboxylic acid cycle intermediates (Yoshida et al. 2013). *Trap1* overexpression is
630 associated with increased expression of genes associated with cell proliferation (Liu et
631 al. 2010a). Since the two Hsp60s are associated to protein folding and cytoplasmic
632 microtubule organization, biological processes also enriched by genes under
633 diversifying selection, this suggests profound changes in the cellular physiology of *T.*
634 *alpestre*. It is also plausible that adapting to a cold environment will depress selection
635 on genes directly related to severe heat shock. As we have shown, *T. alpestre*'s habitat
636 is characterised by temperatures that rarely exceed 20 °C (median = 16.5 °C; 99%
637 confidence interval = 17.8 °C), although rocky habitats may produce warmer
638 microhabitats. This could mean that this species experienced relaxed selection for

639 HSPs associated with the denaturation to protein folding by heat shock, as suggested
640 by the relaxation of purifying selection on *Cct5* and the three Hsp40s associated with
641 responses to heat and temperature stimulus.

642 Although we provide a range of evidence that adaptation to cold environments
643 can be related to changes in protein-coding sequences, it has long been postulated
644 that phenotypic divergence between closely related species, such as *T. alpestre* and *T.*
645 *immigrans*, is primarily driven by quantitative and spatiotemporal changes in gene
646 expression, mediated by alterations in regulatory elements (e.g.: Danko et al., 2018;
647 Prescott et al., 2015). We tested this hypothesis by looking at the expression patterns
648 of *stv* and four other HSPs (*Hsc70-4d*, *Hsc70-4e*, *Hsp83*, *l(2)efl4*), between *T. alpestre*
649 and *T. immigrans*. *Stv* is a member of Bcl-2-associated athanogene (BAG)-family
650 proteins, which interact with Hsc70s and Hsp70s, and can modulate, either positively
651 or negatively, the functions of these chaperones (Doong et al. 2002). It is implicated
652 in the recovery from chill shock in *D. melanogaster* during cold recovery (Colinet and
653 Hoffmann 2010, 2012). The concerted synthesis of *D. melanogaster stv* and Hsp70s
654 suggests cooperation to offset cold injury, possibly by preserving the
655 folding/degradation and by regulating apoptosis (Colinet and Hoffmann 2010). In our
656 experiment, an acute cold stress (-6 °C) was used to knock down individuals. By looking
657 at the differential expression of these genes in coma and non-coma individuals within
658 and between species, we were not only able to show that *stv* may interact with
659 Hsc70s, but also that *T. alpestre* has a quicker response in these genes compared with,
660 *T. immigrans,* the related species not cold adapted. Given these results, which are in
661 line with the selection signature we identified, it is probable that *stv* may play a similar
662 role in the physiology of ants. As we have shown in this study, ants completely lack

663 Hsp70s (Figure S14) and retained only the heat shock cognates (Hsc70s) with mainly
664 two conserved paralogs (*Hsc70-4d* and *Hsc70-4e*), which may have evolved to function
665 as the Dipteran Hsp70s. Although we found an interesting and promising correlation
666 between chill shock, *stv*, and Hsc70s, it will be important to continue exploring other
667 expression patterns in ant physiology to assess whether *stv* has tissue-specific
668 expression patterns, as was found in *D. melanogaster* (Coulson et al. 2005), and to
669 study its expression across different species with different ecological adaptations.

670 One of the most enduring problems of evolutionary biology is explaining how
671 complex adaptive traits originate (Wagner and Lynch 2010). Although it is widely
672 assumed that new traits arise solely from selection on genetic variation, many
673 researchers ask whether phenotypic plasticity precedes and facilitates adaptation
674 (Levis and Pfennig 2016). This alternative scenario, the plasticity-first hypothesis,
675 states that under novel conditions, environmentally induced variation can be refined
676 by selection and, depending on whether plasticity is favoured, become
677 developmentally canalized through genetic assimilation. Under this scenario,
678 environmentally induced phenotypic change can precede and promote the
679 evolutionary origins of a complex adaptive trait.

680 Four criteria have been proposed for verifying this hypothesis (Levis and
681 Pfennig 2016): 1) the focal trait needs to be environmentally induced in the ancestral-
682 proxy lineages; 2) cryptic genetic variation will be uncovered when ancestral-proxy
683 lineages experience the derived environment; 3) the focal trait will exhibit evidence
684 of having undergone an evolutionary change in its regulation, in its form, or in both in
685 the derived lineages; and 4) the focal trait will exhibit evidence of having undergone
686 adaptive refinement in the derived lineages. Although it is methodically difficult to

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

699 Our comparative genomic analysis shows how natural selection could trigger
700 complicated patterns of change in genomes, especially protein-coding sequences,
701 both in terms of diversifying selection and relaxation of purifying selection. Many of
702 these changes are in genes that may be associated with aspects of development,
703 either directly or through the associated complex changes in ecology and natural
704 history, and all of these hypotheses are now open for in-depth scrutiny. Furthermore,
705 this work identifies promising gene expression patterns that are possibly playing an
706 important role in adaptation to new ecological niches. Overall, this study represents
707 the first systematic attempt to provide a framework for the genomic analysis of
708 adaptation to extreme environments and underlines the importance of studying
709 organisms with different ecological niches in understanding the genetic basis of
710 ecological adaptation.

711

712

713 ## *Methods*

714 **Species-locality and climatic data**

715 To define the environment of *T. alpestre* and the other four *Tetramorium* species,
716 georeferenced occurrence data were compiled and complemented with those for the
717 native ranges of other ant species with available genomic data. Occurrence records
718 for 17 species (Table S2) were compiled from the Global Ant Biodiversity Informatics
719 database (GABI, Guénard et al., 2017, accessible via www.antmaps.org, Janicki et al.,
720 2016). The species determinations in the records were evaluated by specialists in ant
721 taxonomy, and erroneous data were excluded. Slightly incorrect coordinates (i.e. in
722 the ocean within 8.5 km of land) were assigned to the nearest land point. Additionally,
723 georeferenced occurrence data of two species (*T. alpestre, T. immigrans*) were
724 compiled from the latest taxonomic revision (Wagner et al. 2017). The occurrence
725 data of each species were spatially rarefied at a 5 km level using the R package SPTHIN
726 v. 0.1.0 (Aiello-Lammens et al. 2015), to counter spatial autocorrelation and to provide
727 some protection against sampling bias issues (Merow et al. 2013). The Bioclim
728 variables bio1-bio19 from the WORLDCLIM database v. 1.4 (Hijmans et al. 2005) were
729 extracted for each unique species locality at a 2.5 arc min resolution (4.65 x 4.65 =
730 21.6225 km^2 at the equator) using ARCGIS v. 10.4 (ESRI, Redland, CA). A Wilcoxon-
731 Mann-Whitney rank sum test as implemented in the R package WILCOX.TEST
732 (http://www.r-project.org) was adopted to test for differences between *T. alpestre*
733 and the other species individually for each of the Bioclim variables. The Bonferroni-

734 Holm sequential rejection procedure (Abdi 2010) was used to control for multiple
735 testing.
736
737 **Biological sample processing, genome size estimation, and genome sequencing**
738 Six colonies of *T. alpestre* ants were sampled, two for flow cytometry (nest 18811:
739 Penser Joch, Italy, 46.83379° N, 11.44652° E, 21 June 2016; nest 18813: Kühtai,
740 Austria, 47.27923, 11.07590° E, 21 June 2016), three for RNA extraction (nest 18586:
741 Zirmbachalm, Austria, 47.21920° N, 11.07530° E, 30 July 2015; nest 18590:
742 Zirmbachalm, 47.21910° N, 11.07620° E, 30 July 2015; nest 18594: Penser Joch,
743 46.82940°N, 11.43830° E, 2 August 2015), and one for DNA extraction (nest 18592:
744 Penser Joch, 46.83380° N, 11.44610° E, 2 August 2015). The ants were brought alive
745 to the laboratory (for flow cytometry) or immediately submerged in RNAlater (for RNA
746 extraction) or 96% ethanol (for DNA extraction) and stored at -70 °C (RNA) or -20 °C
747 (DNA).
748 Flow cytometry was used to estimate relative genome size from single *T.*
749 *alpestre* worker heads, using strain ISO-1 of *Drosophila melanogaster* as internal
750 standard (obtained from Bloomington Drosophila Stock Center, Indiana University,
751 Bloomington, USA; nuclear DNA content: 0.35 pg, Gregory 2018). Six individuals each
752 were measured from *T. alpestre* nests 18811 and 18813. In more detail, single *T.*
753 *alpestre* heads and, separately, heads of 10 ISO-1 females were chopped in 500 µL of
754 ice-cold Otto 1 Buffer (0.1 mol/L citric acid, 0.5% Tween20 (Merck KGaA, Darmstadt,
755 Germany)). The suspensions were filtered using 42-µm nylon meshes and incubated
756 for 5 min. Then, 1 mL of staining solution (DAPI (4 µL/mL) and 2-mercaptoethanol (2
757 µL/mL) in Otto 2 buffer (0.4 mol/L $Na_2HPO_4 \cdot 12$ H2O)) was added to each suspension.

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

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

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

782 size ~500 bp, 2×200 bp read length) were constructed from three stages of workers
783 (larvae, pupae, adult), two stages of gynes (larvae and adult), and adult males. All
784 libraries were sequenced on an Illumina HiSeq 2500 platform in standard (DNA) and
785 RAPID (RNA) mode. For each *T. immigrans*, *T. bicarinatum*, *T. simillimum*, and *T.*
786 *parvispinum*, one TrueSeq library (insert size ~500 bp, 2×100 bp read length) was
787 prepared and sequenced on an Illumina HiSeq 4000 machine at OIST sequencing
788 center.
789
790 **Genomic and mitogenomic assembly and gene prediction**
791 Genomic sequencing reads of *T. alpestre* were filtered to remove low-quality reads
792 and PCR duplicates, corrected, and assembled using ALLPATHS-LG v. 52488 (Gnerre et
793 al. 2011). Contigs were first constructed based on overlapping reads and then
794 scaffolded using paired-end and mate-pair information from all DNA libraries. Given
795 the high coverage of the overlapping reads, based on the estimated genome size,
796 multiple trials were performed using different coverages to establish the best
797 combination of read coverage. The genomes of the other four related species were
798 assembled using SPADES v. 3.10.0 (Bankevich et al. 2012) with enabled read error
799 correction algorithm, a *k*mer range of 21, 33, 55, 77, 99, and mismatch correction.
800 Scaffolds belonging to contaminants were identified and removed using BLOBOLOGY v.
801 2013-09-12 (Kumar et al. 2013).
802 A bioinformatic pipeline that included combinatorial approaches of homology,
803 *ab initio*, and *de novo* methods were implemented to predict gene models.
804 REPEATMASKER v. 4.0.6 (Smit et al. 2013) was used to generate repeat hints. For the
805 homology-based exon hints, protein data sets of *D. melanogaster*, *Apis mellifera*,

806 *Nasonia vitripennis*, and 12 ant species (see Table S1) were aligned to the *T. alpestre*
807 genome using TBLASTN (*e*-value cut-off: $1e^{-5}$) and EXONERATE v. 2.2.0 (Slater and Birney
808 2005). As an *ab initio* procedure, the merged RNA-seq libraries were mapped using
809 STAR v. 020201 (Dobin et al. 2013) (settings: ALIGNENDSTYPE Local; ALIGNINTRONMIN 30;
810 ALIGNINTRONMAX 450,000; OUTSJFILTERINTRONMAXVSREADN 80 100 500 1000 2000 5000
811 20,000; ALIGNSJOVERHANGMIN 10; ALIGNTRANSCRIPTSPERREADNMAX 100,000) to generate
812 the splice-site position-specific scoring matrix (PSSM) and CUFFLINKS v. v2.2.1 (Trapnell
813 et al. 2012) to generate UTR and intron hints. This information was then integrated
814 with GENEMARK-ES v. 4 (Hoff et al. 2015) to train the analysis and use the outcome
815 under the BRAKER1 procedure (Hoff et al. 2015) and AUGUSTUS v. 3.2.1. (Stanke et al.
816 2006). To avoid missing genes due to unassembled regions of the genome, previously
817 unmapped reads were *de novo* assembled using TRINITY v. 2.2.0 (Grabherr et al. 2011)
818 and annotated with FRAMA (Bens et al. 2016). The three results where then merged,
819 and Benchmarking Universal Single Copy Orthologs (BUSCO v. 1.2) (Simão et al. 2015)
820 was adopted to evaluate the annotation completeness together with integrity and
821 completeness evaluation using DELTABLAST (Boratyn et al. 2012). A visual inspection of
822 the annotation and gene integrity was also performed on more than 250 loci.
823 All available complete ant mitochondrial genomes (mtDNA) were downloaded
824 from GENBANK (Benson et al. 2014) (Table S1). Illumina raw SRA data of other ant
825 species were downloaded and processed following Cicconardi et al., (2017b). In brief,
826 reads were quality filtered and assembled using IDBA-UD v. 1.1.1 (Peng et al. 2012)
827 and SPADES v. 3.10.0 (Bankevich et al. 2012), and annotated using MITOS v. 2 web
828 server (Bernt et al. 2013).
829

830 **Phylogenetic analysis**

831 Nuclear single-copy ortholog genes (scOGs) and complete mitochondrial genomes
832 were used to compute the ant species tree. Once scOGs were determined (see below),
833 single-locus trees and a species tree of concatenated scOGs were estimated using
834 Maximum Likelihood (ML) search as implemented in FastTree v. 2.1.8 SSE3 (Price et al.
835 2010). Gene trees were summarized with MP-EST (Liu et al. 2010b) using the STRAW
836 web-server (Shaw et al. 2013).

837 For the mtDNA phylogeny, only protein-coding genes were used. Each gene
838 was aligned using MACSE v. 1.01b (Ranwez et al. 2011), and all alignments were
839 concatenated. The best partitions and models of evolution were identified with
840 PartitionFinder v. 1.1.1_Mac (Lanfear et al. 2014). MtDNA trees were searched with
841 ML and Bayesian Inference (BI) algorithms. For ML, GARLI v. 2.01.1067
842 (http://code.google.com/p/garli) was used performing 20 + 5 runs from random
843 starting trees. The runs were continued until no further improvement in log-likelihood
844 was found, followed by 1000 bootstrap pseudoreplicates. The results were
845 summarized using SumTrees v. 3.3.1 (http://bit.ly/DendroPy), using *A. mellifera*, *N.*
846 *vitripennis*, and *D. melanogaster* as outgroups. For BI, BEAST v. 2 (Bouckaert et al.
847 2014) was adopted to obtain the dated mtDNA tree and to estimate the divergence
848 of the most recent common ancestors (MRCAs) (template: BEAST (Heled and
849 Drummond 2010)). Fossil data were taken mainly from Moreau & Bell (2013) and
850 PaleoBio DB (paleobiodb.org) (see Table S4 for a detailed list of references) and used
851 as minimum calibration points constraining nodes in the topology as listed in Table S4.
852 Lognormal prior distribution was implemented with an offset corresponding to the
853 minimum fossil age, log(mean) of 1.0, and log(SD) of 1.0. Ten independent runs of 8 x

854 10^7 generations were sampled every 100,000th generation. Each partition was
855 modelled with an uncorrelated relaxed clock with a Yule process, species tree prior,
856 and its best substitution model. Evolutionary models not implemented in BEAUTi v2
857 (Bouckaert et al. 2014) were manually edited in the xml file. Tracer v. 1.6
858 (http://beast.bio.ed.ac.uk/Tracer) was used to evaluate convergence and the
859 parameters' effective sampling size (ESS) and a burn-in was manually implemented
860 and roughly 10% from each run were removed. LogCombiner (Bouckaert et al. 2014)
861 and TreeAnnotator (Bouckaert et al. 2014) were used to summarize the results in a
862 single consensus tree. The non-ant species were excluded from the analysis, and no
863 outgroup was adopted. Divergence dates for the maximum-likelihood nuclear DNA
864 (nuDNA) tree (only ants) were also inferred using the penalized likelihood approach
865 (Sanderson 2002) implemented in the R package APE v. 5.1 (Paradis et al. 2004) by
866 calibrating four nodes: the root, the *Tetramorium*, the *Solenopsis*+*Monomorium*, and
867 the Attini branches (Table S4).
868
869 **Functional annotation and orthologous-group dynamic evolution**
870 Genes from 19 ant and three outgroup species (Table S3) were clustered into ortholog
871 groups (OGs) with Hieranoid v. 2 (Kaduk and Sonnhammer 2017), which implements
872 InParanoid v. 8 (Sonnhammer and Östlund 2015) using a guide tree topology based on
873 Moreau & Bell (2013) and BLASTp (*e*-value cutoff of 10^{-5}) to improve the reciprocal hit
874 accuracy (Edgar 2010), while homology relationships were searched with deltaBLAST
875 (Boratyn et al. 2012). The functional annotation of each putative protein coding
876 sequence was performed by identifying both the protein domain architecture and
877 Gene Ontology (GO) terms. For each sequence, first HMMER v. 3.1b2 (hmmscan) (Eddy

878 2011) was used to predict PFAM-A v. 31.0 domains, then Domain Annotation by a
879 Multi-objective Approach (DAMA) v. 2 (Bernardes et al. 2015) was applied to identify
880 architectures combining scores of domain matches, previously observed multi-
881 domain co-occurrence, and domain overlapping. Annotation of GO terms was
882 performed as implemented in the CATH assignments for large sequence datasets (Das
883 et al. 2015). Briefly, each input sequence was scanned against the library of CATH
884 functional families (FUNFAMS) HMMs (Sillitoe et al. 2015) using HMMER3 (Eddy 2011)
885 to assign FUNFAMS to regions on the query sequence (with conditional *E*-value < 0.005).
886 Then the GO annotations for a matching FUNFAM were transferred to the query
887 sequence with its confidence scores, calculated by considering the GO term frequency
888 among the annotated sequences. Finally, a non-redundant set of GO annotations was
889 retained, making up the GO annotations for the query protein sequence (Sillitoe et al.
890 2015).

891 To identify orthologous groups specifically expanded in the *T. alpestre*
892 genome, BADIRATE v. 1.35 (Librado et al. 2012) was performed twice, once for the
893 nuDNA and one for the mtDNA ultrametric tree topologies, reconstructing ancestral
894 family sizes, gain, death, and innovation (GDI), applying stochastic models and
895 allowing estimation of the family turnover rate (λ) by ML. Using this approach, we
896 observed a correlation between the turnover rate (λ) and branch lengths (Spearman
897 correlation: ρ = -0.52, *P*-value = 0), and a bimodal distribution of λ. We interpreted the
898 distribution closer to 0 as background noise and therefore considered only expanded
899 and contracted OGs with λ within the second distribution. The boundary between the
900 two distributions was defined by the valley values between the two, which were

901 computed using OPTIMIZE and APPROXFUN functions implemented in the STATS v. 3.4.0
902 package in R.
903
904 **Heat-shock protein family analysis**
905 From the functional annotation analyses of all Hymenoptera, all sequences bearing a
906 valid protein domain matching HSPs were extracted and aligned using the CLUSTALW
907 v.1.2.1 EBI web server (Larkin et al. 2007) (settings: MBED true; MBEDITERATION true;
908 ITERATIONS 5; GTITERATIONS 5; HMMITERATIONS 5). A phylogenetic tree was computed using
909 ML as implemented in FASTTREE v. 2.1.8 SSE3 (Price et al. 2010) and manually checked
910 to identify and provisionally annotate the five main subfamily proteins (Hsp90, Hsp70,
911 Hsp60, Hsp40, and sHsp) and their OG. Sequences falling outside clusters and with low
912 bootstrap values were manually checked with the DELTABLAST (Boratyn et al. 2012)
913 web-server and with the CONSERVED DOMAIN DATABASE (CDD) web-server (Marchler-
914 Bauer et al. 2017) to identify and remove bacterial contaminants and
915 spurious/incomplete sequences. Then, nucleotide sequences belonging to each
916 subfamily were revers translated into amino acids, aligned using CLUSTALW v.1.2.1, and
917 the obtained protein alignment was used to derive the nucleotide one. FASTTREE was
918 used again to compute phylogenetic trees for the five HSP subfamilies.
919
920 **Selection on single-copy orthologous groups and gene families**
921 All selected scOGs and HSP-OGs were scanned to evaluate selection signature on
922 coding regions of *T. alpestre*. This was done by computing the mean ω and the
923 relaxation of each branch of the phylogeny of *Tetramorium* spp. + *Vollenhovia emeryi*.
924 More specifically, adopting a pipeline similar to that in Cicconardi et al. (2017a,

925 2017b), the signatures of diversifying selection were searched in codon-based aligning
926 groups of one-to-one orthologous ant genes with MACSE v. 1.01b (Ranwez et al.
927 2011), filtering with GBLOCKS v. 0.91b (Castresana 2000) under a "relaxed" condition
928 (Parker et al. 2013; Cicconardi et al. 2017a, 2017b) and using the aBSREL algorithm as
929 implemented in the HYPHY batch language (Kosakovsky Pond et al. 2005) using a batch
930 script (BRANCHSITEREL) in HYPHY (http://github.com/veg/hyphy).

931 In parallel, the same scOGs and HSP-OGs were scanned using the RELAX test
932 (Wertheim et al. 2014) to search for putative signals of relaxation. In brief, RELAX tests
933 the hypothesis of evolutionary-rate relaxation in selected branches of a phylogenetic
934 tree compared with reference branches. A *k* value is computed to evaluate whether
935 the selective strength ω shifts towards neutrality. The rate of d_N/d_S (ω) can relax ($k <$
936 1), stay stable ($k = 1$), or intensify ($k > 1$). For both scOGs and HSP-OGs, and for both
937 the ABSREL and RELAX tests, the Bonferroni-Holm sequential rejection procedure
938 (Abdi 2010) was used to control the false discovery rate with a very stringent cutoff
939 value of 0.005 for the adjusted *P*-values applied (Benjamin et al. 2017). Hsp60s and
940 Hsp40s under relaxed selection were analyzed using the GENEMANIA prediction server
941 (Warde-Farley et al. 2010) to predict their functions by creating an interaction
942 network with genes by including protein and genetic interactions, pathways, co-
943 expression, co-localization, and protein domain similarity.

944 Finally, the GOSTATS package for R (Falcon and Gentleman 2007) (settings:
945 ANNOTATION org.Dm.eg.db; CONDITIONAL TRUE; TESTDIRECTION over) was used to check for
946 GO terms of biological processes and KEGG pathway enrichments of scOGs under
947 selection and relaxation using the whole set of tested scOGs as background. *P*-values
948 < 0.005 and < 0.05 were implemented for GO terms and KEGG pathways, respectively.

949

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

951 Following identification as under diversifying selection in *T. alpestre* (see results
952 section "Dual signature of evolutionary pressures in single-copy OGs in the *T. alpestre*
953 genome"), synthesis of genes and enzymes was attempted for phosphofructokinase,
954 transketolase, phosphoglucosemutase, and Hex-t1 was successful just for the Hex-t1.
955 In more detail, the Hex-t1 genes of *Tetramorium alpestre* (1398 bp) and *T. immigrans*
956 (1386 bp) were synthesized by ThermoFisher lifetechnologies (Carlsbad, CA, USA) and
957 Eurofins Genomics (Ebersberg, Germany), respectively. Both constructs carried a 5'
958 EcoRI and a 3' HindIII recognition site for subsequent cloning as well as a C-terminal
959 His-tag. The genes were excised from their carrier backbones by enzymatic digestion.
960 The fragments were purified by agarose gel excision and sub-cloned into the
961 pACEBac1 expression vector (MULTIBACTM, Geneva Biotech, Geneva, Switzerland).
962 Correct insertion was verified by Sanger sequencing, and transformation into
963 *Drosophila melanogaster* cells and protein expression followed the standard protocols
964 of the expression vector manufacturer. After cell harvesting, the presence of the
965 recombinant protein in the cell lysates was verified by a western blot targeting the
966 His-tag. The relative amounts of protein were quantified from the western blot using
967 ImageJ (https://imagej.net/Welcome). The cell lysates were centrifuged at 4000 rpm
968 at 4 °C for 5 min, and the supernatant was directly used as protein source.

969 Enzyme activity assays followed a modified protocol of Crabtree & Newsholme
970 (1972). Briefly, 1000 µl assay medium containing 75 mM Tris pH=7.5 (Merck Millipore,
971 Burlington, MA, USA), 7.5 mM $MgCl_2$ (Merck), 0.8 mM EDTA (Sigma-Aldrich, St. Louis,
972 MO, USA), 1.5 mM KCl (Merck), 4.0 mM mercaptoethanol (Sigma), 0.4 mM NADP+

973 (Abcam, Cambridge, UK), 2.5 mM ATP (Abcam), and 10 mM creatine phosphate
974 (Abcam) were placed in a disposable cuvette (Brand, Germany) in a SPECORD 210
975 PLUS UV/VIS spectrophotometer (Analytic Jena AG, Germany) with an attached 1157P
976 programmable thermostate (VWR, Radnor, PA, USA). Variable amounts of D-glucose
977 (VWR) and synthetic protein were added, and the cuvette was allowed to reach the
978 assay temperature. Then, the reaction was started by adding 15 U creatine
979 phosphokinase (Sigma) and 5 U glucose 6-phospahte dehydrogenase (Sigma).
980 The Hex-t1 activity was measured in triplicate as the rate of reduction of $NADP^+$
981 causing an extinction at 340 nm. Each measurement lasted for 60 s, and every second,
982 an extinction value was recorded. For the determination of the Michaelis-Menten
983 constant (Km-value) of both enzymes, assays were conducted along a substrate
984 gradient from 0.00167 to 24.30000 mM D-glucose, achieving saturated extinction
985 curves. Assay temperatures ranged from 6 °C to 36 °C with 10 K increments. The
986 software PRISM v. 8 (GraphPad Software, San Diego, CA, USA) was used to calculate the
987 Km-values.
988
989 **Hex-t1 structure modelling and molecular dynamics simulations**
990 The three-dimensional structures of *T. alpestre* and *T. immigrans* Hex-t1 proteins are
991 not available. To study the effect of the amino acid changes on the proteins'
992 function/dynamics, we modeled the structures of these two proteins via homology
993 modelling. The X-ray structure of the Hex-t1 isoform from *Schistosoma mansoni*
994 (Mulichak et al. 2002), solved at 2.6 Å resolution was used as a template and the
995 models were generated using the software MODELLER v. 9.21 (Webb and Sali 2014).
996 Once we obtained the models, we built the simulative systems to run molecular

997 dynamics simulations. The two proteins TalpHex-t1 and TimmHex-t1, in complex with
998 the glucose substrate, were immersed in a triclinic box filled with TIP3P water
999 molecules (Jorgensen et al. 1983) and rendered electroneutral by the addition of
1000 cloride counterions. The topology of the two systems, consisting of ~ 80.000 atoms
1001 each, was built using the AMBER14 force field (Case et al. 2014), further converted in
1002 GROMACS v. 4.6 (Hess et al. 2008) format using ACPYPE (Sousa Da Silva and Vranken
1003 2012). Simulations were run on a GPU cluster using GROMACS v. 4.6 with the following
1004 protocol: 1) 25000 steps of steepest descent followed by 25000 steps of coniugate
1005 gradient minimization; 2) 5*100 ps of equilibration runs in the NVT environment
1006 starting at 50 K and performed by increasing the temperature of 50 K after each run
1007 untill a final value of 250 K; 3) 5*100 ps of equilibration runs in the NPT environment
1008 starting at 50 K and performed by increasing the temperature of 50 K after each run
1009 untill a final value of 250 K; 4) the equilibrated systems were simulated for 500 ns at
1010 279 or 300 K, i.e. at 6 or 26 °C, for a total of four simulations and 2 µs of sampling.
1011 Electrostatic interactions were considered by means of the Particle Mesh Ewald
1012 method (PME) with a cut off of 1.2 nm for the real space and Van der Waals
1013 interactions (Darden et al. 1993). Bond lengths and angles were constrained via the
1014 LINCS algorithm (Hess et al. 1997). The temperatures were kept constant at 279 or
1015 300 ˚K by using the velocity rescale with a coupling constant of 0.1 ps, and the pressure
1016 was kept at 1 bar using the Parrinello-Rahman barostat with a coupling constant of
1017 1.0 ps during sampling (Parrinello and Rahman 1981). The four trajectories were
1018 collected and comparative analyses performed with the GROMACS suite or with in
1019 house written code.
1020

1021 **Chill coma assay and quantitative qRT-PCR assay**

1022 The cold hardiness and expression of selected genes were compared between *T.*
1023 *alpestre* and *T. immigrans* using chill coma assays. From four colonies additional to
1024 those used for other experiments in this study, two per species, about 500 workers
1025 per colony were collected in August 2017; *T. alpestre*: Jaufenpass (18980: 46.83791°
1026 N, 11.29768° E) and Penser Joch (18978: 46.81402° N, 11.44198° E); *T. immigrans*:
1027 Vienna (18977: 48.12448° N, 16.43522° E.; 18983: 48.31341° N, 16.42529° E). Workers
1028 were collected alive using aspirators, transported to the laboratory in Innsbruck, and
1029 kept in polypropylene boxes (18.5 cm x 11 cm) with chambers of various sizes. The
1030 walls of the boxes were Fluon-coated (GP1, De Monchy International BV, Rotterdam,
1031 Netherlands) to prevent workers from escaping. Food (sugar-honey-water and deep-
1032 frozen *Drosophila* flies) and tap water were provided *ad libitum* three times a week.
1033 Workers were kept in a climate chamber (MIR-254, Panasonic, Etten Leur,
1034 Netherlands) at a 12L:12D photoperiod at constant 18 °C. The temperature of 18 °C
1035 was the average of the monthly mean temperatures experienced at 200 m a.s.l. (*T.*
1036 *immigrans*, approx. 23.7 °C) and at 2000 m a.s.l. (*T. alpestre*, approx. 12.3 °C) in July
1037 2017. Workers were acclimatized to 18 °C for at least three weeks before the chill
1038 coma assays.

1039 After the acclimation period, from each of the four colonies, 90 workers were
1040 randomly chosen and divided into 18 equivalent pools, each were used for the chill
1041 coma and for the control assays. For the chill coma assay, worker pools were
1042 transferred into empty 5-ml glass vials, sealed, and immersed for 6.5 hours in a
1043 water:ethane-1,2-diol mix (1:1) bath set at -6 °C; in pilot experiments, this
1044 temperature had been determined as the highest temperature at which the ants had

1045 fallen into chill coma after 6.5 hours exposition; the temperature was identical for
1046 both species. Temperature was monitored using an electronic thermometer (TFX 430,
1047 ebro Electronic GmbH; Ingolstadt, Germany) with an accuracy of 0.05 °C inserted into
1048 an additional, empty vial treated in the same way as the vials with ants. After chill
1049 coma, the vials were transferred to a climate chamber at 18 °C, and the workers were
1050 allowed to recover for 0.5, 2.0, and 4.0 hours. For each recovery time, three replicate
1051 pools were used. After recovery, the workers were transferred to RNAse-free reaction
1052 tubes and killed in liquid nitrogen. Tubes were stored at -70 °C. For the control assay,
1053 worker pools were transferred to empty glass vials and placed in the climate chamber
1054 at 18 °C. After 6.5 hours plus respective recovery time, the workers were killed and
1055 stored as described above.

1056 For each recovery time and replicate, complete worker pools were used for the
1057 molecular analyses. RNA was extracted using the NucleoSpin® RNA Kit (Macherey-
1058 Nagel, Düren, Germany) following the instructions of the manufacturer. First-strand
1059 cDNA was synthesized using 200 U RevertAid reverse transcriptase (all reagents by
1060 Thermo Fisher Scientific, Waltham, USA), 40 U RiboLock RNase inhibitor, 5 µM random
1061 hexamer primers, 100 µM dNTPs, and 2 µl RNA extract in a total volume of 40 µl. The
1062 mixture was incubated for 5 min at 25 °C, 60 min at 42 °C, and 5 min at 70 °C on a
1063 UnoCycler 1200 (VWR, Radnor, USA). Quantitative PCR was conducted on a Rotorgene
1064 Q (Qiagen) PCR system. Each reaction contained 1× Rotor-Gene SYBR Green PCR
1065 Mastermix (Qiagen), 0.2 µM of primers, specifically designed for target genes (Table
1066 S12), and 1 µl cDNA in a total volume of 10 µl. All qRT-PCR reactions were performed
1067 as triplicates. Cycling conditions were 95 °C for 5 min, followed by 40 cycles of 94 °C
1068 for 15 s, 58 °C for 10 s, and 72 °C for 15 s. Fluorescence was acquired at the end of

1069 each elongation step. PCR was followed by a melting curve analysis from 60 to 95 °C
1070 with 0.1 °C increments held for 5 s before fluorescence acquisition. Cycle threshold
1071 (Ct) values were subtracted from total number of cycles. On these values, linear-
1072 regression analyses were based separately for each target gene, treatment, recovery
1073 time, and species (in each regression analysis, two populations and three replicate
1074 values were used, i.e., n = 6) using the housekeeping gene as independent variable
1075 and the target gene as dependent variable. For each target gene / recovery-time
1076 combination one-way analysis of covariance (ANCOVA) followed by Bonferroni-Holm
1077 correction was then used to identify significant differences between treatments
1078 within species and between species within treatment.
1079
1080

1081 ***Data access***

1082 The sequence data from this study have been submitted to the National Center for
1083 Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject
1084 numbers PRJNA532334, PRJNA533534, PRJNA533535, PRJNA533536 and
1085 PRJNA533537 (https://www.ncbi.nlm.nih.gov/bioproject/). The mtDNA genome
1086 assemblies have been submitted to the NCBI (GenBank accessions: MK861047 -
1087 MK861070). All accession codes of deposited and retrieved data are provided in Table
1088 S1.
1089
1090

1091 ***Acknowledgements***

1092 Philipp Andesner and Stefan Gross for valuable work in the laboratory. Research was
1093 supported by the Austrian Science Fund (FWF, P23409, P30861). The computational
1094 results presented here have been achieved in part by using the HPC infrastructure of
1095 the University of Innsbruck (LEO), the Vienna Scientific Cluster (VSC), the
1096 collaborative system of the Universities Innsbruck and Linz (MACH), and the Swiss
1097 National Supercomputing Centre (CSCS) hosted in Lugano.

1098

1099

1100 ## *References*

1101 Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. 2015. spThin:
1102 An R package for spatial thinning of species occurrence records for use in
1103 ecological niche models. *Ecography (Cop)* **38**: 541–545.

1104 Archambault V, Lépine G, Kachaner D. 2015. Understanding the Polo Kinase
1105 machine. *Oncogene* **34**: 4799–4807.

1106 Bankevich A, Nurk S, Antipov D, Gurevich A a., Dvorkin M, Kulikov AS, Lesin VM,
1107 Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: A New Genome
1108 Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput*
1109 *Biol* **19**: 455–477.

1110 Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, Bollen
1111 KA, Brembs B, Brown L, Camerer C, et al. 2017. Redefine statistical significance.
1112 *Nat Hum Behav*. http://www.nature.com/articles/s41562-017-0189-z.

1113 Bens M, Sahm A, Groth M, Jahn N, Morhart M, Holtze S, Hildebrandt TB, Platzer M,
1114 Szafranski K. 2016. FRAMA : from RNA-seq data to annotated mRNA assemblies.
1115 *BMC Genomics* 1–12. http://dx.doi.org/10.1186/s12864-015-2349-8.

1116 Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2014.
1117 GenBank. *Nucleic Acids Res* **43**: D30–D35.
1118 http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku1216.
1119 Bernardes JS, Vieira FRJ, Zaverucha G, Carbone A. 2015. A multi-objective
1120 optimization approach accurately resolves protein domain architectures.
1121 *Bioinformatics* **32**: 345–353.
1122 Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J,
1123 Middendorf M, Stadler PF. 2013. MITOS: Improved de novo metazoan
1124 mitochondrial genome annotation. *Mol Phylogenet Evol* **69**: 313–319.
1125 http://dx.doi.org/10.1016/j.ympev.2012.08.023.
1126 Bertelsmeier C, Ollier S, Liebhold A, Keller L. 2017. Recent human history governs
1127 global ant invasion dynamics. *Nat Ecol Evol* **1**.
1128 https://www.nature.com/articles/s41559-017-0184.pdf.
1129 Bolton B. 2018. An online catalog of the ants of the world. Available from
1130 http://antcat.org.
1131 Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012.
1132 Domain enhanced lookup time accelerated BLAST. *Biol Direct* **7**: 12.
1133 http://www.ncbi.nlm.nih.gov/pubmed/22510480%5Cnhttp://www.pubmedcen
1134 tral.nih.gov/articlerender.fcgi?artid=PMC3438057.
1135 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard M a, Rambaut A,
1136 Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary
1137 analysis. *PLoS Comput Biol* **10**: e1003537.
1138 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3985171&tool=pm
1139 centrez&rendertype=abstract (Accessed May 27, 2014).

1140 Case DAA, Babin V, Berryman JTT, Betz RMM, Cai Q, Cerutti DSS, Cheatham TE, III,
1141 Darden TAA, Duke REE, et al. 2014. AMBER14. *Univ California, San Fr*.
1142 Christian KA, Morton SR. 1992. Extreme thermophilia in a central Australian ant,
1143 Melophorus bagoti. *Physiol Zool* **65**: 885–905.
1144 Cicconardi F, Di Marino D, Olimpieri PP, Arthofer W, Schlick-Steiner BC, Steiner FM.
1145 2017a. Chemosensory adaptations of the mountain fly Drosophila nigrosparsa
1146 (Insecta: Diptera) through genomics' and structural biology's lenses. *Sci Rep* **7**:
1147 43770. http://dx.doi.org/10.1038/srep43770.
1148 Cicconardi F, Marcatili P, Arthofer W, Schlick-Steiner BC, Steiner FM. 2017b. Positive
1149 diversifying selection is a pervasive adaptive force throughout the Drosophila
1150 radiation. *Mol Phylogenet Evol* **112**: 230–243.
1151 Clark MS, Worland MR. 2008. How insects survive the cold: molecular mechanisms—
1152 a review. *J Comp Physiol B* **178**: 917–933.
1153 http://link.springer.com/10.1007/s00360-008-0286-4.
1154 Colinet H, Hoffmann A. 2010. Gene and protein expression of *Drosophila* Starvin
1155 during cold stress and recovery from chill coma. *Insect Biochem Mol Biol* **40**:
1156 425–428. http://dx.doi.org/10.1016/j.ibmb.2010.03.002.
1157 Colinet H, Hoffmann AA. 2012. Comparing phenotypic effects and molecular
1158 correlates of developmental, gradual and rapid cold acclimation responses in
1159 Drosophila melanogaster. *Funct Ecol* **26**: 84–93.
1160 Coulson M, Robert S, Saint R. 2005. Drosophila starvin encodes a tissue-specific BAG-
1161 domain protein required for larval food uptake. *Genetics* **171**: 1799–1812.
1162 Danko CG, Choate LA, Marks BA, Rice EJ, Wang Z, Chu T, Martins AL, Dukler N,
1163 Coonrod SA, Tait Wojno ED, et al. 2018. Dynamic evolution of regulatory

1164 element ensembles in primate CD4+ T cells. *Nat Ecol Evol* **2**.
1165 http://www.nature.com/articles/s41559-017-0447-5.
1166 Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An N·log(N) method for
1167 Ewald sums in large systems. *J Chem Phys*.
1168 Das S, Sillitoe I, Lee D, Lees JG, Dawson NL, Ward J, Orengo CA. 2015. CATH
1169 FunFHMMer web server: Protein functional annotations using functional family
1170 assignments. *Nucleic Acids Res* **43**: W148–W153.
1171 Dobin A, Davis C a, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
1172 Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:
1173 15–21.
1174 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3530905&tool=pm
1175 centrez&rendertype=abstract.
1176 Doong H, Vrailas A, Kohn EC. 2002. What's in the "BAG"? - A functional domain
1177 analysis of the BAG-family proteins. *Cancer Lett* **188**: 25–32.
1178 Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* **7**.
1179 Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST.
1180 *Bioinformatics* **26**: 2460–2461.
1181 Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term
1182 association. *Bioinformatics* **23**: 257–258.
1183 Fischer MK, Shingleton AW. 2001. Host plant and ants influence the honeydew
1184 sugar. *Funct Ecol* **15**: 544–550.
1185 Gerken AR, Eller OC, Hahn D a, Morgan TJ. 2015. Constraints, independence, and
1186 evolution of thermal plasticity: Probing genetic architecture of long- and short-
1187 term thermal acclimation. *Proc Natl Acad Sci U S A* **112**: 4399–404.

1188 http://www.ncbi.nlm.nih.gov/pubmed/25805817.

1189 Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall
1190 G, Shea TP, Sykes S, et al. 2011. High-quality draft assemblies of mammalian
1191 genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* **108**:
1192 1513–1518.

1193 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson D a, Amit I, Adiconis X, Fan L,
1194 Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from
1195 RNA-Seq data without a reference genome. *Nat Biotechnol* **29**: 644–52.
1196 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3571712&tool=pm
1197 centrez&rendertype=abstract (Accessed March 19, 2014).

1198 Guénard B, Weiser MD, Gómez K, Narula N, Economo EP. 2017. The Global Ant
1199 Biodiversity Informatics (GABI) database: Synthesizing data on the geographic
1200 distribution of ant species (Hymenoptera: Formicidae). *Myrmecological News*
1201 **24**: 83–89.

1202 Harpur BA, Kent CF, Molodtsova D, Lebon JMD, Alqarni AS, Owayss AA, Zayed A.
1203 2014. Population genomics of the honey bee reveals strong signatures of
1204 positive selection on worker traits. *Proc Natl Acad Sci* **111**: 2614–2619.
1205 http://www.pnas.org/cgi/doi/10.1073/pnas.1315506111.

1206 Haslbeck M, Vierling E. 2015. A first line of stress defense: Small heat shock proteins
1207 and their function in protein homeostasis. *J Mol Biol* **427**: 1537–1548.
1208 http://dx.doi.org/10.1016/j.jmb.2015.02.002.

1209 Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus
1210 data. *Mol Biol Evol* **27**: 570–80.
1211 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2822290&tool=pm

1212 centrez&rendertype=abstract (Accessed February 29, 2012).
1213 Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. 1997. LINCS: A Linear Constraint
1214 Solver for molecular simulations. *J Comput Chem*.
1215 Hess B, Kutzner C, Van Der Spoel D, Lindahl E. 2008. GRGMACS 4: Algorithms for
1216 highly efficient, load-balanced, and scalable molecular simulation. *J Chem*
1217 *Theory Comput*.
1218 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution
1219 interpolated climate surfaces for global land areas. *Int J Climatol* **25**: 1965–
1220 1978.
1221 Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2015. BRAKER1:
1222 Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and
1223 AUGUSTUS. *Bioinformatics* **32**: 767–769.
1224 Hölldobler B, Wilson EO. 1990. *The Ants*.
1225 Holt RD. 2009. Bringing the Hutchinsonian niche into the 21st century: Ecological and
1226 evolutionary perspectives. *Proc Natl Acad Sci* **106**: 19659–19665.
1227 http://www.pnas.org/cgi/doi/10.1073/pnas.0905137106.
1228 Ivens ABF, Kronauer DJC, Pen I, Weissing FJ, Boomsma JJ. 2012. Ants farm
1229 subterranean aphids mostly in single clone groups - An example of prudent
1230 husbandry for carbohydrates and proteins? *BMC Evol Biol* **12**.
1231 Janicki J, Narula N, Ziegler M, Guénard B, Economo EP. 2016. Visualizing and
1232 interacting with large-volume biodiversity data using client-server web-mapping
1233 applications: The design and implementation of antmaps.org. *Ecol Inform* **32**:
1234 185–193.
1235 Jayakumar PC, Shouche YS, Patole MS. 2007. Functional analysis of *Drosophila*

1236 *melanogaster* hexokinase Hex-A locus : multiple Initiator-like elements enhance
1237 DPE containing promoter activity. **16**: 3–13.
1238 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison
1239 of simple potential functions for simulating liquid water. *J Chem Phys*.
1240 Kaduk M, Sonnhammer E. 2017. Improved orthology inference with Hieranoid 2.
1241 *Bioinformatics* **5**: btw774.
1242 http://bioinformatics.oxfordjournals.org/lookup/doi/10.1093/bioinformatics/bt
1243 w774.
1244 Keeling CI, Yuen MM, Liao NY, Roderick Docking T, Chan SK, Taylor G a, Palmquist DL,
1245 Jackman SD, Nguyen A, Li M, et al. 2013. Draft genome of the mountain pine
1246 beetle, Dendroctonus ponderosae Hopkins, a major forest pest. *Genome Biol*
1247 **14**: R27.
1248 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4053930&tool=pm
1249 centrez&rendertype=abstract.
1250 Kelley JL, Peyton JT, Fiston-Lavier A-S, Teets NM, Yee M-C, Johnston JS, Bustamante
1251 CD, Lee RE, Denlinger DL. 2014. Compact genome of the Antarctic midge is likely
1252 an adaptation to an extreme environment. *Nat Commun* **5**: 4611.
1253 http://www.nature.com/ncomms/2014/140812/ncomms5611/full/ncomms561
1254 1.html.
1255 Körner C, Paulsen J, Spehn EM. 2011. A definition of mountains and their bioclimatic
1256 belts for global comparisons of biodiversity data. *Alp Bot* **121**: 73–78.
1257 Kosakovsky Pond SL, Frost SDW, Muse S V. 2005. HyPhy: Hypothesis testing using
1258 phylogenies. *Bioinformatics* **21**: 676–679.
1259 Krapf P, Russo L, Arthofer W, Möst M, Steiner FM, Schlick-Steiner BC. 2018. An

1260 Alpine ant's behavioural polymorphism: monogyny with and without internest
1261 aggression in Tetramorium alpestre. *Ethol Ecol Evol* **30**: 220–234.
1262 https://www.tandfonline.com/doi/full/10.1080/03949370.2017.1343868.
1263 Kuhlemann J. 2007. Paleogeographic and paleotopographic evolution of the Swiss
1264 and Eastern Alps since the Oligocene. *Glob Planet Change* **58**: 224–236.
1265 Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. 2013. Blobology: exploring
1266 raw genome data for contaminants, symbionts and parasites using taxon-
1267 annotated GC-coverage plots. *Front Genet* **4**: 237.
1268 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3843372&tool=pm
1269 centrez&rendertype=abstract.
1270 Lamprecht A, Semenchuk PR, Steinbauer K, Winkler M, Pauli H. 2018. Climate change
1271 leads to accelerated transformation of high-elevation vegetation in the central
1272 Alps. *New Phytol* 447–459. http://doi.wiley.com/10.1111/nph.15290.
1273 Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. 2014. Selecting optimal
1274 partitioning schemes for phylogenomic datasets. *BMC Evol Biol* **14**: 82.
1275 http://www.biomedcentral.com/1471-2148/14/82.
1276 Larkin M a., Blackshields G, Brown NP, Chenna R, Mcgettigan P a., McWilliam H,
1277 Valentin F, Wallace IM, Wilm a., Lopez R, et al. 2007. Clustal W and Clustal X
1278 version 2.0. *Bioinformatics* **23**: 2947–2948.
1279 Levis NA, Pfennig DW. 2016. Evaluating ' Plasticity-First ' Evolution in Nature : Key
1280 Criteria and Empirical Approaches. *Trends Ecol Evol* **31**: 563–574.
1281 http://dx.doi.org/10.1016/j.tree.2016.03.012.
1282 Librado P, Vieira FG, Rozas J. 2012. BadiRate: Estimating family turnover rates by
1283 likelihood-based methods. *Bioinformatics* **28**: 279–281.

1284 Liu C, Guénard B, Garcia FH, Yamane S, Blanchard B, Yang DR, Economo E. 2015. New
1285 records of ant species from Yunnan, China. *Zookeys* **78**: 17–78.
1286 Liu D, Hu J, Agorreta J, Cesario A, Zhang Y, Harris AL, Gatter K, Pezzella F. 2010a.
1287 Tumor necrosis factor receptor-associated protein 1(TRAP1) regulates genes
1288 involved in cell cycle and metastases. *Cancer Lett* **296**: 194–205.
1289 Liu L, Yu L, Edwards S V. 2010b. A maximum pseudo-likelihood approach for
1290 estimating species trees under the coalescent model. *BMC Evol Biol* **10**: 302.
1291 http://www.biomedcentral.com/1471-
1292 2148/10/302%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=
1293 2976751&tool=pmcentrez&rendertype=abstract.
1294 Macdonald HC, Cunha L, Bruford MW, Florestas E, Macdonald HC, Sir T, Evans M,
1295 Avenue M. 2016. Development of genomic resources for four potential
1296 environmental bioindicator species: *Isoperla grammatica*, *Amphinemura*
1297 *sulcicollis*, *Oniscus asellus* and *Baetis rhodani*. *bioRxivorg*.
1298 Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK,
1299 Geer RC, Gonzales NR, et al. 2017. CDD/SPARCLE: Functional classification of
1300 proteins via subfamily domain architectures. *Nucleic Acids Res* **45**: D200–D203.
1301 Merow C, Smith MJ, Silander JA. 2013. A practical guide to MaxEnt for modeling
1302 species' distributions: What it does, and why inputs and settings matter.
1303 *Ecography (Cop)* **36**: 1058–1069.
1304 Moreau CS, Bell CD. 2013. Testing The Museum Versus Cradle Tropical Biological
1305 Diversity Hypothesis: Phylogeny, Diversification, And Ancestral Biogeographic
1306 Range Evolution Of The Ants. *Evolution (N Y)* **67**: 2240–2257.
1307 Mulichak AM, Wilson JE, Padmanabhan K, Garavito RM. 2002. The structure of

1308 mammalian hexokinase-1. *Nat Struct Biol*.

1309 Nguyen AD, Gotelli NJ, Cahan SH. 2016. The evolution of heat shock protein

1310 sequences, cis-regulatory elements, and expression profiles in the eusocial

1311 Hymenoptera. *BMC Evol Biol* **16**: 15. http://www.biomedcentral.com/1471-

1312 2148/16/15.

1313 Nygaard S, Hu H, Li C, Schiøtt M, Chen Z, Yang Z, Xie Q, Ma C, Deng Y, Dikow R, et al.

1314 2016. Reciprocal genomic evolution in the ant-fungus agricultural symbiosis.

1315 *Nat Commun* 1–9.

1316 Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of phylogenetics and evolution

1317 in R language. *Bioinformatics* **20**: 289–290.

1318 Parker DJ, Wiberg RAW, Trivedi U, Tyukmaeva VI, Gharbi K, Butlin RK, Hoikkala A,

1319 Kankare M, Ritchie MG. 2018. Inter and Intraspecific Genomic Divergence in

1320 Drosophila montana Shows Evidence for Cold Adaptation. **10**: 2086–2101.

1321 Parker J, Tsagkogeorga G, Cotton J a, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013.

1322 Genome-wide signatures of convergent evolution in echolocating mammals.

1323 *Nature* **502**: 1–9.

1324 http://dx.doi.org/10.1038/nature12511%5Cnpapers2://publication/doi/10.103

1325 8/nature12511.

1326 Parrinello M, Rahman A. 1981. Polymorphic transitions in single crystals: A new

1327 molecular dynamics method. *J Appl Phys*.

1328 Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a *de novo* assembler for

1329 single-cell and metagenomic sequencing data with highly uneven depth.

1330 *Bioinformatics* **28**: 1420–1428.

1331 http://www.ncbi.nlm.nih.gov/pubmed/22495754 (Accessed July 24, 2014).

1332 Prescott SL, Srinivasan R, Marchetto MC, Grishina I, Narvaiza I, Selleri L, Gage FH,
1333 Swigut T, Wysocka J. 2015. Enhancer Divergence and cis-Regulatory Evolution in
1334 the Human and Chimp Neural Crest. *Cell* **163**: 68–84.
1335 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood
1336 trees for large alignments. *PLoS One* **5**: e9490.
1337 Pringle EG, Novo A, Ableson I, Barbehenn R V, Vannette RL. 2014. Plant-derived
1338 differences in the composition of aphid honeydew and their effects on colonies
1339 of aphid-tending ants.
1340 Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple alignment of
1341 coding SEquences accounting for frameshifts and stop codons. *PLoS One* **6**:
1342 e22594.
1343 Rogora M, Frate L, Carranza ML, Freppaz M, Stanisci A, Bertani I, Bottarin R,
1344 Brambilla A, Canullo R, Carbognani M, et al. 2018. Assessment of climate
1345 change effects on mountain ecosystems through a cross-site analysis in the Alps
1346 and Apennines. *Sci Total Environ* **624**: 1429–1442.
1347 https://doi.org/10.1016/j.scitotenv.2017.12.155.
1348 Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M, Keller L. 2014. Patterns
1349 of positive selection in seven ant genomes. *Mol Biol Evol* **31**: 1661–1685.
1350 Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and
1351 divergence times: a penalized likelihood approach. *Mol Biol Evol* **19**: 101–109.
1352 http://www.ncbi.nlm.nih.gov/pubmed/11752195.
1353 Seifert B. 2018. *The ants of Central and North Europe*. ed. Tauer. Lutra Verlags- und
1354 Vertriebsgesellschaft.
1355 Shaw TI, Ruan Z, Glenn TC, Liu L. 2013. STRAW: Species TRee Analysis Web server.

Nucleic Acids Res **41**: 238–241.

Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG, et al. 2015. CATH: Comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res* **43**: D376–D381.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 1–11.

Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0. 2013-2015 . *http://www.repeatmasker.org*. http://repeatmasker.org.

Soberón J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol Lett* **10**: 1115–1123.

Sonnhammer ELL, Östlund G. 2015. InParanoid 8: Orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res* **43**: D234–D239.

Sousa Da Silva AW, Vranken WF. 2012. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res Notes*.

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**: 62.

Steiner FM, Seifert B, Moder K, Schlick-Steiner BC. 2010. A multisource solution for a complex problem in biodiversity research: Description of the cryptic ant species Tetramorium alpestre sp.n. (Hymenoptera: Formicidae). *Zool Anzeiger - A J Comp Zool* **249**: 223–254.

1380 http://linkinghub.elsevier.com/retrieve/pii/S0044523110000537 (Accessed
1381 April 9, 2014).
1382 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL,
1383 Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of
1384 RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**: 562–578.
1385 http://www.nature.com/doifinder/10.1038/nprot.2012.016 (Accessed March 1,
1386 2012).
1387 Uchida Y, Uesaka M, Yamamoto T, Takeda H, Irie N. 2018. Embryonic lethality is not
1388 sufficient to explain hourglass-like conservation of vertebrate embryos. *Evodevo*
1389 **9**: 1–11. https://doi.org/10.1186/s13227-018-0095-0.
1390 Wagner GP, Lynch VJ. 2010. Evolutionary novelties. *Curr Biol* **20**: 48–52.
1391 https://pdf.sciencedirectassets.com/272099/1-s2.0-S0960982210X00021/1-
1392 s2.0-S0960982209019459/main.pdf?x-amz-security-
1393 token=AgoJb3JpZ2luX2VjECIaCXVzLWVhc3QtMSJHMEUCIDsUh2jRKJL1EeOiMrN
1394 glWa%2BljWFLis7tEu%2FF65JtK%2FrAiEA3Ge%2FEhYUyff41n0NLuT%2F6f3OMY
1395 i9qUIAiE.
1396 Wagner HC, Arthofer W, Seifert B, Muster C, Steiner FM, Schlick-Steiner BC. 2017.
1397 Light at the end of the tunnel: Integrative taxonomy delimits cryptic species in
1398 the Tetramorium caespitum complex (Hymenoptera: Formicidae).
1399 *Myrmecological News* **25**: 95–129.
1400 Wallis GP, Waters JM, Upton P, Craw D. 2016. Transverse Alpine Speciation Driven by
1401 Glaciation. *Trends Ecol Evol* **31**: 916–926.
1402 http://dx.doi.org/10.1016/j.tree.2016.08.009.
1403 Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M,

1404 Grouios C, Kazi F, Lopes CT, et al. 2010. The GeneMANIA prediction server:
1405 biological network integration for gene prioritization and predicting gene
1406 function eds. J. Weaver, K. Anija, and T. Daspit. *Nucleic Acids Res* **38**: W214–
1407 W220. http://nar.oxfordjournals.org/content/38/suppl_2/W214.short
1408 (Accessed March 1, 2013).
1409 Webb B, Sali A. 2014. Protein structure modeling with MODELLER. *Methods Mol Biol*
1410 **1137**: 1–15.
1411 http://www.springerlink.com/openurl.asp?genre=book&id=doi:10.1385/15925
1412 93682%5Cnhttp://link.springer.com/10.1007/978-1-4939-0366-
1413 5%5Cnhttp://link.springer.com/protocol/10.1007/978-1-60327-058-
1414 8_8%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/24573470.
1415 Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2014. RELAX:
1416 Detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol* **32**: 1–
1417 13. http://www.ncbi.nlm.nih.gov/pubmed/25540451.
1418 Wilson EO. 1990. *Success and Dominance in Ecosystems: The Case of Social Insects*.
1419 Xian-Wu L, Wei-Hua X. 2016. Hexokinase is a key regulator of energy metabolism and
1420 ROS activity in insect lifespan extension. *Aging (Albany NY)* **8**: 245–258.
1421 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4789580/pdf/aging-08-
1422 245.pdf.
1423 Yoshida S, Tsutsumi S, Muhlebach G, Sourbier C, Lee M-J, Lee S, Vartholomaiou E,
1424 Tatokoro M, Beebe K, Miyajima N, et al. 2013. Molecular chaperone TRAP1
1425 regulates a metabolic switch between mitochondrial respiration and aerobic
1426 glycolysis. *Proc Natl Acad Sci* **110**: E1604–E1612.
1427 http://www.pnas.org/cgi/doi/10.1073/pnas.1220659110.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, Rythms, and Aberration in Global Climate 65 Ma to Present. *Science (80-)* **292**: 686–693.