Chromosome-level genome assembly and annotation of two lineages of the ant *Cataglyphis hispanica*: steppingstones towards genomic studies of hybridogenesis and thermal adaptation in desert ants

Hugo DARRAS [1,2,a,*], Natalia de Souza ARAUJO[1,3,*], Lyam BAUDRY[4,6], Nadège GUIGLIELMONI[1,7], Pedro LORITE[5], Martial MARBOUTY[6], Fernando RODRIGUEZ[8], Irina ARKHIPOVA [8], Romain KOSZUL[6], Jean-François FLOT[1,3], Serge ARON[1]

[1] Unit of Evolutionary Biology & Ecology, Université libre de Bruxelles (ULB), Brussels, Belgium
[2] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[3] Interuniversity Institute of Bioinformatics in Brussels – (IB)² Brussels, Belgium
[4] Centre for Integrative Genomics, University of Lausanne, Lausanne, Switzerland
[5] Department of Biologia Experimental, Universidad de Jaén, Jaén, Spain
[6] Unité Régulation Spatiale des Génomes, CNRS, Institut Pasteur, Paris, France
[7] Universität zu Köln, Köln, Germany
[8] Marine Biological Laboratory, Woods Hole, Massachusetts, USA

[a] corresponding author: Darras, H. hgdarras@gmail.com
[*] Authors contributed equally to this study

ABSTRACT

Cataglyphis are thermophilic ants that forage during the day when temperatures are highest and sometimes close to their critical thermal limit. Several *Cataglyphis* species have evolved unusual reproductive systems such as facultative queen parthenogenesis or social hybridogenesis, which have not yet been investigated in detail at the molecular level. We generated high-quality genome assemblies for two hybridogenetic lineages of the Iberian ant *Cataglyphis hispanica* using long-read Nanopore sequencing and exploited chromosome conformation capture (3C) sequencing to assemble contigs into 26 and 27 chromosomes, respectively. Further karyotype analyses confirm this difference in chromosome numbers between lineages; however, they also suggest it may not be fixed among lineages. We obtained transcriptomic data to assist gene annotation and built custom repeat libraries for each of the two assemblies. Comparative analyses with 19 other published ant genomes were also conducted. These new genomic resources pave the way for exploring the genetic mechanisms underlying the remarkable thermal adaptation and the molecular mechanisms associated with transitions between different genetic systems characteristics of the ant genus *Cataglyphis*.

KEYWORDS

Social insects; *Cataglyphis*; Genome assembly; Genome annotation; karyotype; social hybridogenesis

BACKGROUND

Ants of the genus *Cataglyphis* inhabit arid regions throughout the Old World, including inhospitable deserts such as the Sahara (Boulay *et al.* 2017; Lenoir *et al.* 1990). Their foraging activities are strictly diurnal, with most species being active during the hottest hours of the day (Cerda *et al.* 1998; Wehner *et al.* 1992). Some *Cataglyphis* species even forage at temperatures close to their critical thermal limits (Cerda *et al.* 1998). For instance, workers of the silver ant *Cataglyphis bombycina* have been observed to forage when ground temperatures exceed 60°C (Wehner *et al.* 1992), which supposedly provides a competitive advantage against lizard predators who avoid such harsh conditions. The high thermal tolerance seen in *Cataglyphis* species relies on a range of behavioral, morphological, physiological and molecular adaptations, such as exploitation of thermal refuges, elongated legs, high speed of movement and intense recruitment of heat-shock chaperone proteins (Aron and Wehner 2021; Gehring and Wehner 1995; Perez and Aron 2020; Perez *et al.* 2021; Pfeffer *et al.* 2019; Sommer and Wehner 2012; Willot *et al.* 2017).

In addition to their impressive heat tolerance, *Cataglyphis* ants are prominent social insect models because of their amazing diversity of reproductive traits: the number of queens per colony, the mating frequencies, the dispersal strategies and the modes of production of different castes all vary greatly among species (Aron *et al.* 2016a, 2016b; Mardulyn and Leniaud 2016; Peeters and Aron 2017). Unusual reproductive systems relying on conditional use of sex to produce different female castes have evolved repeatedly in different *Cataglyphis* groups. Under these systems, non-reproductive workers are sexually generated, while reproductive queens are asexually produced by thelytokous parthenogenesis – a strategy that allows queens to increase the transmission rate of their genes to their reproductive female offspring while maintaining genetic diversity in the worker force (Kuhn *et al.* 2020; Pearcy *et al.* 2004). Males arise from arrhenotokous parthenogenesis, as is usually the case in Hymenoptera. In several species, the conditional use of sex evolved into a unique reproductive system, named clonal social hybridogenesis, whereby male and female sexuals are produced by parthenogenesis while workers are produced exclusively from interbreeding between two sympatric, yet non-recombining genetic lineages (Darras *et al.* 2014; Eyer *et al.* 2013; Kuhn *et al.* 2020; Leniaud *et al.* 2012).

The unique characteristics of *Cataglyphis* make this ant genus an interesting model to investigate the genetic mechanisms underlying thermal adaptation and the evolution of alternative reproductive strategies. To date, only one incomplete assembly of the genome of *Cataglyphis niger*, a species characterized by classical haplodiploid reproduction, is available for genomic analyses (Yahav and Privman, 2019). To fill this gap, we combined Oxford Nanopore long reads, Illumina short reads and chromosome conformation capture (3C) sequencing (Flot *et al.* 2015; Lieberman-Aiden *et al.* 2009; Marie-Nelly *et al.* 2014) to generate high-quality chromosome-scale genome assemblies of two lineages of the Iberian ant *Cataglyphis hispanica* (Figure 1). We also annotated and compared the repeats and gene sets of this species with those of other ant genera.

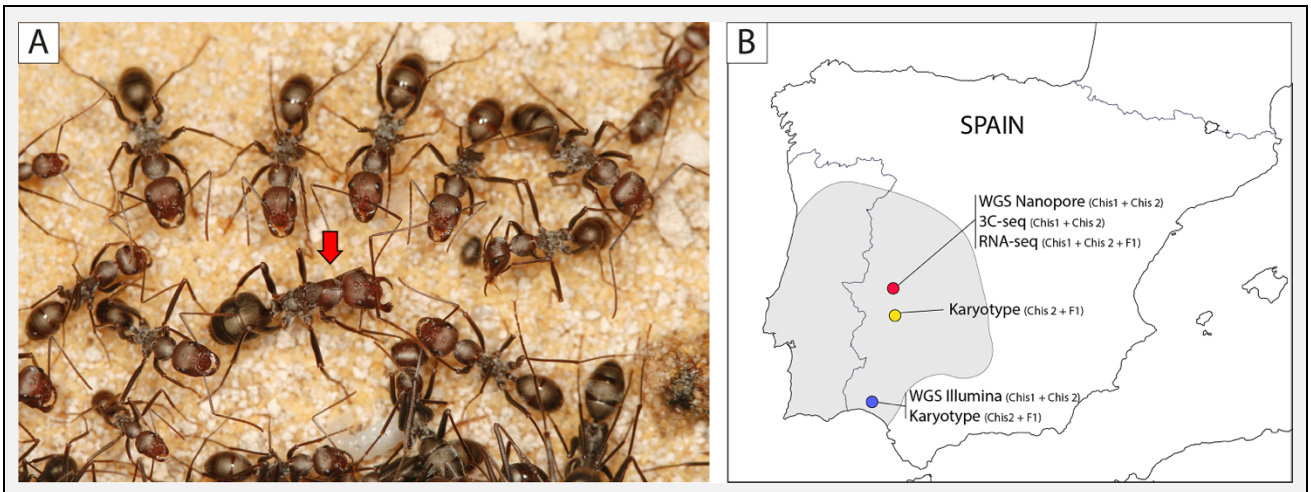


Figure 1. The ant *Cataglyphis hispanica.* (A) A queen of *C. hispanica* (red arrow) surrounded by workers. (B) Sampled sites in southwest Spain. The two interdependent lineages of the species, Chis1 and Chis2, were collected in Caceres (red), Merida (yellow) and Bonares (blue). For each lineage, a male from Bonares was used for whole genome short read sequencing (WGS Illumina) and queens from Caceres were used for both long read sequencing (WGS Nanopore) and chromosome conformation capture sequencing (3C-seq). Karyotypes of two Chis2 males and three hybrid (F1) workers were obtained from Merida and Bonares. To assist gene annotation, transcriptomes (RNA-seq) were generated from Chis1 and Chis2 individuals from Caceres. The complete range of the species *C. hispanica* is shown in grey.

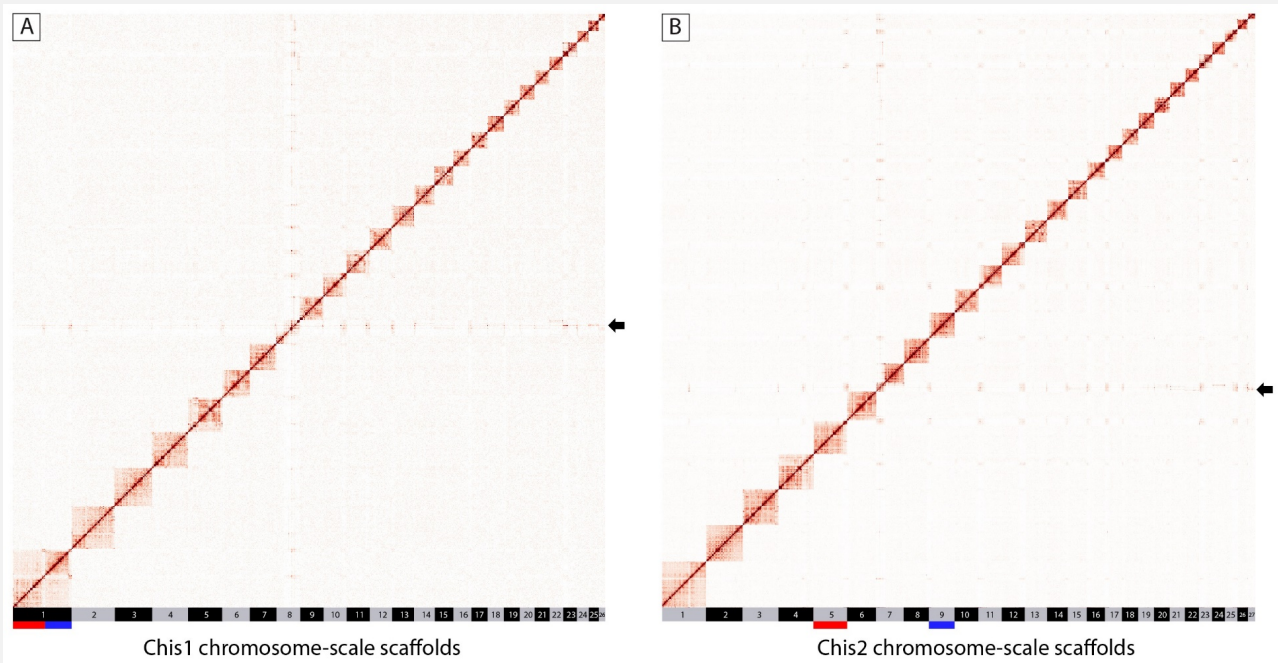
RESULTS AND DISCUSSION

Genome assemblies

Cataglyphis hispanica inhabits the most arid habitats of the Iberian Peninsula. Two sympatric hybridogenetic lineages (Chis1 and Chis2) co-occur as a complementary pair across the distribution range of the species (Leniaud *et al.* 2012; Darras *et al.* 2014). Queens of each lineage mate with males from the other lineage and produce non-reproductive workers by sexual reproduction. By contrast, male and female reproductive individuals are produced clonally through arrhenotokous and thelytokous parthenogenesis, respectively. As a result, all workers in the colonies are inter-lineage hybrids, but the two reproductive lineages are maintained divergent.

The genomes of the two hybridogenetic lineages were assembled independently (see Figure S1 for a schematic drawing of the assembly pipeline). For each of the Chis1 and Chis2 lineages, we generated respectively 5.7 and 5.1 Gbp of Nanopore reads from a pool of sister clonal queens (for *de novo* long-read assemblies); 32.2 and 34.2 Gbp of PE 2 x 100 bp Illumina reads with insert sizes ranging from 170 bp to 800 bp from a single male (for short read error correction/polishing); and 8.7 and 7.0 Gbp of 3C-seq PE 2 x 66 bp (after demultiplexing) Illumina reads from a single queen (for scaffolding). The long-read assembler Flye (Kolmogorov *et al.* 2019) generated assemblies consisting of several hundreds of contigs (439 and 929, respectively). The contigs were scaffolded using the 3C data (Marie-Nelly *et al.* 2014; Baudry *et al.* 2020): 99.7% of the Chis1 assembly was scaffolded into 26 chromosome-scale (> 2.4 Mb in length) scaffolds (Figure 2A), while 98.8 % of the Chis2 assembly was scaffolded into 27 chromosome-scale scaffolds (Figure 2B). These chromosome-scale scaffolds were numbered by decreasing size. The remaining 0.3 – 1.2% unscaffolded sequences were all relatively small (<40 kb for Chis1, <120 kb for Chis2). The overall sizes of the two scaffolded assemblies were 206 Mb and 209 Mb, respectively. Assembly completeness, as estimated by BUSCO scores (Manni *et al.* 2021), was very high: among the 5,991 highly conserved single-copy genes of the Hymenoptera odb10 database, 96.8% (Chis1) and 96.1% (Chis2) were complete in each assembly. In addition, only 0.5-0.4% of the BUSCO genes appeared duplicated for both assemblies, suggesting that our assemblies did not contain much uncollapsed haplotypes, if any. In line with these results, KAT analyses based on the Illumina reads

98 of each lineage showed a single peak of k-mer multiplicity, which were almost all represented
99 exactly once in the assemblies as expected for high-quality genomes (Figure S2); k-mer
100 completeness was estimated as 98.86% for Chis1 and 98.45%for Chis2 (Mapleson *et al.* 2016). For
101 each assembly, a region with no large-scale synteny pattern was assembled at the extremity of one
102 scaffold (the first 5.4 Mb of scaffold #9 of Chis1 and the first 3.1 Mb of scaffold #7 of Chis2). Each
103 of these regions consisted of a collection of small contigs (mostly in the 2-10 Kb range) with 2 to 5
104 times higher average coverage compared to other genomic regions. These sequences exhibited
105 microsyntenies with the extremities of other large scaffolds (Figure 2 and S3) suggesting that they
106 correspond to repeat sequences that were improperly assembled into fragmented contigs.
107
108



109

110 **Figure 2.** Assembly of the *Cataglyphis hispanica* Chis1 (A) and Chis2 (B) genomes into
111 chromosomes. Hi-C interaction map revealing the presence of 26 and 27 linkage groups. The color
112 scale represents the interaction frequencies. The positions of the rearranged chromosome are
113 indicated, and the arrows show the assembly artefact found in each genome (see main text). The
114 longest chromosome of Chis1 is split in two chromosomes in Chis2 (scaffolds 5 and 9, shown with
115 red and blue colors).
116

117 Comparison of the Chis1 and Chis2 assemblies revealed that 25 of the chromosome-scale
118 scaffolds had a one-to-one homolog in each of the two lineages. In addition, and by contrast, the
119 largest scaffold of Chis1 (#1) was split into two chromosome-scale scaffolds (# 5 and #9) in the
120 Chis2 assembly (Figure S3). The 3C contact maps of both lineages showed that these scaffolds
121 (Chis1 #1 and Chis2 #5, #9) correspond to well-individualized 3D features, thereby ruling out a
122 scaffolding error (Figure 2). These observations support that a centric fusion or fission
123 (Robertsonian translocation) took place in one of the two lineages studied. Robertsonian
124 translocations are the main mechanism of karyotype evolution in many animal groups, including
125 ants (Lorite and Palomeque, 2010) and can promote speciation through the suppression of genetic
126 recombination in the vicinity of rearranged centromeric regions or the reduction of fertility in
127 karyotypic hybrids (Davisson and Akeson, 1993; Faria and Navarro, 2010). Intrachromosomal
128 rearrangements between the lineages, consisting in large translocations and inversions, were also
129 observed for 6 of the 25 large orthologous scaffolds (Figure S3), but these could not be confirmed
130 independently with the current data.
131
132 **Karyotyping**
133 The numbers of chromosomes inferred for the Chis1 and Chis2 assemblies (n=26 and 27,
134 respectively) are within the range observed in karyotypes of *Cataglyphis bicolor* (n=26), *Cataglyphis*
135 *iberica* (n=26) and *Cataglyphis setipes* (n=26), as well as other Formicine species of the genera
136 *Formica* (n=26-27), *Iberoformica* (n=26) and *Polyergus* (n=27) (Hauschteck-Jungen and Jungen,
137 1983; Imai *et al.* 1984; Lorite and Palomeque 2010). To determine whether the two lineages of *C.*
138 *hispanica* are fixed for different chromosomal arrangements, we inspected metaphase chromosome
139 slides from male and worker pupae from different populations (Figure 1B). In ants, as in other social
140 Hymenoptera, males are haploid (*n*) whereas workers are diploid females (2*n*). Two males of the
141 Chis2 lineage from Merida and Bonares were analyzed (Figure 3A and S4A-D). Both male
142 karyotypes carried 27 chromosomes as was inferred with 3C data for the Chis2 lineage from the
143 Caceres population. The precise morphology of the chromosomes could not be determined due to
144

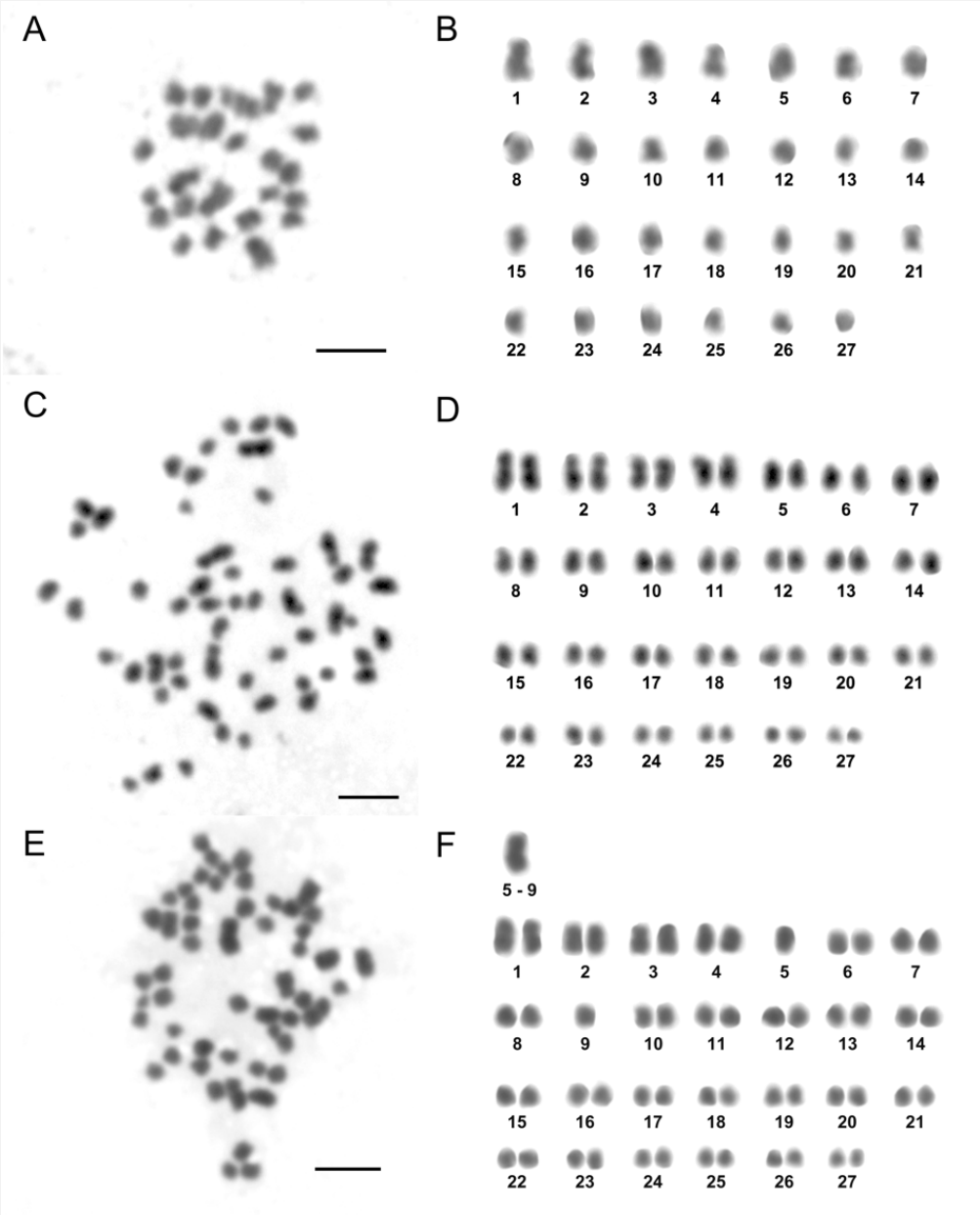


Figure 3. Karyotype analyses of *Cataglyphis hispanica.* (A,C and E) Metaphase chromosome slides of one haploid Chis2 male from Merida (A) and two F1 hybrid workers from Bonares (C) and Merida (E). (B,D and F) Corresponding karyotypes showing that the haploid chromosome number varies across populations from an haploid number of 26 (F) to 27 (B, D and F). The bar in all the images is 2 µm.

their small sizes (Figure 3). No male or queen pupa of the Chis1 lineage could be obtained for karyotyping. Instead, we indirectly inferred the karyotype variation in the Chis1 lineage using worker samples. Workers of *C. hispanica* are first generation hybrids and would, therefore, be expected to

156 carry odd chromosome numbers (i.e. 2n=26+27) if the two lineages were fixed for different
157 karyotypes. Workers from Bonares (N=2 from different colonies) and Merida (N=1) were analyzed.
158 The two workers from Bonares carried odd number of chromosomes (2n=54, Figure 3C and S4E)
159 suggesting that the parental lineages carry the same number of chromosomes in this population
160 (i.e. n=27). By contrast, the worker from Merida carried 53 chromosomes consistent with
161 expectations based on genome assemblies (Figure 3E and S5). If our assumptions are correct,
162 these results indicate that the number of chromosomes in the Chis1 lineage may vary in different
163 populations from n=26 (in the population used for 3C sequencing) to n=27 (in the population used
164 for karyotyping). The chromosomal polymorphism observed between our Chis1 and Chis2 genome
165 assemblies is therefore unlikely to be linked to the long-term maintenance of the two lineages.
166 Without clear karyotypes of pure Chis 1 individuals at hand, we were however unable to verify this
167 hypothesis. An alternative scenario would be that during the generation of the hybrid individuals a
168 fission sometimes occurs in the large chromosome #1 of Chis 1 producing 27 Chis1 chromosomes
169 in workers instead of 26.
170
171 **Gene annotation**
172 We annotated the genome of the Chis2 lineage (see Figure S1 for a schematic drawing describing
173 the genome annotation pipeline). *Ab initio* gene prediction using AUGUSTUS and homology-based
174 predictions using GenomeThreader (Gremme *et al.* 2005) identified 16,993 and 8,234 gene models,
175 respectively. A total of 40,969 models (including isoforms) were also predicted by the
176 PASA/Transdecoder (Haas *et al.* 2003) pipeline using direct evidence from 13 Gbp of Illumina RNA-
177 seq data. The three annotation sets were validated and combined into a single annotation of 16,146
178 non-overlapping models using EvidenceModeler (Haas *et al.* 2008). Among these, 11,101 gene
179 models showed significant similarity to proteins predicted in other ant species (blastp against 18 ant
180 proteomes from the RefSeq collection) and 10,543 had functional information inferred through
181 sequence orthology with the eggnog v5.0 database which covers more than five thousands
182 organisms (Huerta-Cepas *et al.* 2017, 2019). We filtered out all gene models non validated by at
183 least one of these databases to obtain a final dataset of 11,290 high quality gene models, 11,033

(98%) of which are placed within the 27 chromosome-scale scaffolds. This gene set is comparable in size to those annotated by the NCBI Eukaryotic Genome Annotation Pipeline for other ant genomes (range: 10,491-15,668; N= 18 different RefSeq ant genera; Table S1). We compared the obtained gene set of *C. hispanica* (Chis2) with 19 published ant annotations. Out of the 258,587 protein-coding genes analyzed using OrthoFinder (Emms and Kelly 2019), 96.82% (250,353) were placed in 13,698 orthogroups. Of these, 1,407 were species-specific and 6,199 were found in all species including 3,365 single-copy genes. The orthogroup profile of *C. hispanica* was overall comparable to that of other ants (Figure 4). However, our annotation had one of the smallest number of genes placed in orthogroups (10,918), and one of the largest proportions of unassigned genes (3.3%).

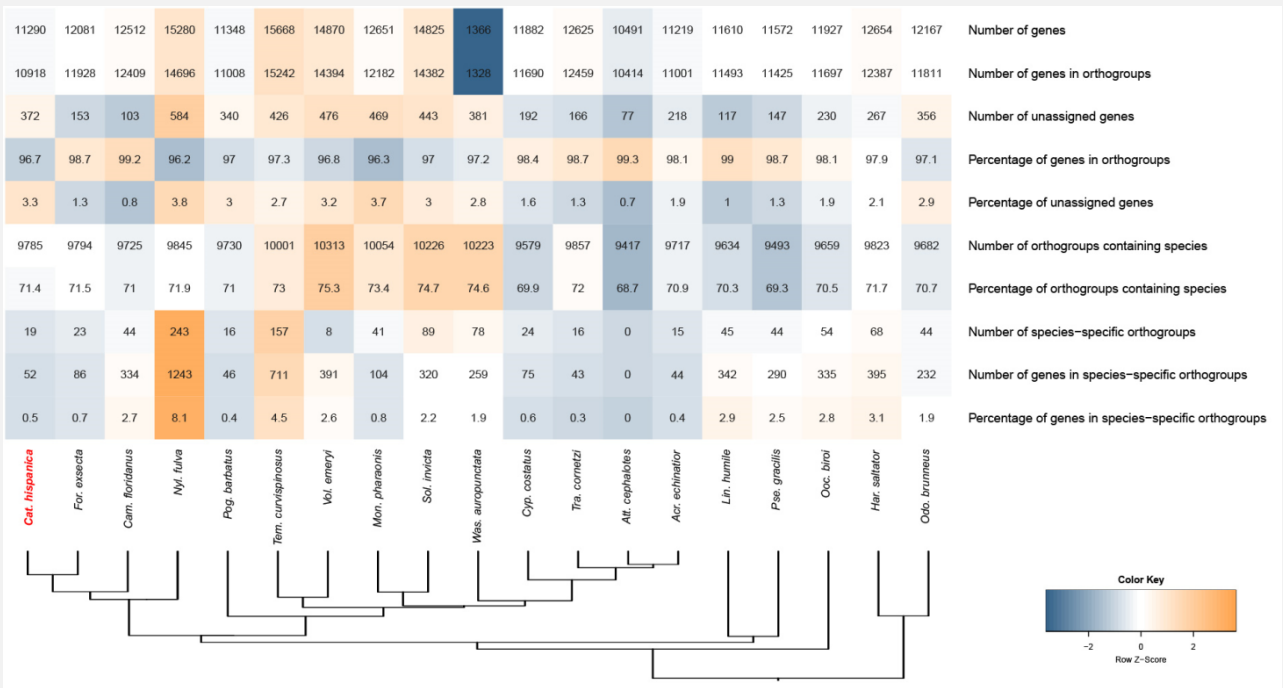
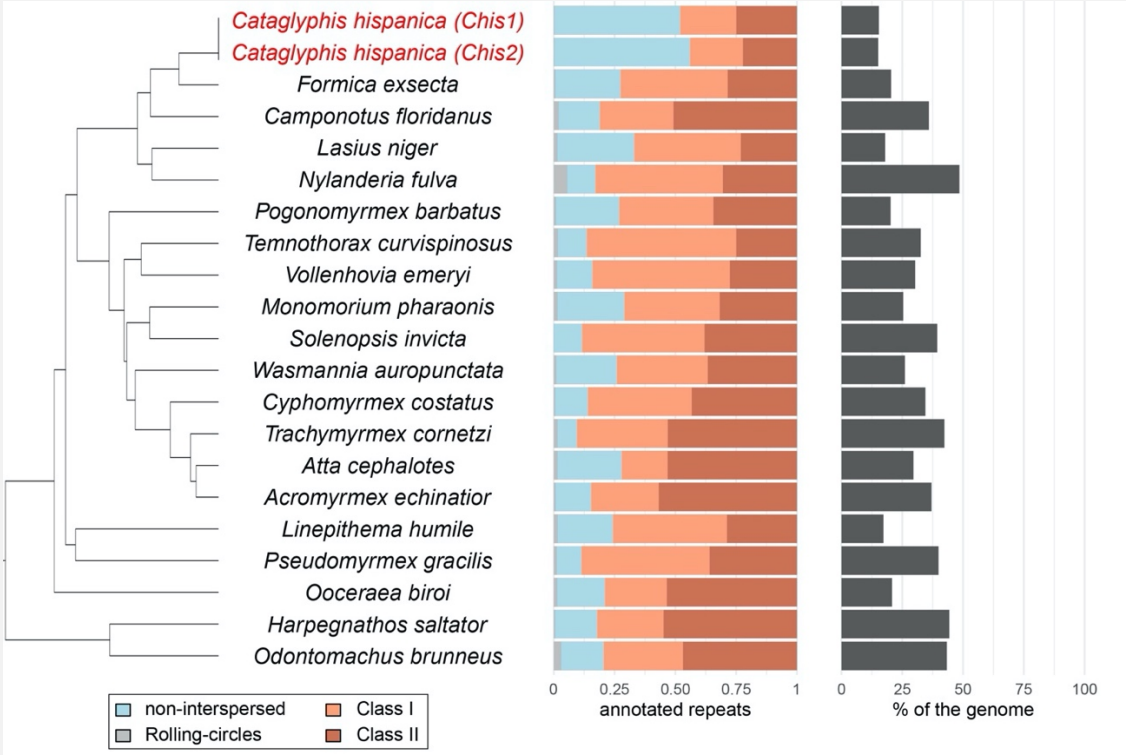


Figure 4. Summary values from the ortholog analyses. The color intensity indicates the z-score of variation (deviation from the mean) among all species, from the smallest value (blue) to the highest value (orange). Species are ordered according to their phylogenetic positions inferred from a concatenated alignment of single-copy orthologs. The published *Lasius niger* assembly was removed from this comparison due to its low quality.

203 **Repeat annotation**

204 We built custom repeat libraries for each of the two assemblies of *C. hispanica* and for the 19
205 published ant genomes (see genome accessions in Table S1). The Chis1 and Chis2 assemblies
206 contained 1,708 and 1,673 different repetitive elements, which accounted for 15.43% (31,851,170
207 bp) and 15.1% (31,512,815 bp) of their assembly sizes, respectively (Figure 5). A large proportion
208 of these corresponded to unclassified interspersed repeats (6.7% / 6.78% of the genomes; Figure
209 S6). The two genomes also contained 2.0% / 1.8% of Class I (retroelements), and 2.18% / 1.85%
210 of Class II elements (DNA transposons). In total, 56 different families of repetitive elements were
211 annotated in *C. hispanica*. LTR/Gypsy were the most frequent transposable elements of Class I in
212 the genomes (0.53% / 0.82%), while large Polintons / Mavericks were the most abundant Class II
213 transposable elements (0.98% / 0.67%).



214

215 **Figure 5.** Summary of the repetitive elements' categories annotated in 20 different ant species using
216 our custom pipeline. The ratios of the major categories of repetitive elements identified in each
217 species is shown on the left. The total proportion of repetitive elements found in each genome is
218 shown on the right. Species are ordered accordingly to their phylogenetic positions inferred from a
219 concatenated alignment of single-copy orthologs.

220 Across published ant assemblies, the total proportion of transposable elements appeared quite
221 variable irrespective of their phylogenetic relationships (range: 17.27 – 48.47%; N= 19 ant species;
222 Figure 5; Table S2). The *C. hispanica* assemblies had smaller proportions of repetitive elements
223 (15.1% - 15.43%) than any of these assemblies, including that of *Formica exsecta* (18.53 %), the
224 closest species available for comparison. The relatively low proportion of transposable elements
225 observed in the genomes of *C. hispanica* may be due to the fact that it was assembled primarily
226 from noisy nanopore long-reads, possibly leading to a collapse of repeated regions. Alternatively,
227 *C. hispanica* may resist the invasion and proliferation of transposable elements more efficiently than
228 other species. Whether its unusual reproductive system, combining both diploid and haploid
229 parthenogenesis for queen and male production, could help keep transposable elements at bay
230 deserves further exploration.
231
232 **Lineage comparison**
233 We previously showed that *Cataglyphis hispanica* consists of two divergent lineages that are readily
234 identifiable using microsatellite markers (Darras *et al.* 2014). Individuals from Chis1 and Chis2
235 lineages can however not be distinguished based on external traits: they share virtually the same
236 morphologies for the queen and male castes, co-occur in the same localities and do not differ in
237 any obvious colony characteristics. Furthermore, although queens mate with a partner originating
238 from the alternative lineage to successfully produce workers, we have no evidence that lineages
239 can recognize each other and avoid assortative mating. This apparent lack of differences among
240 lineages suggests low overall genomic divergence. The interdependent nature of the lineages could
241 stem from a small number of recessive mutations biasing development toward the queen caste in
242 each lineage. Such "royal cheats" (Hughes and Boomsma 2008) seem common in eusocial
243 Hymenoptera and have been hypothesized to be at the origin of caste determination and possibly
244 social hybridogenesis (Anderson *et al.* 2008; Weyna *et al.* 2021; Withrow and Tarpy 2018). In line
245 with this prediction of low inter-lineage genomic divergence, assemblies of the two lineages appear
246 highly similar and syntenic (Figure S3). Large indel (>10kb) variation among lineages account for
247 6.6 % (13.6 Mb, Chis1) and 6.4 % (13.2 Mb, Chis2) of the chromosome-scale scaffolds. These

248 “lineage-specific” indels are scattered across the assemblies (Figure S7) and are gene-deprived;
249 only 35 of the 11,033 (0.3%) genes models from the Chis2 chromosome-scale scaffolds turned
250 missing from Chis1 when performing an annotation lift-over using Liftoff (Table S3). Small inter-
251 lineage polymorphism (i.e. SNPs and indels smaller than 100 bp) also appear uniformly distributed
252 across chromosomes, with no large portion of chromosomes showing elevated divergence among
253 assemblies (Figure S7). This later result contradicts previous hypotheses that hybridogenetic
254 lineage pairs might be determined by ancient non-recombining regions, as found in other dimorphic
255 system such as sex chromosomes or social chromosomes (Darras *et al.* 2014; Linksvayer *et al.*
256 2013; Schwander *et al.* 2014).

257 We additionally estimated divergence between the two genomes sequenced analyzing
258 polymorphism at four-fold-degenerate sites, which are expected to be neutrally evolving since every
259 mutation at a four-fold site is synonymous. Our annotation of the Chis2 genome contained
260 2,620,448 four-fold-degenerate sites. Among these, 13,048 had a different allele in the Chis1 and
261 Chis2 males used to obtain haploid genome consensus. Assuming no recombination and a typical
262 insect mutation rate of approximately 3×10^{-9} mutations per neutral site per haploid genome per
263 generation (Keightley *et al.* 2014, 2015; Yang *et al.* 2015; Liu *et al.* 2017; Oppold & Pfenninger,
264 2017), this proportion of mutated four-fold-degenerate sites translated into an average divergence
265 time of about 830,000 generations between the alleles of the two males sequenced (Obbard *et al.*
266 2012). Hence, the two genomes sequenced may have diverged almost 1 million years ago
267 (assuming one generation per year) - a divergence time similar to that observed between closely
268 related species of fire ants (Cohen & Privman, 2019). The origin of the hybridogenetic lineages
269 themselves could be much younger though, considering they might have emerged from two
270 divergent populations or shared ancestral polymorphism (Darras *et al.* 2019).

271

272 **CONCLUSIONS**

273 We generated high-quality chromosome level genome assemblies of the two lineages of the
274 hybridogenetic ant *C. hispanica*, a representative species of the thermophilic ant genus *Cataglyphis.*
275 Using chromosome conformation capture, we identified a Robertsonian translocation between the

276 two queens sequenced, resulting in 26 and 27 chromosomes, respectively. However, this difference
277 in chromosome numbers seem not fixed between lineages, suggesting that this chromosome
278 rearrangement was not pivotal in the origin and maintenance of social hybridogenesis in *C.*
279 *hispanica*. The two lineage assemblies were overall very similar with no large-scale region showing
280 high divergence. Future work using population genomic approaches and genomic comparisons with
281 other *Cataglyphis* species exhibiting social hybridogenesis will be necessary to identifying
282 polymorphic genes or regulatory regions that are involved in the differentiation of queens and
283 workers during development.
284
285 **METHODS**
286 **Biological samples**
287 Permits were obtained to collect colonies of *Cataglyphis hispanica* in three Spanish locations
288 (Bonares, Caceres and Merida; Figure 1B). Male samples from Bonares were used for Illumina
289 DNA sequencing. Shortly after sampling, the Bonares population was wiped out by human activities.
290 Consequently, samples from another locality (Caceres) were used for subsequent Nanopore
291 sequencing, 3C-seq and RNA-seq. Male and worker pupae from two distant localities (Bonares and
292 Merida) were used for karyotyping. Twelve diagnostic microsatellite loci were genotyped prior to
293 sequencing and karyotyping to assess the lineage membership of each queen and male and to
294 confirm that workers were all first generation hybrids (Darras *et al.* 2014).
295
296 **DNA and RNA-Sequencing**
297 Genomic resources were generated for both the Chis1 and the Chis2 lineages. High-molecular-
298 weight DNA was extracted from pure lineage queen and male individuals using QIAGEN Genomic-
299 tips. For each lineage, two queen clones originating from the same nest were used for Nanopore
300 sequencing. Queens of *C. hispanica* are produced through automictic parthenogenesis with central
301 fusion which results into diploid individuals that are highly homozygous (Darras *et al.* 2014; Pearcy
302 *et al.* 2011) and thus suitable for genome assembly. Nanopore libraries were prepared using rapid
303 sequencing kits (SQK-RAD001 and SQK-RAD004). The resulting long read libraries were

304 sequenced on MIN106 flow cells and basecalled using Albacore v2.1.10. For each lineage, three
305 Illumina libraries were generated from whole-genome amplified DNA extracted from a single male
306 with mean insert sizes of 170 bp, 500 bp and 800 bp, and sequenced with a HiSeq2000 (paired-
307 end 2 x 100 bp mode).

308 3C-seq libraries were prepared according to the protocol described in (Marie-Nelly *et al.* 2014).
309 Briefly, queens from both lineages had their gut removed and were immediately suspended in 30
310 mL of formaldehyde solution (Sigma Aldrich; 3% final concentration in 1X tris-EDTA buffer). After
311 one hour of incubation, quenching of the remaining formaldehyde was done by adding 10 mL of
312 glycine (0.25 M final concentration) to the mix for during 20 min. The cross-linked tissues were
313 pelleted and stored at −80°C until further use. The 3C-seq libraries were prepared using the *Dpn*II
314 enzyme and sequenced using an Illumina NextSeq 500 apparatus (paired-end 2×75 bp; first ten
315 bases corresponding to custom-made tags). 3C-seq libraries are similar to Hi-C libraries except that
316 they contain a higher percentage of paired-end reads due to the lack of an enrichment step (Flot *et*
317 *al.* 2015).

318 To help annotate the genomes, three normalized RNA-seq cDNA Illumina libraries were obtained:
319 one from an adult Chis1 queen, one from a Chis2 queen and one from a brood pool comprising
320 multiple developmental stages and adult workers originated from colonies of the two lineages
321 (HiSeq2000, paired-end 2 x 100 bp mode).

322

323 **Genomes assembly**

324 The genome of each hybridogenetic lineage was assembled independently following the pipeline
325 depicted in Figure S1. Nanopore data were assembled using Flye v2.7 with four iterations of
326 polishing based on long reads (Kolmogorov *et al.* 2019). Raw Illumina reads were trimmed for
327 quality and adapters were removed using Trimmomatic v0.32 with options
328 ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
329 MINLEN:36 (Bolger *et al.* 2014). The trimmed reads were then aligned to the long-read assemblies
330 using BWA-MEM v0.7.15 (Li and Durbin 2009). SNPs and indels with at least three supporting

331 observations were called using freebayes v1.2 (Garrison and Marth 2012), and error-corrected
332 consensus sequences were obtained using BCFtools v1.4 (Li *et al.* 2009).
333 To obtain chromosome-scale assemblies, we scaffolded the polished contigs with the 3C reads
334 using instaGRAAL, a MCMC, proximity-ligation based scaffolder (Baudry *et al.* 2020; Marie-Nelly
335 *et al.* 2014). The 3C reads were trimmed using cutadapt (Martin 2011) and subsequently processed
336 using hicstuff (https://zenodo.org/record/4722873) with the following parameters *–aligner bowtie2 –*
337 *iterative –enzyme DpnII*. The instaGRAAL scaffolder was run on the pre-processed data for 100
338 cycles (parameters: level 4, with options *--coverage-std 1 –level 4 –cycles 100*) (Baudry *et al.* 2020)
339 and final scaffolds were obtained using the instaGRAAL *-polish* script, with all corrective procedures
340 at once (only one parameter: *-m polish*). Briefly, instaGRAAL explores the chromosome structures
341 by testing the relative positions and/or orientations of DNA segments (or bins) according to the
342 contacts expected given a simple three-parameter power-law model. These modifications take the
343 form of a fixed set of operations (swapping, flipping, inserting, merging, etc.) of bins corresponding
344 to 3^4 = 81 *Dpn*II restriction fragments. The likelihood of the model is then maximized by sampling
345 the parameters using a MCMC approach (Marie-Nelly *et al.* 2014). After 100 iterations (i.e., a likely
346 position for each bin is tested 100 times), the genome structure converges towards a relatively
347 stable structure that does not evolve anymore when more iterations are added, resulting in
348 chromosome-level scaffolds. The algorithm is probabilistic and ignores initially part of the intrinsic
349 structure of the original contigs in order to sample a larger range of genome space (Baudry *et al.*
350 2020). Therefore, some trustworthy information contained in the initial polished assembly can be
351 lost, or modified, along the way. The final correction step of instaGRAAL consists in reintegrating
352 this lost information into the final assembly, to correct for instance local untrustworthy tiny inversions
353 of individual bins within a contig. The contact maps of the scaffolded assemblies were built using
354 hicstuff. Gaps created during the scaffolding process were closed using Nanopore data with four
355 iterations of TGS-GapCloser (Xu *et al.* 2019) and new polished consensus sequences were
356 obtained using BCFtools (see method above). Completeness of the assemblies were assessed at
357 each step using BUSCO v5.2.2 with the Hymenoptera odb10 lineage (Simão *et al.* 2015;
358 Waterhouse *et al.* 2017). We also ran KAT v2.4.1 to compare the k-mer frequencies of Illumina

359 reads to final assemblies (Mapleson *et al.* 2016). To investigate differences in chromosomal
360 arrangement among lineages, the two genome assemblies were aligned with minimap2 v2.17
361 (exact preset: -x asm5) and alignments were visualized using dot plots obtained with D-GENIES
362 (Cabanettes & Klopp, 2018).
363
364 **Karyotyping**
365 To validate the number of chromosomes inferred from 3C contact information, chromosome
366 preparations were made from brains of male and worker larvae following the protocol described by
367 (Lorite *et al.* 1996), with some modifications. Briefly, larvae at the last instar stage were dissected
368 and their cerebral ganglia were transferred to microplate wells with 0.05% colchicine in distilled
369 water. After 30 min, samples were transferred to a fixative solution (acetic acid:ethanol, 3:1) and
370 incubated for 45 min. Ganglia cells were disaggregated in a drop of 50% acetic acid on a clean
371 slide, new fixative solution was added and the slides were dried at 60ºC. Chromosome preparations
372 were stained with 10% Giemsa in phosphate buffer (pH 7). Microscopy images were captured with
373 a CCD camera (Olympus DP70) coupled to a microscope (Olympus BX51) and were processed
374 using Adobe Photoshop.
375
376 **Gene annotation**
377 We used the Chis2 chromosome-level assembly for gene annotation. A repeat library was
378 constructed using the REPET package v2.5 (Flutre *et al.* 2011; Quesneville *et al.* 2005). This repeat
379 library was cleaned up manually to remove bacterial genes, mitochondrial genes and genes with
380 hits to the gene set of the ant *Cardiocondyla obscurior* (v1.4) which had been purged of
381 transposable elements (Schrader *et al.* 2014). The fraction of the genome classified by
382 RepeatClassifier as "Unknown" was reduced from 2.2% to 0.9% as a result of this procedure.
383 Repeats were soft-masked using RepeatMasker v4.0.7 (Smit and Hubley,
384 http://www.repeatmasker.org) prior to *de novo* gene prediction.
385 Gene models were inferred from RNA-seq, homology data and *ab initio* predictions. The three
386 RNA-seq libraries were aligned to the *Chis2* genome using STAR v2.6.0 (Dobin *et al.* 2013) with

387 the multi-sample 2-pass mapping strategy. Transcripts were then assembled using Trinity v2.10.0
388 (Grabherr *et al.* 2011; Haas *et al.* 2013)(options --genome_guided_max_intron 100000 --
389 jaccard_clip) and combined into gene models using PASA (Haas *et al.* 2003). Ant proteomes
390 annotated using the NCBI Eukaryotic Genome Annotation pipeline (RefSeq, taxid:36668) were
391 aligned to the genome using GenomeThreader v1.5.10 (Gremme *et al.* 2005) in order to predict
392 gene structures. AUGUSTUS *ab initio* predictions were generated using BRAKER v2.1.02 (Hoff *et*
393 *al.* 2016, 2019) based on hints from RNA-seq data and GenomeThreader protein alignments (--
394 etpmode). BRAKER was first run with preliminary AUGUSTUS parameters trained by running
395 BUSCO v3.0.2 on the genome assembly (--long option; Hymenoptera odb9 database). To refine
396 the training of AUGUSTUS, the most accurate gene models inferred by BRAKER were then
397 identified using GeneValidator (Drăgan *et al.* 2016) with RefSeq ant proteomes as references and
398 an arbitrary quality threshold of Q89. To avoid biases, predicted proteins with more than 70%
399 sequence identity to another protein in the set were removed from the selected gene models using
400 the aa2nonred.pl script provided with BRAKER. The resulting gene models were used to train
401 AUGUSTUS again, and BRAKER was run with the new parameter set. *Ab initio*, RNA-seq-based
402 and homology-based gene predictions were combined into a single gene set using
403 EvidenceModeler v1.1.1 (Haas *et al.* 2008) with the following weight settings: PASA alignments:
404 10; GenomeThreader alignments: 3, Augustus predictions: 1, PASA/Transdecoder predictions: 1,
405 GenomeThreader predictions: 1. Functional information was obtained from eggNOG-mapper v2
406 (Huerta-Cepas *et al.* 2017, 2019) with the options "taxonomic scope adjusted per query" and
407 "annotations transferred from any ortholog". Protein sequences with similarity to RefSeq ant
408 proteins (as of July 2019) were identified using blastp and an E-value threshold of 10^{-5}. Annotations
409 with no known functional information and no hits to any RefSeq ant proteins were filtered out.
410

411 **Comparative analyses**

412 To identify orthologous and taxonomically restricted genes, we compared the proteomes of *C.*
413 *hispanica*, of 18 ants annotated by the NCBI Eukaryotic Genome Annotation Pipeline (Table S1)
414 and of *Lasius niger* (Konorov *et al.* 2017) using OrthoFinder v2.3.12 (Emms and Kelly, 2019) with

415 its standard DEndroBLAST workflow. We used the feature annotation tables from RefSeq
416 annotations to select the longest isoform of each gene annotated by NCBI prior to analysis. The
417 published genome of *L. Niger* is highly incomplete (no more than 65% of the 4,415 highly conserved
418 single-copy genes of BUSCO's Hymenoptera odb9 database are found in this assembly).
419 Consequently, it was only used to guide phylogenetic analyses due to its relative proximity with
420 *Cataglyphis.* A preliminary catalog of single-copy orthologs was obtained from a first run of
421 OrthoFinder. Single-copy sequences were aligned with Mafft v7.310 (Katoh and Standley, 2013)
422 and the alignments were trimmed with trimAL v1.4.1(options "-gt 0.8 -st 0.001") (Capella-Gutiérrez
423 *et al.* 2009). The concatenated alignments were then passed to IQ-TREE v1.7.17 (option "-m
424 LG+R4") (Nguyen et al. 2015) to infer a species tree. The tree was converted to an ultrametric
425 topology with the r8s program with options "mrca root Obir Hsal; fixage taxon=root age=150; divtime
426 method=LF algorithm=TN" (Sanderson 2003). The resulting species tree was used for a second,
427 more precise run of OrthoFinder.
428
429 **Repeat annotation**
430 To compare the frequency of repetitive elements found in the genome of *C. hispanica* to the
431 frequencies found in the genomes of other ant species available (Table S2), we constructed
432 optimized repeat libraries for each species using a custom pipeline
433 (https://github.com/nat2bee/repetitive_elements_pipeline). Shortly, repeat libraries were built with
434 RepeatModeler v1.0.11 (http://www.repeatmasker.org/RepeatModeler/), TransposonPSI
435 (http://transposonpsi.sourceforge.net/) and LTRharvest from GenomeTools v1.6.1 (Ellinghaus *et al.*
436 2008). For each species, the different libraries were merged into a non-redundant library (<80%
437 identity) using USEARCH v11.0.667 (Edgar 2010). Library annotations were obtained with
438 RepeatClassifier. Each custom library was concatenated with the Dfam v3.1 Hymenoptera library
439 of RepeatMasker v4.1.0 and used to annotate repeats in the genome of the corresponding species
440 using RepeatMasker. Summary statistics of the annotated repeats were obtained with
441 RepeatMasker_stats.py (https://github.com/nat2bee/repetitive_elements_pipeline).
442

443 **Lineage comparison**

444 The two assemblies were aligned with minimap v2.19 (-cx asm5 –cs) and variants were called with
445 paftools (paftools.js call -L5000 -l1000). The distribution of large indels (>10 kb) and the density of
446 small polymorphisms (SNPs and indels no larger than 100 bp) across the genomes were calculated
447 using custom scripts. Annotation lift-over from the Chis2 assembly on to the Chis1 assembly was
448 performed with Liftoff v1.6.3 (Shumate and Salzberg 2020). To verify if missing annotations did not
449 result from misassemblies, we also lift these on a consensus Chis1 assembly derived from
450 alignment of the Chis1 haploid short reads on the Chis2 assemblies using BCFtools as described
451 above (see Genomes assembly) with regions not covered by reads masked to avoid reference bias
452 (--mask --mask-with N).

453 To estimate the divergences of the two lineages of *C. hispanica*, we investigated the
454 polymorphism at 4-fold-degenerate sites, which we assumed to be neutrally evolving. The Illumina
455 read of the Chis1 lineage were mapped onto the Chis2 reference genome and single-nucleotide
456 variants were called using MapCaller v0.9.9.41 (Lin and Hsu 2019). The resulting vcf file was filtered
457 to keep only single-nucleotide variants with two alleles and a 'PASS' quality filter. To determine the
458 proportion of 4-fold sites that were polymorphic among our male samples of the two lineages, the
459 positions of 4-fold sites in coding sequences of our annotation were identified using a custom script
460 (T. Sackton, github.com/tsackton/linked-selection.git).

461

462 **DECLARATIONS**

463 **Data Availability**

464 All the raw sequencing data and genome assemblies generated during this study have been
465 deposited at NCBI (Accession numbers: SRR17481978 - SRR17481992). The genomes of *C.*
466 *hispanica* were deposited in NCBI (Accession numbers: JAJUXC000000000 and
467 JAJUXE000000000). Supplementary figures, tables, gene annotations, TE repeat libraries and
468 reports can be accessed at figshare (https://doi.org/10.6084/m9.figshare.17964695.v7).

469

Funding

NSA and SA are supported by the Belgian Fonds National pour la Recherche Scientifique (FRS-FNRS). NG was supported by the Horizon 2020 research and innovation program of the European Union under the Marie Skłodowska-Curie grant agreement No. 764840 (ITN IGNITE, www.itn-ignite.eu) to JFF. This project was funded by the FRS-FNRS Grants # J.0151.16 and T.0140.18 (to SA). HD received financial support from the Jean-Marie Delwart Foundation. Computational resources were provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Belgian Fund for Scientific Research-FNRS (F.R.S.-FNRS; grant No. 2.5020.11).

Authors' Contributions

HD collected the ants, prepared DNA/RNA, assembled and annotated the genome. NSA performed TE analyses and genomic comparisons. Both HD and NSA prepared first manuscript draft. PL performed karyotyping. LB optimized 3C scaffolding parameters. NG performed 3C scaffolding. MM prepared the 3C libraries. FR constructed the TE library. IA supervised the construction of the TE library. RK supervised 3C library generation and scaffolding. JFF supervised genome assembly. SA collected the ants, designed and supervised the study. All authors read and approved the final manuscript.

Acknowledgments

We thank A. Cohanim, E. Privman and R. Faure for their advices on early genome assemblies, F. Rodriguez and L. Grumiau for their assistance with genotyping and Q. M. Pan for her comments on the manuscript.

Conflict of interest disclosure

The authors of this article declare that they have no financial conflict of interest with the content of this article. J.F. Flot is one of the PCI Genomics recommenders.

498 **REFERENCES**

499 Anderson, Kirk E., Timothy A. Linksvayer, and Chris R. Smith. 2008. "The Causes and
500 Consequences of Genetic Caste Determination in Ants (Hymenoptera: Formicidae)."
501 *Myrmecological News* 11: 119–32

502 Aron, Serge, Pascale Lybaert, Claire Baudoux, Morgane Vandervelden, and Denis Fournier. 2016a.
503 "Sperm Production Characteristics Vary with Level of Sperm Competition in *Cataglyphis*
504 Desert Ants." *Functional Ecology* 30 (4): 614–24.

505 Aron, Serge, Patrick Mardulyn, and Laurianne Leniaud. 2016b. "Evolution of Reproductive Traits in
506 *Cataglyphis* Desert Ants: Mating Frequency, Queen Number, and Thelytoky." *Behavioral*
507 *Ecology and Sociobiology* 70 (8): 1367–79.

508 Aron, Serge, and Rüdiger Wehner. 2021. "Cataglyphis." In *Encyclopedia of Social Insects*, 217–23.
509 Cham: Springer International Publishing.

510 Baudry, Lyam, Nadège Guiglielmoni, Hervé Marie-Nelly, Alexandre Cormier, Martial Marbouty,
511 Komlan Avia, Yann Loe Mie, *et al.* 2020. "InstaGRAAL: Chromosome-Level Quality
512 Scaffolding of Genomes Using a Proximity Ligation-Based Scaffolder." *Genome Biology* 21
513 (1): 148.

514 Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for
515 Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20.

516 Boulay, Raphaël, Serge Aron, Xim Cerdá, Claudie Doums, Paul Graham, Abraham Hefetz, and
517 Thibaud Monnin. 2017. "Social Life in Arid Environments: The Case Study of *Cataglyphis*
518 Ants." *Annual Review of Entomology* 62 (January): 305–21.

519 Cabanettes, Floréal, and Christophe Klopp. 2018. "D-GENIES: Dot Plot Large Genomes in an
520 Interactive, Efficient and Simple Way." *PeerJ* 6 (June): e4958.

521 Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. "TrimAl: A Tool for
522 Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25
523 (15): 1972–73.

524 Cerda, X., J. Retana, and S. Cros. 1998. “Critical Thermal Limits in Mediterranean Ant Species:
525 Trade-off between Mortality Risk and Foraging Performance.” *Functional Ecology* 12 (1):
526 45–55.
527 Cohen, Pnina, and Eyal Privman. 2019. “Speciation and Hybridization in Invasive Fire Ants.” *BMC*
528 *Evolutionary Biology* 19 (1): 111.
529 Darras, Hugo, Alexandre Kuhn, and Serge Aron. 2019. “Evolution of Hybridogenetic Lineages in
530 *Cataglyphis* Ants.” *Molecular Ecology* 28 (12): 3073–88.
531 Darras, Hugo, Laurianne Leniaud, and Serge Aron. 2014. “Large-Scale Distribution of
532 Hybridogenetic Lineages in a Spanish Desert Ant.” *Proceedings. Biological Sciences / The*
533 *Royal Society* 281 (1774): 20132396.
534 Davisson, M. T., and E. C. Akeson. 1993. “Recombination Suppression by Heterozygous
535 Robertsonian Chromosomes in the Mouse.” *Genetics* 133 (3): 649–67.
536 Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha,
537 Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal
538 RNA-Seq Aligner.” *Bioinformatics* 29 (1): 15–21.
539 Drăgan, Monica-Andreea, Ismail Moghul, Anurag Priyam, Claudio Bustos, and Yannick Wurm.
540 2016. “GeneValidator: Identify Problems with Protein-Coding Gene Predictions.”
541 *Bioinformatics* 32 (10): 1559–61.
542 Edgar, Robert C. 2010. “Search and Clustering Orders of Magnitude Faster than BLAST.”
543 *Bioinformatics* 26 (19): 2460–61.
544 Eyer, P. A., L. Leniaud, H. Darras, and S. Aron. 2013. “Hybridogenesis through Thelytokous
545 Parthenogenesis in Two *Cataglyphis* Desert Ants.” *Molecular Ecology* 22 (4): 947–55.
546 Ellinghaus, David, Stefan Kurtz, and Ute Willhoeft. 2008. “LTRharvest , an Efficient and Flexible
547 Software for de Novo Detection of LTR Retrotransposons.” *BMC Bioinformatics* 9 (1): 1–14.
548 Emms, David M., and Steven Kelly. 2019. “OrthoFinder: Phylogenetic Orthology Inference for
549 Comparative Genomics.” *Genome Biology* 20 (1): 238.

550 Faria, Rui, and Arcadi Navarro. 2010. “Chromosomal Speciation Revisited: Rearranging Theory with
551 Pieces of Evidence.” *Trends in Ecology & Evolution* 25 (11): 660–69.
552 Flot, Jean-François, Hervé Marie-Nelly, and Romain Koszul. 2015. “Contact Genomics: Scaffolding
553 and Phasing (Meta)Genomes Using Chromosome 3D Physical Signatures.” *FEBS Letters*
554 589 (20 Pt A): 2966–74.
555 Flutre, Timothée, Elodie Duprat, Catherine Feuillet, and Hadi Quesneville. 2011. “Considering
556 Transposable Element Diversification in de Novo Annotation Approaches.” *PloS One* 6 (1):
557 e16526.
558 Garrison, Erik, and Gabor Marth. 2012. “Haplotype-Based Variant Detection from Short-Read
559 Sequencing.” *ArXiv [q-Bio.GN]*. arXiv. http://arxiv.org/abs/1207.3907.
560 Gehring, W. J., and R. Wehner. 1995. “Heat Shock Protein Synthesis and Thermotolerance in
561 Cataglyphis, an Ant from the Sahara Desert.” *Proceedings of the National Academy of*
562 *Sciences of the United States of America* 92 (7): 2994–98.
563 Grabherr, Manfred G., Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido
564 Amit, Xian Adiconis, *et al.* 2011. “Full-Length Transcriptome Assembly from RNA-Seq Data
565 without a Reference Genome.” *Nature Biotechnology*. https://doi.org/10.1038/nbt.1883.
566 Gremme, Gordon, Volker Brendel, Michael E. Sparks, and Stefan Kurtz. 2005. “Engineering a
567 Software Tool for Gene Structure Prediction in Higher Organisms.” *Information and Software*
568 *Technology* 47 (15): 965–78.
569 Haas, Brian J., Arthur L. Delcher, Stephen M. Mount, Jennifer R. Wortman, Roger K. Smith Jr, Linda
570 I. Hannick, Rama Maiti, *et al.* 2003. “Improving the Arabidopsis Genome Annotation Using
571 Maximal Transcript Alignment Assemblies.” *Nucleic Acids Research* 31 (19): 5654–66.
572 Haas, Brian J., Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D. Blood, Joshua
573 Bowden, Matthew Brian Couger, *et al.* 2013. “De Novo Transcript Sequence Reconstruction
574 from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis.” *Nature*
575 *Protocols* 8 (8): 1494–1512.

576 Haas, Brian J., Steven L. Salzberg, Wei Zhu, Mihaela Pertea, Jonathan E. Allen, Joshua Orvis,
577 Owen White, C. Robin Buell, and Jennifer R. Wortman. 2008. "Automated Eukaryotic Gene
578 Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced
579 Alignments." *Genome Biology* 9 (1): R7.
580 Hauschteck-Jungen, E., and H. Jungen. 1983. "Ant Chromosomes." *Insectes Sociaux* 30 (2): 149–
581 64.
582 Hoff, Katharina J., Simone Lange, Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke.
583 2016. "BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET
584 and AUGUSTUS: Table 1." *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btv661.
585 Hoff, Katharina J., Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2019. "Whole-
586 Genome Annotation with BRAKER." *Methods in Molecular Biology*.
587 https://doi.org/10.1007/978-1-4939-9173-0_5.
588 Huerta-Cepas, Jaime, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl
589 Jensen, Christian von Mering, and Peer Bork. 2017. "Fast Genome-Wide Functional
590 Annotation through Orthology Assignment by EggNOG-Mapper." *Molecular Biology and*
591 *Evolution* 34 (8): 2115–22.
592 Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund,
593 Helen Cook, Daniel R. Mende, *et al.* 2019. "EggNOG 5.0: A Hierarchical, Functionally and
594 Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502
595 Viruses." *Nucleic Acids Research* 47 (D1): D309–14.
596 Hughes, William O. H., and Jacobus J. Boomsma. 2008. "Genetic Royal Cheats in Leaf-Cutting Ant
597 Societies." *Proceedings of the National Academy of Sciences of the United States of*
598 *America* 105 (13): 5150–53.
599 Imai, B. Y. H. T., C. Baroni Urbani, M. Kubota, G. P. Sharma, M. N. Narasimhanna, B. C. Das, A.
600 K. Sharma, *et al.* 1984. "Karyological Survey of Indian Ants." *The Japanese Journal of*
601 *Genetics*. https://doi.org/10.1266/jjg.59.1.

602 Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software
603 Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30
604 (4): 772–80.
605 Keightley, Peter D., Rob W. Ness, Daniel L. Halligan, and Penelope R. Haddrill. 2014. “Estimation
606 of the Spontaneous Mutation Rate per Nucleotide Site in a Drosophila Melanogaster Full-
607 Sib Family.” *Genetics* 196 (1): 313–20.
608 Keightley, Peter D., Ana Pinharanda, Rob W. Ness, Fraser Simpson, Kanchon K. Dasmahapatra,
609 James Mallet, John W. Davey, and Chris D. Jiggins. 2015. “Estimation of the Spontaneous
610 Mutation Rate in Heliconius Melpomene.” *Molecular Biology and Evolution* 32 (1): 239–43.
611 Kolmogorov, Mikhail, Jeffrey Yuan, Yu Lin, and Pavel A. Pevzner. 2019. “Assembly of Long, Error-
612 Prone Reads Using Repeat Graphs.” *Nature Biotechnology* 37 (5): 540–46.
613 Konorov, Evgenii A., Mikhail A. Nikitin, Kirill V. Mikhailov, Sergey N. Lysenkov, Mikhail Belenky,
614 Peter L. Chang, Sergey V. Nuzhdin, and Victoria A. Scobeyeva. 2017. “Genomic Exaptation
615 Enables Lasius Niger Adaptation to Urban Environments.” *BMC Evolutionary Biology* 17
616 (Suppl 1): 39.
617 Kuhn, Alexandre, Hugo Darras, Omid Paknia, and Serge Aron. 2020. “Repeated Evolution of Queen
618 Parthenogenesis and Social Hybridogenesis in *Cataglyphis* Desert Ants.” *Molecular Ecology*
619 29 (3): 549–64.
620 Leniaud, Laurianne, Hugo Darras, Raphael Boulay, and Serge Aron. 2012. “Social Hybridogenesis
621 in the Clonal Ant Cataglyphis Hispanica.” *Current Biology: CB* 22 (13): 1188–93.
622 Lenoir, A., E. Nowbahari, L. Querard, N. Pondicq, and C. Delalande. 1990. “Habitat Exploitation and
623 Intercolonial Relationships in the Ant *Cataglyphis* Cursor (Hymenoptera: Formicidae).” *Acta*
624 *Oecologica* 11 (1): 3–18.
625 Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-
626 Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.

Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, *et al.* 2009. “Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.” *Science (New York, N.Y.)* 326 (5950): 289–93.

Linksvayer, Timothy A., Jeremiah W. Busch, and Chris R. Smith. 2013. “Social Supergenes of Superorganisms: Do Supergenes Play Important Roles in Social Evolution?” *BioEssays* 35 (8): 683–89.

Liu, Haoxuan, Yanxiao Jia, Xiaoguang Sun, Dacheng Tian, Laurence D. Hurst, and Sihai Yang. 2017. “Direct Determination of the Mutation Rate in the Bumblebee Reveals Evidence for Weak Recombination-Associated Mutation and an Approximate Rate Constancy in Insects.” *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msw226.

Lorite, P., E. Chica, and T. Palomeque. 1996. “Cytogenetic Studies of Ant Linepithema Humile Shattuck (= Iridomyrmex Humilis Mayr) in European Populations.” *Caryologia* 49 (2): 199–205.

Lorite, Pedro, and Teresa Palomeque. 2010. “Karyotype Evolution in Ants (Hymenoptera: Formicidae), with a Review of the Known Ant Chromosome Numbers.” *Myrmecological News / Osterreichische Gesellschaft Fur Entomofaunistik* 13 (1): 89–102.

Manni, Mosè, Matthew R. Berkeley, Mathieu Seppey, Felipe A. Simão, and Evgeny M. Zdobnov. 2021. “BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.” *Molecular Biology and Evolution* 38 (10): 4647–54.

Mapleson, Daniel, Gonzalo Garcia Accinelli, George Kettleborough, Jonathan Wright, and Bernardo J. Clavijo. 2016. “KAT: A K-Mer Analysis Toolkit to Quality Control NGS Datasets and Genome Assemblies.” *Bioinformatics*, October, btw663.

654 Marie-Nelly, Hervé, Martial Marbouty, Axel Cournac, Jean-François Flot, Gianni Liti, Dante Poggi
655 Parodi, Sylvie Syan, *et al.* 2014. "High-Quality Genome (Re)Assembly Using Chromosomal
656 Contact Data." *Nature Communications* 5 (December): 5695.
657 Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing
658 Reads." *EMBnet.Journal* 17 (1): 10–12.
659 Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE:
660 A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies."
661 *Molecular Biology and Evolution* 32 (1): 268–74.
662 Obbard, Darren J., John Maclennan, Kang-Wook Kim, Andrew Rambaut, Patrick M. O'Grady, and
663 Francis M. Jiggins. 2012. "Estimating Divergence Dates and Substitution Rates in the
664 Drosophila Phylogeny." *Molecular Biology and Evolution* 29 (11): 3459–73.
665 Oppold, Ann-Marie, and Markus Pfenninger. 2017. "Direct Estimation of the Spontaneous Mutation
666 Rate by Short-Term Mutation Accumulation Lines in Chironomus Riparius." *Evolution Letters*
667 1 (2): 86–92.
668 Pearcy, O. J. Hardy, and S. Aron. 2011. "Automictic Parthenogenesis and Rate of Transition to
669 Homozygosity." *Heredity* 107 (2): 187–88.
670 Pearcy, Morgan, Serge Aron, Claudie Doums, and Laurent Keller. 2004. "Conditional Use of Sex
671 and Parthenogenesis for Worker and Queen Production in Ants." *Science* 306 (5702): 1780–
672 83.
673 Peeters, Christian, and Serge Aron. 2017. "Evolutionary Reduction of Female Dispersal in
674 *Cataglyphis* Desert Ants." *Biological Journal of the Linnean Society. Linnean Society of*
675 *London* 122 (1): 58–70.
676 Perez, Rémy, and Serge Aron. 2020. "Adaptations to Thermal Stress in Social Insects: Recent
677 Advances and Future Directions." *Biological Reviews of the Cambridge Philosophical*
678 *Society* 95 (6): 1535–53.

679 Perez, Rémy, Natalia de Souza Araujo, Matthieu Defrance, and Serge Aron. 2021. “Molecular
680 Adaptations to Heat Stress in the Thermophilic Ant Genus *Cataglyphis*.” *Molecular Ecology*
681 30 (21): 5503–16.
682 Pfeffer, Sarah Elisabeth, Verena Luisa Wahl, Matthias Wittlinger, and Harald Wolf. 2019. “High-
683 Speed Locomotion in the Saharan Silver Ant,.” *The Journal of Experimental Biology* 222 (Pt
684 20). https://doi.org/10.1242/jeb.198705.
685 Quesneville, Hadi, Casey M. Bergman, Olivier Andrieu, Delphine Autard, Danielle Nouaud, Michael
686 Ashburner, and Dominique Anxolabehere. 2005. “Combined Evidence Annotation of
687 Transposable Elements in Genome Sequences.” *PLoS Computational Biology* 1 (2): 166–
688 75.
689 Sanderson, Michael J. 2003. “R8s: Inferring Absolute Rates of Molecular Evolution and Divergence
690 Times in the Absence of a Molecular Clock.” *Bioinformatics* 19 (2): 301–2.
691 Schrader, Lukas, Jay W. Kim, Daniel Ence, Aleksey Zimin, Antonia Klein, Katharina Wyschetzki,
692 Tobias Weichselgartner, *et al.* 2014. “Transposable Element Islands Facilitate Adaptation to
693 Novel Environments in an Invasive Species.” *Nature Communications*.
694 Schwander, Tanja, Romain Libbrecht, and Laurent Keller. 2014. “Supergenes and Complex
695 Phenotypes.” *Current Biology* 24 (7): 288–94.
696 Shumate, Alaina, and Steven L. Salzberg. 2020. “Liftoff: Accurate Mapping of Gene Annotations.”
697 *Bioinformatics* 37 (12): 1639–1643.
698 Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny
699 M. Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness
700 with Single-Copy Orthologs.” *Bioinformatics* 31 (19): 3210–12.
701 Sommer, Stefan, and Rüdiger Wehner. 2012. “Leg Allometry in Ants: Extreme Long-Leggedness in
702 Thermophilic Species.” *Arthropod Structure & Development* 41 (1): 71–77.
703 Waterhouse, Robert M., Mathieu Seppey, Felipe A. Simão, Mosè Manni, Panagiotis Ioannidis,
704 Guennadi Klioutchnikov, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2017. “BUSCO

705 Applications from Quality Assessments to Gene Prediction and Phylogenomics." *Molecular*
706 *Biology and Evolution*, December. https://doi.org/10.1093/molbev/msx319.
707 Wehner, R., A. C. Marsh, and S. Wehner. 1992. "Desert Ants on a Thermal Tightrope." *Nature* 357
708 (6379): 586–87.
709 Weyna, Arthur, Jonathan Romiguier, and Charles Mullon. 2021. "Hybridization Enables the Fixation
710 of Selfish Queen Genotypes in Eusocial Colonies." *Evolution Letters* 5 (6): 582–94.
711 Willot, Quentin, Cyril Gueydan, and Serge Aron. 2017. "Proteome Stability, Heat Hardening and
712 Heat-Shock Protein Expression Profiles in Desert Ants." *The Journal of Experimental Biology*
713 220 (Pt 9): 1721–28.
714 Withrow, James M., and David R. Tarpy. 2018. "Cryptic 'Royal' Subfamilies in Honey Bee (Apis
715 Mellifera) Colonies." *PloS One* 13 (7): e0199124.
716 Xu, Mengyang, Lidong Guo, Shengqiang Gu, Ou Wang, Rui Zhang, Guangyi Fan, Xun Xu, Li Deng,
717 and Xin Liu. 2019. "TGS-GapCloser: Fast and Accurately Passing through the Bermuda in
718 Large Genome Using Error-Prone Third-Generation Long Reads." *BioRxiv*. bioRxiv.
719 https://doi.org/10.1101/831248.
720 Yahav, Tal, and Eyal Privman. 2019. "A Comparative Analysis of Methods for de Novo Assembly of
721 Hymenopteran Genomes Using Either Haploid or Diploid Samples." *Scientific Reports* 9 (1).
722 https://doi.org/10.1038/s41598-019-42795-6.
723 Yang, Sihai, Long Wang, Ju Huang, Xiaohui Zhang, Yang Yuan, Jian-Qun Chen, Laurence D. Hurst,
724 and Dacheng Tian. 2015. "Parent-Progeny Sequencing Indicates Higher Mutation Rates in
725 Heterozygotes." *Nature* 523 (7561): 463–67.