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. Males of one lineage were karyotyped to confirm the number of chromosomes inferred from 3C data. 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 (Lenoir et al. 1990; Boulay et al. 2017). Their foraging activities are strictly diurnal, with most species being active during the hottest hours of the day (Wehner *et al.*, 1992; Cerda *et al.*, 1998). 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 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 (Gehring & Wehner 1995; Perez & Aron 2020; Pfeffer et al. 2019; Sommer & Wehner 2012; Willot et al. 2017; Perez et al. 2021; Aron & Wehner 2021).

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, Mardulyn, and Leniaud 2016; Peeters and Aron 2017; Aron et al. 2016). Unusual reproductive systems relying on conditional use of sex for the production of different female castes have evolved repeatedly in different *Cataglyphis* groups. Under these systems, non-reproductive workers are sexually produced, 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 (Pearcy et al. 2004; Kuhn et al. 2020). 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 (Leniaud et al. 2012; Eyer et al. 2013; Kuhn et al. 2020).

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 one species (*Cataglyphis niger*) is available for genomic analyses (Yahav & Privman, 2019). To fill this gap, we combined Oxford Nanopore long reads, Illumina short reads and chromosome conformation capture (3C) sequencing (Lieberman-Aiden et al. 2009; Marie-Nelly et al. 2014; Flot, Marie-Nelly, and Koszul 2015) 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.

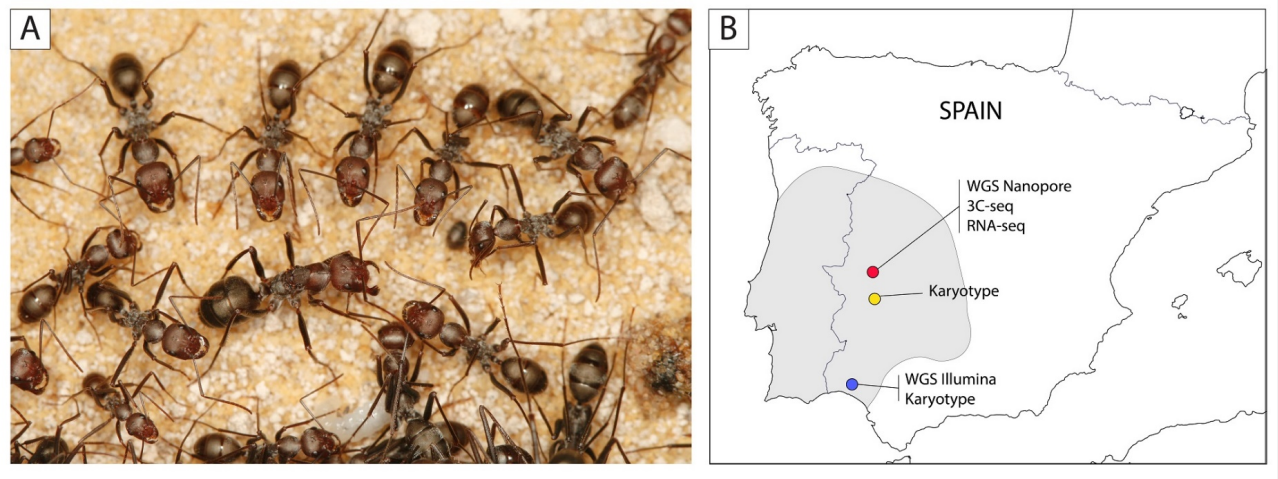

Figure 1. The ant *Cataglyphis hispanica.* (A) A queen of *C. hispanica* (red arrow) surrounded by workers. (B) Sampled sites in southwest Spain (red: Caceres, yellow: Merida and blue: Bonares). The complete range of the species *C. hispanica* is shown in grey. WGS: whole genome sequencing.

RESULTS AND DISCUSSION

Genome assembly

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.

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 203 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 do not contain much uncollapsed haplotypes, if any. In line with these results, KAT analyses based on the Illumina reads of each lineage showed a single peak of k-mer multiplicity, which were almost all represented exactly once in the assemblies as expected for high-quality genomes (Figure S1); k-mer completeness was estimated as 98.86% for Chis1 and 98.45%for Chis2 (Mapleson *et al.*, 2016). For each assembly, a region with no large-scale synteny pattern was assembled at the extremity of one scaffold (the first 5.4 Mb of scaffold #9 of Chis1 and the first 3.1 Mb of scaffold #7 of Chis2). Each of these regions consisted of a collection of small contigs (mostly in the 2-10 Kb range) with 2 to 5 times higher average coverage compared to other genomic regions. These problematic sequences exhibited microsyntenies with the extremities of other large

scaffolds (Figure 2; Figure S2) suggesting that they correspond to repeat sequences that were improperly assembled into fragmented contigs.

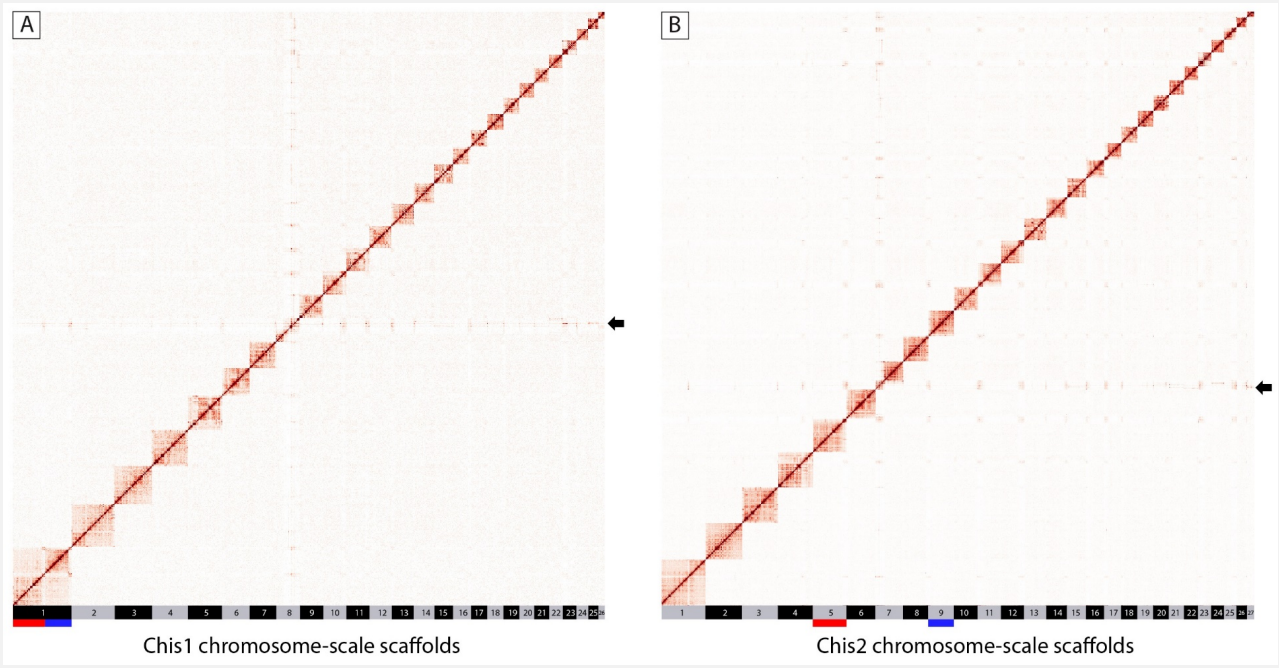

Figure 2. Assembly of the *Cataglyphis hispanica* Chis1 (A) and Chis2 (B) genomes into chromosomes. Hi-C interaction map revealing the presence of 26 and 27 linkage groups. The color scale represents the interaction frequencies. The positions of the rearranged chromosome are indicated, and the arrows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

Comparison of the Chis1 and Chis2 assemblies revealed that 25 of the chromosome-scale scaffolds had a one-to-one homolog in each of the two lineages. In addition, and by contrast, the largest scaffold of Chis1 (#1) was split into two chromosome-scale scaffolds (# 5 and #10) in the Chis2 assembly (Figure S2). The 3C contact maps of both lineages showed that these scaffolds (Chis2 #5, #10 and Chis1 #1) correspond to well-individualized 3D features, thereby ruling out a scaffolding error (Figure 2). These observations support that a chromosome centric (Robertsonian) translocation took place in one of the two lineages studied. Centric fusions and fissions are the main

mechanisms of karyotype evolution in many animal groups, including ants (Lorite & Palomeque, 2010). Robertsonian translocations have been reported to promote speciation through the suppression of genetic recombination in the vicinity of rearranged centromeric regions or the reduction of fertility in karyotypic hybrids (Davisson & Akeson, 1993; Faria & Navarro, 2010). The significance of this genomic rearrangement for the origin of social hybridogenesis in *Cataglyphis* deserves further investigation. A different number of chromosomes among lineages could potentially reduce the fertility of inter-lineage hybrids and may contribute to the long-term maintenance of divergent lineages (Schwander, Keller, and Cahan 2007). Notably, two species closely related to *C. hispanica* also reproduce by social hybridogenesis involving two interdependent lineages (Eyer et al., 2013). We have previously speculated that this recurrent genetic system may have evolved prior to the speciation of these three species. Under this scenario, social hybridogenesis would be controlled by an ancient biallelic non-recombining region maintained by chromosomal rearrangement(s), with two haplotypes accounting for the occurrence of two genetic lineages in each species (Darras et al., 2014).

Intrachromosomal rearrangements between the lineages, consisting in large translocations and inversions, were also observed for 6 of the 25 large orthologous scaffolds (Figure S2), but these could not be confirmed independently with the current data.

Karyotyping

To validate the number of chromosomes scaffolded, we inspected metaphase chromosome plates from haploid males. In ants, as in other Hymenoptera, females are diploid (2*n*) whereas males are haploid (*n*). Two males of the Chis2 lineage from two distant localities were analyzed (Figure S3A). Both male karyotypes carried 27 chromosomes. The precise morphology of the chromosomes could not be determined due to their small sizes (Figure 3). This number is within the range reported for *Cataglyphis bicolor*, *Cataglyphis iberica* and *Cataglyphis setipes* species and other Formicine species of the genera *Formica*, *Iberoformica* and *Polyergus* (n= 26-27) (Hauschteck-Jungen & Jungen, 1983; Imai *et al.*, 1984; Lorite & Palomeque, 2010). These results confirm the 3C

scaffolding of the Chis2 genome into 27 chromosomes. No male of the Chis1 lineage could be obtained for karyotyping.

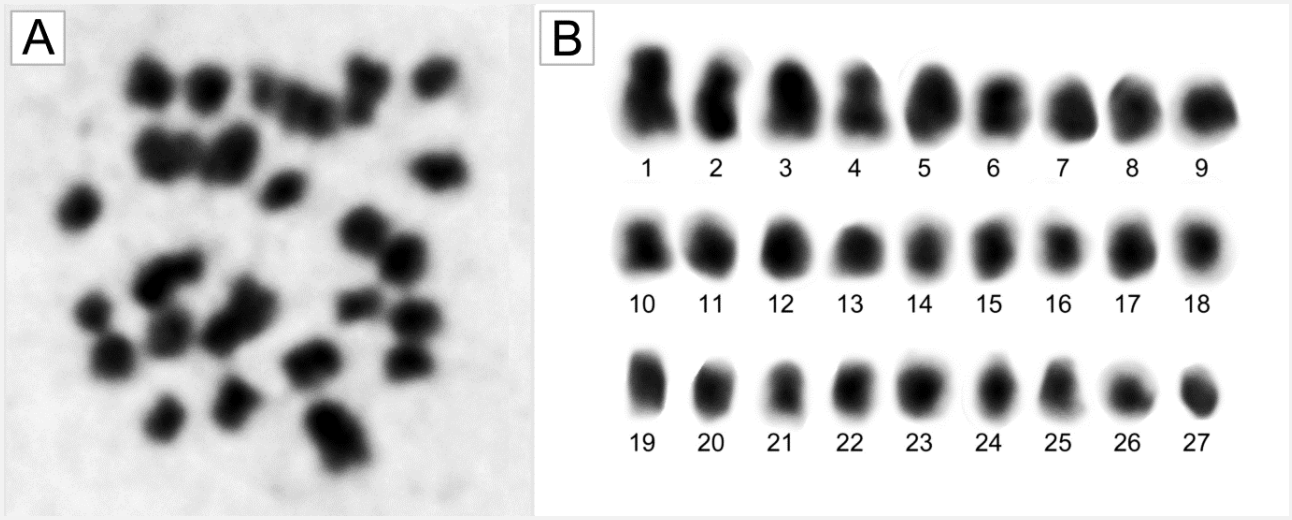

Figure 3. Chromosomes of *Cataglyphis hispanica* (Chis2). (A) Metaphase chromosome plate and (B) karyotype of a male from Merida (Spain), showing that the haploid chromosome number in this lineage is n=27.

Gene annotation

We annotated the genome of the Chis2 lineage. *Ab initio* gene prediction using AUGUSTUS and homology-based predictions using GenomeThreader (Gremme *et al.*, 2005) identified 16,993 and 8,234 gene models, respectively. A total of 40,969 models (including isoforms) were also predicted by the PASA/Transdecoder (Haas *et al.*, 2003) pipeline using direct evidence from 13 Gbp of Illumina RNA-seq data. The three annotation sets were validated and combined into a single annotation of 16,146 non-overlapping models using EvidenceModeler (Haas *et al.*, 2008). Among these, 11,101 gene models showed significant similarity to at least one RefSeq ant protein and 10,543 had a functional eggNOG hit (Huerta-Cepas et al. 2017, 2019). After filtering out models with no protein similarity and no known functional information, we obtained a final set of 11,290 high quality gene models. 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 gene set of *C. hispanica* (Chis2) with 19 published ant

annotations. Out of the 258,587 protein-coding genes analyzed using OrthoFinder (Emms & 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%).

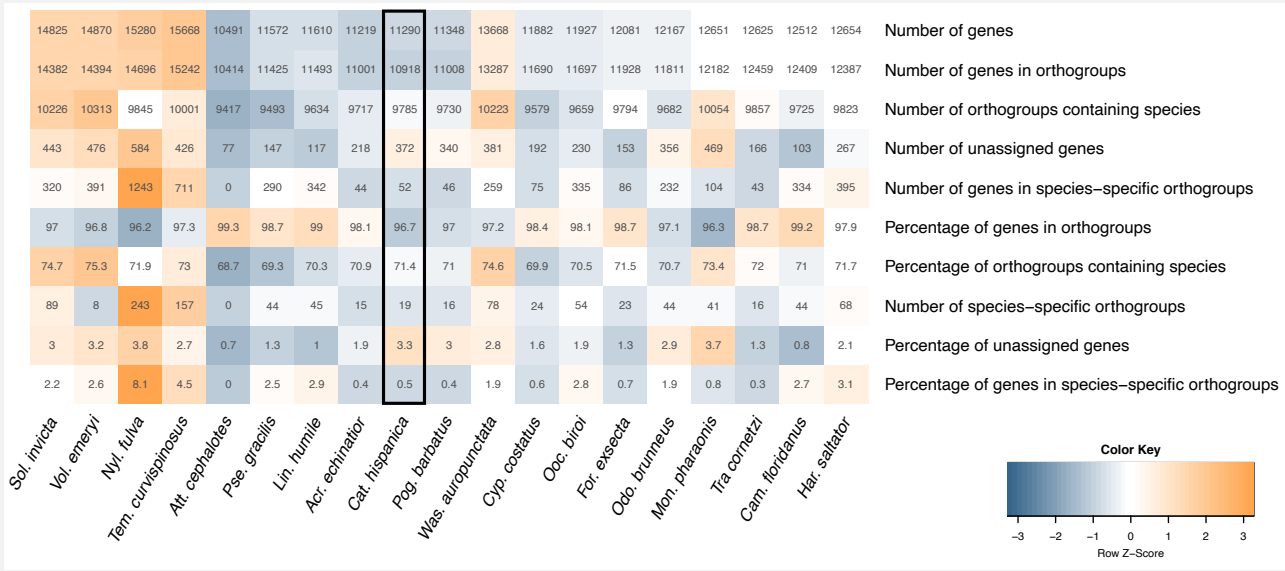

Figure 4. Summary values from the ortholog analyses. The color intensity indicates the z-score of variation (deviation from the mean) among all species, from the smallest value (blue) to the highest value (orange). The *Lasius niger* assembly was removed from this comparison due to its low quality.

Repeat annotation

We built custom repeat libraries for each of the two assemblies of *C. hispanica* and for 19 published ant genomes (Table S1). The Chis1 and Chis2 assemblies contained 1,708 and 1,673 different repetitive elements, which accounted for 15.43% (31,851,170 bp) and 15.1% (31,512,815 bp) of their assembly sizes, respectively (Figure 5). A large proportion of these corresponded to unclassified interspersed repeats (6.7% / 6.78% of the genomes; Figure S4). The two genomes also contained 2.0% / 1.8% of Class I (retroelements), and 2.18% / 1.85% of Class II elements (DNA transposons). In total, 56 different families of repetitive elements were annotated in *C.*

hispanica. LTR/Gypsy were the most frequent transposable elements of Class I in the genomes (0.53% / 0.82%), while large Polintons / Mavericks were the most abundant Class II transposable elements (0.98% / 0.67%).

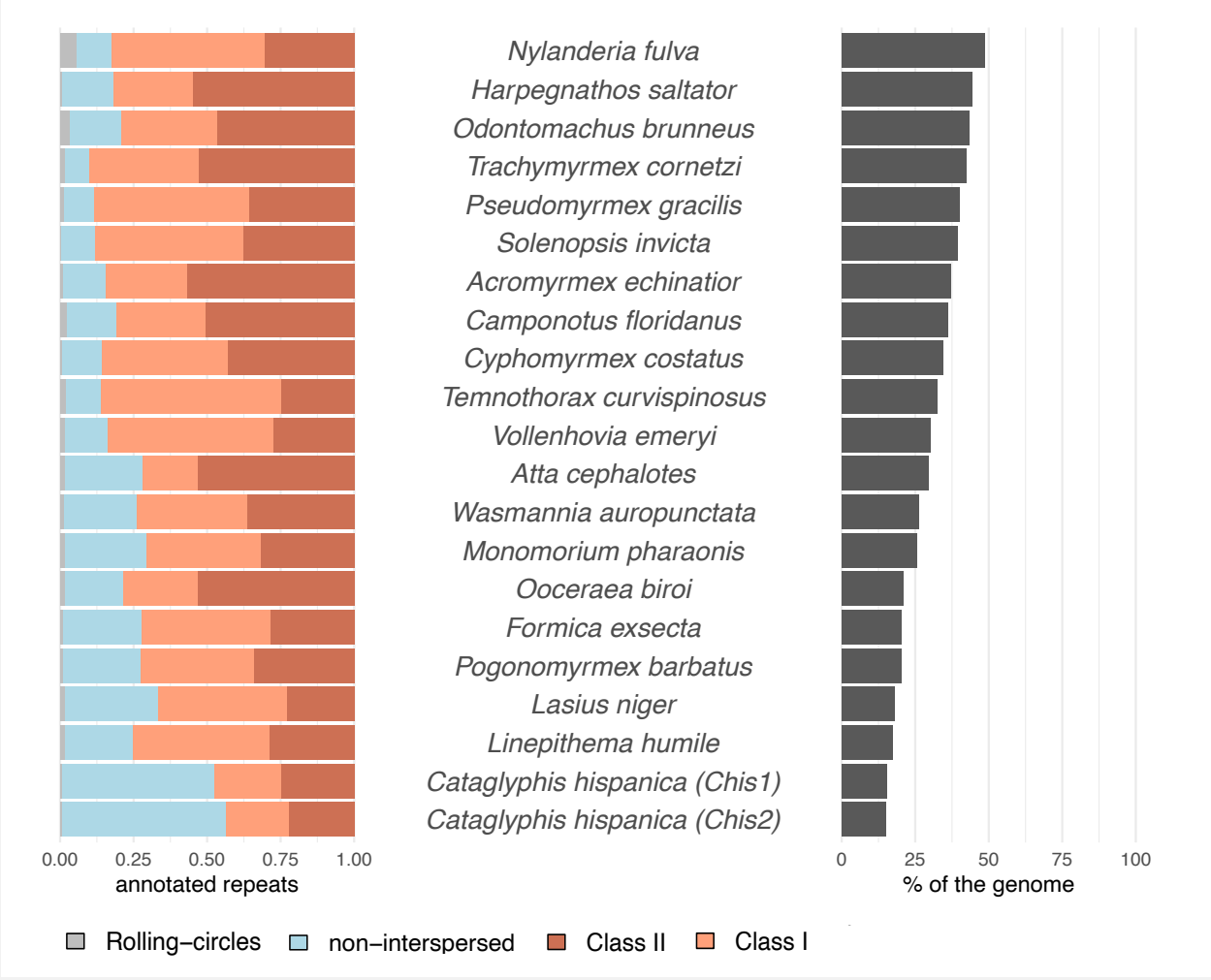

Figure 5. Summary of the repetitive elements' categories annotated in 20 different ant species using our custom pipeline. The ratios of the major categories of repetitive elements identified in each species is shown on the left. The total proportion of repetitive elements found in each genome is shown on the right.

Across published ant assemblies, the total proportion of transposable elements appeared quite variable (range: 17.27 – 48.47%; N= 19 ant species; Figure 5; Table S2). The *C. hispanica* assemblies had a smaller proportion of repetitive elements (15.1% - 15.43%) than any of these assemblies. This difference in repeat content was slightly smaler when compared to the assembly

of *Formica exsecta* (18.53 %), the closest species available for comparison. The relatively low proportion of transposable elements observed in the genomes of *C. hispanica* may be due to the fact that it was assembled primarily from noisy nanopore long-reads, possibly leading to a collapse of repeated regions. Alternatively, *C. hispanica* may resist the invasion and proliferation of transposable elements more efficiently than other species. Whether its unusual reproductive system, combining both diploid and haploid parthenogenesis for queen and male production, could help keep transposable elements at bay deserves further exploration.

Lineage divergence estimation

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

CONCLUSIONS

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

two lineages, resulting in 26 and 27 chromosomes in Chis1 and Chis 2 lineages, respectively. Whether this rearrangement played a role in the origin and maintenance of social hybridogenesis in *C. hispanica* and across other *Cataglyphis* species deserves further investigation (Darras *et al.*, 2014).

METHODS

Biological samples

Permits were obtained to collect colonies of *Cataglyphis hispanica* in three Spanish locations (Bonares, Caceres and Merida; Figure 1B). Male samples from Bonares were used for Illumina DNA sequencing. Shortly after sampling, the Bonares population was wiped out by human activities. Consequently, samples from another locality (Caceres) were used for subsequent Nanopore sequencing, 3C-seq and RNA-seq. Male pupae from two distant localities (Bonares and Merida) were used for karyotyping. Twelve diagnostic microsatellite loci were genotyped to assess the lineage membership of each queen and male prior to sequencing and karyotyping (Darras *et al.*, 2014).

DNA and RNA-Sequencing

Genomic resources were generated for both the Chis1 and the Chis2 lineages. High-molecular-weight DNA was extracted from pure lineage queen and male individuals using QIAGEN Genomic-tips. For each lineage, two queen clones originating from the same nest were used for Nanopore sequencing. Queens of *C. hispanica* are produced through automictic parthenogenesis with central fusion which results into diploid individuals that are highly homozygosus (Pearcy *et al.*, 2011; Darras *et al.*, 2014) and thus suitable for genome assembly. Nanopore libraries were prepared using rapid sequencing kits (SQK-RAD001 and SQK-RAD004). The resulting long read libraries were sequenced on MIN106 flow cells and basecalled using Albacore v2.1.10. For each lineage, three Illumina libraries were generated from whole-genome amplified DNA extracted from a single male

with mean insert sizes of 170 bp, 500 bp and 800 bp, and sequenced with a HiSeq2000 (paired-end 2 x 100 bp mode).

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

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

Genomes assembly

The genome of each hybridogenetic lineage was assembled independently. Nanopore data were assembled using Flye v2.7 with four iterations of polishing using long reads (Kolmogorov *et al.*, 2019). Raw Illumina reads were trimmed for quality and adapters were removed using Trimmomatic v0.32 with options ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 (Bolger *et al.*, 2014). The trimmed reads were then aligned to the long-read assemblies using BWA-MEM v0.7.15 (Li & Durbin, 2009). SNPs and indels with at least three supporting observations were called using freebayes v1.2 (Garrison & Marth, 2012), and error-corrected consensus sequences were obtained using BCFtools v1.4 (Li *et al.*, 2009).

To obtain chromosome-scale assemblies, we scaffolded the polished contigs with the 3C reads using instaGRAAL, a MCMC, proximity-ligation based scaffolder (Marie-Nelly *et al.*, 2014; Baudry

et al., 2020). The 3C reads were trimmed using cutadapt (Martin, 2011) and subsequently processed using hicstuff (https://zenodo.org/record/4722873) with the following parameters *–aligner bowtie2 –iterative –enzyme DpnII*. The instaGRAAL scaffolder was run on the pre-processed data for 100 cycles (parameters: level 4, with options *--coverage-std 1 –level 4 –cycles 100*) (Baudry *et al.*, 2020) and final scaffolds were obtained using the instaGRAAL *-polish* script, with all corrective procedures at once (only one parameter: *-m polish*). Briefly, instaGRAAL explores the chromosome structures by testing the relative positions and/or orientations of DNA segments (or bins) according to the contacts expected given a simple three-parameter power-law model. These modifications take the form of a fixed set of operations (swapping, flipping, inserting, merging, etc.) of bins corresponding to $3^4 = 81$ *Dpn*II restriction fragments. The likelihood of the model is then maximized by sampling the parameters using a MCMC approach (Marie-Nelly *et al.*, 2014). After 100 iterations (i.e., a likely position for each bin is tested 100 times), the genome structure converges towards a relatively stable structure that does not evolve anymore when more iterations are added, resulting in chromosome-level scaffolds. The algorithm is probabilistic and ignores initially part of the intrinsic structure of the original contigs in order to sample a larger range of genome space (Baudry *et al.*, 2020). Therefore, some trustworthy information contained in the initial polished assembly can be lost, or modified, along the way. The final correction step of instaGRAAL consists in reintegrating this lost information into the final assembly, to correct for instance local untrustworthy tiny inversions of individual bins within a contig. The contact maps of the scaffolded assemblies were built using hicstuff. Gaps created during the scaffolding process were closed using Nanopore data with four iterations of TGS-GapCloser (Xu *et al.*, 2019) and new polished consensus sequences were obtained using BCFtools (see method above). Completeness of the assemblies were assessed at each step using BUSCO v5.2.2 with the Hymenoptera odb10 lineage (Simão *et al.*, 2015; Waterhouse *et al.*, 2017). We also ran KAT v2.4.1 to compare the k-mer frequencies of Illumina reads to final assemblies (Mapleson *et al.*, 2016). To investigate differences in chromosomal arrangement among lineages, the two genome assemblies were aligned with minimap2 v2.17 (exact preset: -x asm5) and alignments were visualized using dot plots obtained with D-GENIES (Cabanettes & Klopp, 2018).

308

309 **Karyotyping**

310 To validate the number of chromosomes inferred from 3C contact information, chromosome
311 preparations were made from brains of male larvae following the protocol described by (Lorite *et*
312 *al.*, 1996), with some modifications. Briefly, larvae at the last instar stage were dissected and their
313 cerebral ganglia were transferred to microplate wells with 0.05% colchicine in distilled water. After
314 30 min, samples were transferred to a fixative solution (acetic acid:ethanol, 3:1) and incubated for
315 45 min. Ganglia cells were disaggregated in a drop of 50% acetic acid on a clean slide, new fixative
316 solution was added and the slides were dried at 60ºC. Chromosome preparations were stained with
317 10% Giemsa in phosphate buffer (pH 7). Microscopy images were captured with a CCD camera
318 (Olympus DP70) coupled to a microscope (Olympus BX51) and were processed using Adobe
319 Photoshop.

320

321 **Gene annotation**

322 We used the Chis2 chromosome-level assembly for gene annotation. A repeat library was
323 constructed using the REPET package v2.5 (Quesneville *et al.*, 2005; Flutre *et al.*, 2011). The
324 repeat library was cleaned up manually to remove bacterial genes, mitochondrial genes and genes
325 with hits to the gene set of the ant *Cardiocondyla obscurior* (v1.4) which had been purged of
326 transposable elements (Schrader *et al.*, 2014). The fraction of the genome classified by
327 RepeatClassifier as "Unknown" was reduced from 2.2% to 0.9% as a result of this procedure.
328 Repeats were soft-masked using RepeatMasker v4.0.7 (Smit and Hubley,
329 http://www.repeatmasker.org) prior to *de novo* gene prediction.

330 Gene models were inferred from RNA-seq, homology data and *ab initio* predictions. The three
331 RNA-seq libraries were aligned to the *Chis2* genome using STAR v2.6.0 (Dobin *et al.*, 2013) with
332 the multi-sample 2-pass mapping strategy. Transcripts were then assembled using Trinity v2.10.0
333 (Grabherr *et al.*, 2011; Haas *et al.*, 2013)(options --genome_guided_max_intron 100000 --
334 jaccard_clip) and combined into gene models using PASA (Haas *et al.*, 2003). Ant proteomes
335 annotated using the NCBI Eukaryotic Genome Annotation pipeline (RefSeq, taxid:36668) were

aligned to the genome using GenomeThreader v1.5.10 (Gremme *et al.*, 2005) in order to predict gene structures. AUGUSTUS *ab initio* predictions were generated using BRAKER v2.1.02 (Hoff *et al.*, 2016, 2019) based on hints from RNA-seq data and GenomeThreader protein alignments (--etpmode). BRAKER was first run with preliminary AUGUSTUS parameters trained by running BUSCO v3.0.2 on the genome assembly (--long option; Hymenoptera odb9 database). To refine the training of AUGUSTUS, the most accurate gene models inferred by BRAKER were then identified using GeneValidator (Drăgan *et al.*, 2016) with RefSeq ant proteomes as references and an arbitrary quality threshold of Q89. To avoid biases, predicted proteins with more than 70% sequence identity to another protein in the set were removed from the selected gene models using the aa2nonred.pl script provided with BRAKER. The resulting gene models were used to train AUGUSTUS again, and BRAKER was run with the new parameter set. *Ab initio*, RNA-seq-based and homology-based gene predictions were combined into a single gene set using EVidenceModeler v1.1.1 (Haas *et al.*, 2008) with the following weight settings: PASA alignments: 10; GenomeThreader alignments: 3, Augustus predictions: 1, PASA/Transdecoder predictions: 1, GenomeThreader predictions: 1. Functional information was obtained from eggNOG-mapper v2 (Huerta-Cepas et al. 2017, 2019) with the options "taxonomic scope adjusted per query" and "annotations transferred from any ortholog". Protein sequences with similarity to RefSeq ant proteins (as of July 2019) were identified using blastp and an E-value threshold of 10^{-5}. Annotations with no known functional information and no hits to any RefSeq ant proteins were filtered out.

Comparative analyses

To identify orthologous and taxonomically restricted genes (or orphan genes), we compared the proteomes of *C. hispanica*, of 18 ants annotated by the NCBI Eukaryotic Genome Annotation Pipeline (Table S1) and of *Lasius niger* (Konorov *et al.*, 2017) using OrthoFinder v2.3.12 (Emms & Kelly, 2019) with its standard DEndroBLAST workflow. We used the feature annotation tables from RefSeq annotations to select the longest isoform of each gene annotated by NCBI prior to analysis. The published genome of *L. niger* is highly incomplete (no more than 65% of the 4,415 highly conserved single-copy genes of BUSCO's Hymenoptera odb9 database are found in this

assembly). Consequently, it was only used to guide phylogenetic analyses due to its relative proximity with *Cataglyphis.* A preliminary catalog of single-copy orthologs was obtained from a first run of OrthoFinder. Single-copy sequences were aligned with Mafft v7.310 (Katoh & Standley, 2013) and the alignments were trimmed with trimAL v1.4.1(options "-gt 0.8 -st 0.001"; (Capella-Gutiérrez *et al.*, 2009). The concatenated alignments were then passed to IQ-TREE v1.7.17 (Nguyen et al. 2015) "-m LG+R4"; (Nguyen *et al.*, 2015) to infer a species tree. The tree was converted to an ultrametric topology with the r8s program with options "mrca root Obir Hsal; fixage taxon=root age=150; divtime method=LF algorithm=TN" (Sanderson, 2003). The resulting species tree was used for a second, more precise run of OrthoFinder.

Repeat annotation

To compare the frequency of repetitive elements found in the genome of *C. hispanica* to the frequencies found in the genomes of other ant species available (Table S4), we constructed optimized repeat libraries for each species using a custom pipeline (https://github.com/nat2bee/repetitive_elements_pipeline). Shortly, repeat libraries were built with RepeatModeler v1.0.11 (http://www.repeatmasker.org/RepeatModeler/), TransposonPSI (http://transposonpsi.sourceforge.net/) and LTRharvest from GenomeTools v1.6.1 (Ellinghaus *et al.*, 2008). For each species, the different libraries were merged into a non-redundant library (<80% identity) using USEARCH v11.0.667 (Edgar, 2010). Library annotations were obtained with RepeatClassifier. Each custom library was concatenated with the Dfam v3.1 Hymenoptera library of RepeatMasker v4.1.0 and used to annotate repeats in the genome of the corresponding species using RepeatMasker. Summary statistics of the annotated repeats was obtained with RepeatMasker_stats.py (https://github.com/nat2bee/repetitive_elements_pipeline).

Lineage divergence estimation

To estimate the divergences of the two lineages of *C. hispanica*, we investigated the polymorphism at 4-fold-degenerate sites, which we assumed to be neutrally evolving. The Illumina read of the Chis1 lineage were mapped onto the Chis2 reference genome and single-nucleotide variants were

called using MapCaller v0.9.9.41 (Lin and Hsu 2019). The resulting vcf file was filtered to keep only single-nucleotide variants with two alleles and a 'PASS' quality filter. To determine the proportion of 4-fold sites that were polymorphic among our male samples of the two lineages, the positions of 4-fold sites in coding sequences of our annotation were identified using a custom script (T. Sackton, github.com/tsackton/linked-selection.git).

DECLARATIONS

Data Availability

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

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-ig-nite.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 re-sources 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

419 performed karyotyping. LB optimized 3C scaffolding parameters. NG performed 3C scaffolding. MM
420 prepared the 3C libraries. FR constructed the TE library. IA supervised the construction of the TE
421 library. RK supervised 3C library generation and scaffolding. JFF supervised genome assembly.
422 SA collected the ants, designed and supervised the study. All authors read and approved the final
423 manuscript.
424
425 **Competing Interests**
426 The authors declare that they have no competing interests.
427
428 **Acknowledgments**
429 We thank A. Cohanim and E. Privman for their advices on early genome assemblies and Qiaowei
430 M. Pan for her comments on the manuscript.
431 ---
432
433 **REFERENCES**
434 Aron, Serge, Pascale Lybaert, Claire Baudoux, Morgane Vandervelden, and Denis Fournier. 2016.
435 "Sperm Production Characteristics Vary with Level of Sperm Competition in *Cataglyphis*
436 Desert Ants." *Functional Ecology* 30 (4): 614–24.
437 Aron, Serge, Patrick Mardulyn, and Laurianne Leniaud. 2016. "Evolution of Reproductive Traits in
438 *Cataglyphis* Desert Ants: Mating Frequency, Queen Number, and Thelytoky." *Behavioral*
439 *Ecology and Sociobiology* 70 (8): 1367–79.
440 Aron, Serge, and Rüdiger Wehner. 2021. "Cataglyphis." In *Encyclopedia of Social Insects*, 217–23.
441 Cham: Springer International Publishing.
442 Baudry, Lyam, Nadège Guiglielmoni, Hervé Marie-Nelly, Alexandre Cormier, Martial Marbouty,
443 Komlan Avia, Yann Loe Mie, et al. 2020. "InstaGRAAL: Chromosome-Level Quality
444 Scaffolding of Genomes Using a Proximity Ligation-Based Scaffolder." *Genome Biology* 21
445 (1): 148.

446 Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for
447 Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20.
448 Boulay, Raphaël, Serge Aron, Xim Cerdá, Claudie Doums, Paul Graham, Abraham Hefetz, and
449 Thibaud Monnin. 2017. “Social Life in Arid Environments: The Case Study of *Cataglyphis*
450 Ants.” *Annual Review of Entomology* 62 (January): 305–21.
451 Cabanettes, Floréal, and Christophe Klopp. 2018. “D-GENIES: Dot Plot Large Genomes in an
452 Interactive, Efficient and Simple Way.” *PeerJ* 6 (June): e4958.
453 Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. “TrimAl: A Tool for
454 Automated Alignment Trimming in Large-Scale Phylogenetic Analyses.” *Bioinformatics* 25
455 (15): 1972–73.
456 Cerda, X., J. Retana, and S. Cros. 1998. “Critical Thermal Limits in Mediterranean Ant Species:
457 Trade-off between Mortality Risk and Foraging Performance.” *Functional Ecology* 12 (1):
458 45–55.
459 Cohen, Pnina, and Eyal Privman. 2019. “Speciation and Hybridization in Invasive Fire Ants.” *BMC*
460 *Evolutionary Biology* 19 (1): 111.
461 Darras, Hugo, Alexandre Kuhn, and Serge Aron. 2019. “Evolution of Hybridogenetic Lineages in
462 *Cataglyphis* Ants.” *Molecular Ecology* 28 (12): 3073–88.
463 Darras, Hugo, Laurianne Leniaud, and Serge Aron. 2014. “Large-Scale Distribution of
464 Hybridogenetic Lineages in a Spanish Desert Ant.” *Proceedings. Biological Sciences / The*
465 *Royal Society* 281 (1774): 20132396.
466 Davisson, M. T., and E. C. Akeson. 1993. “Recombination Suppression by Heterozygous
467 Robertsonian Chromosomes in the Mouse.” *Genetics* 133 (3): 649–67.
468 Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha,
469 Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal
470 RNA-Seq Aligner.” *Bioinformatics* 29 (1): 15–21.

471 Drăgan, Monica-Andreea, Ismail Moghul, Anurag Priyam, Claudio Bustos, and Yannick Wurm.
472 2016. "GeneValidator: Identify Problems with Protein-Coding Gene Predictions."
473 *Bioinformatics* 32 (10): 1559–61.
474 Edgar, Robert C. 2010. "Search and Clustering Orders of Magnitude Faster than BLAST."
475 *Bioinformatics* 26 (19): 2460–61.
476 Ellinghaus, David, Stefan Kurtz, and Ute Willhoeft. 2008. "LTRharvest , an Efficient and Flexible
477 Software for de Novo Detection of LTR Retrotransposons." *BMC Bioinformatics* 9 (1): 1–14.
478 Emms, David M., and Steven Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for
479 Comparative Genomics." *Genome Biology* 20 (1): 238.
480 Eyer, P. A., L. Leniaud, H. Darras, and S. Aron. 2013. "Hybridogenesis through Thelytokous
481 Parthenogenesis in Two *Cataglyphis* Desert Ants." *Molecular Ecology* 22 (4): 947–55.
482 Faria, Rui, and Arcadi Navarro. 2010. "Chromosomal Speciation Revisited: Rearranging Theory with
483 Pieces of Evidence." *Trends in Ecology & Evolution* 25 (11): 660–69.
484 Flot, Jean-François, Hervé Marie-Nelly, and Romain Koszul. 2015. "Contact Genomics: Scaffolding
485 and Phasing (Meta)Genomes Using Chromosome 3D Physical Signatures." *FEBS Letters*
486 589 (20 Pt A): 2966–74.
487 Flutre, Timothée, Elodie Duprat, Catherine Feuillet, and Hadi Quesneville. 2011. "Considering
488 Transposable Element Diversification in de Novo Annotation Approaches." *PloS One* 6 (1):
489 e16526.
490 Garrison, Erik, and Gabor Marth. 2012. "Haplotype-Based Variant Detection from Short-Read
491 Sequencing." *ArXiv [q-Bio.GN]*. arXiv. http://arxiv.org/abs/1207.3907.
492 Gehring, W. J., and R. Wehner. 1995. "Heat Shock Protein Synthesis and Thermotolerance in
493 Cataglyphis, an Ant from the Sahara Desert." *Proceedings of the National Academy of*
494 *Sciences of the United States of America* 92 (7): 2994–98.
495 Grabherr, Manfred G., Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido
496 Amit, Xian Adiconis, et al. 2011. "Full-Length Transcriptome Assembly from RNA-Seq Data
497 without a Reference Genome." *Nature Biotechnology*. https://doi.org/10.1038/nbt.1883.

Gremme, Gordon, Volker Brendel, Michael E. Sparks, and Stefan Kurtz. 2005. “Engineering a Software Tool for Gene Structure Prediction in Higher Organisms.” *Information and Software Technology* 47 (15): 965–78.

Haas, Brian J., Arthur L. Delcher, Stephen M. Mount, Jennifer R. Wortman, Roger K. Smith Jr, Linda I. Hannick, Rama Maiti, et al. 2003. “Improving the Arabidopsis Genome Annotation Using Maximal Transcript Alignment Assemblies.” *Nucleic Acids Research* 31 (19): 5654–66.

Haas, Brian J., Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D. Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. “De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis.” *Nature Protocols* 8 (8): 1494–1512.

Haas, Brian J., Steven L. Salzberg, Wei Zhu, Mihaela Pertea, Jonathan E. Allen, Joshua Orvis, Owen White, C. Robin Buell, and Jennifer R. Wortman. 2008. “Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments.” *Genome Biology* 9 (1): R7.

Hauschteck-Jungen, E., and H. Jungen. 1983. “Ant Chromosomes.” *Insectes Sociaux* 30 (2): 149–64.

Hoff, Katharina J., Simone Lange, Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2016. “BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS: Table 1.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btv661.

Hoff, Katharina J., Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2019. “Whole-Genome Annotation with BRAKER.” *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-9173-0_5.

Huerta-Cepas, Jaime, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. 2017. “Fast Genome-Wide Functional Annotation through Orthology Assignment by EggNOG-Mapper.” *Molecular Biology and Evolution* 34 (8): 2115–22.

524 Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund,
525 Helen Cook, Daniel R. Mende, et al. 2019. “EggNOG 5.0: A Hierarchical, Functionally and
526 Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502
527 Viruses.” *Nucleic Acids Research* 47 (D1): D309–14.
528 Imai, B. Y. H. T., C. Baroni Urbani, M. Kubota, G. P. Sharma, M. N. Narasimhanna, B. C. Das, A.
529 K. Sharma, et al. 1984. “Karyological Survey of Indian Ants.” *The Japanese Journal of*
530 *Genetics*. https://doi.org/10.1266/jjg.59.1.
531 Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software
532 Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30
533 (4): 772–80.
534 Keightley, Peter D., Rob W. Ness, Daniel L. Halligan, and Penelope R. Haddrill. 2014. “Estimation
535 of the Spontaneous Mutation Rate per Nucleotide Site in a Drosophila Melanogaster Full-
536 Sib Family.” *Genetics* 196 (1): 313–20.
537 Keightley, Peter D., Ana Pinharanda, Rob W. Ness, Fraser Simpson, Kanchon K. Dasmahapatra,
538 James Mallet, John W. Davey, and Chris D. Jiggins. 2015. “Estimation of the Spontaneous
539 Mutation Rate in Heliconius Melpomene.” *Molecular Biology and Evolution* 32 (1): 239–43.
540 Kolmogorov, Mikhail, Jeffrey Yuan, Yu Lin, and Pavel A. Pevzner. 2019. “Assembly of Long, Error-
541 Prone Reads Using Repeat Graphs.” *Nature Biotechnology* 37 (5): 540–46.
542 Konorov, Evgenii A., Mikhail A. Nikitin, Kirill V. Mikhailov, Sergey N. Lysenkov, Mikhail Belenky,
543 Peter L. Chang, Sergey V. Nuzhdin, and Victoria A. Scobeyeva. 2017. “Genomic Exaptation
544 Enables Lasius Niger Adaptation to Urban Environments.” *BMC Evolutionary Biology* 17
545 (Suppl 1): 39.
546 Kuhn, Alexandre, Hugo Darras, Omid Paknia, and Serge Aron. 2020. “Repeated Evolution of Queen
547 Parthenogenesis and Social Hybridogenesis in *Cataglyphis* Desert Ants.” *Molecular Ecology*
548 29 (3): 549–64.
549 Leniaud, Laurianne, Hugo Darras, Raphael Boulay, and Serge Aron. 2012. “Social Hybridogenesis
550 in the Clonal Ant Cataglyphis Hispanica.” *Current Biology: CB* 22 (13): 1188–93.

Lenoir, A., E. Nowbahari, L. Querard, N. Pondicq, and C. Delalande. 1990. "Habitat Exploitation and Intercolonial Relationships in the Ant *Cataglyphis* Cursor (Hymenoptera: Formicidae)." *Acta Oecologica* 11 (1): 3–18.

Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-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.

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.

Marie-Nelly, Hervé, Martial Marbouty, Axel Cournac, Jean-François Flot, Gianni Liti, Dante Poggi Parodi, Sylvie Syan, et al. 2014. “High-Quality Genome (Re)Assembly Using Chromosomal Contact Data.” *Nature Communications* 5 (December): 5695.

Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” *EMBnet.Journal* 17 (1): 10–12.

Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. “IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.” *Molecular Biology and Evolution* 32 (1): 268–74.

Obbard, Darren J., John Maclennan, Kang-Wook Kim, Andrew Rambaut, Patrick M. O’Grady, and Francis M. Jiggins. 2012. “Estimating Divergence Dates and Substitution Rates in the Drosophila Phylogeny.” *Molecular Biology and Evolution* 29 (11): 3459–73.

Oppold, Ann-Marie, and Markus Pfenninger. 2017. “Direct Estimation of the Spontaneous Mutation Rate by Short-Term Mutation Accumulation Lines in Chironomus Riparius.” *Evolution Letters* 1 (2): 86–92.

Pearcy, O. J. Hardy, and S. Aron. 2011. “Automictic Parthenogenesis and Rate of Transition to Homozygosity.” *Heredity* 107 (2): 187–88.

Pearcy, Morgan, Serge Aron, Claudie Doums, and Laurent Keller. 2004. “Conditional Use of Sex and Parthenogenesis for Worker and Queen Production in Ants.” *Science* 306 (5702): 1780–83.

Peeters, Christian, and Serge Aron. 2017. “Evolutionary Reduction of Female Dispersal in *Cataglyphis* Desert Ants.” *Biological Journal of the Linnean Society. Linnean Society of London* 122 (1): 58–70.

602 Perez, Rémy, and Serge Aron. 2020. “Adaptations to Thermal Stress in Social Insects: Recent
603 Advances and Future Directions.” *Biological Reviews of the Cambridge Philosophical*
604 *Society* 95 (6): 1535–53.
605 Perez, Rémy, Natalia de Souza Araujo, Matthieu Defrance, and Serge Aron. 2021. “Molecular
606 Adaptations to Heat Stress in the Thermophilic Ant Genus *Cataglyphis*.” *Molecular Ecology*
607 30 (21): 5503–16.
608 Pfeffer, Sarah Elisabeth, Verena Luisa Wahl, Matthias Wittlinger, and Harald Wolf. 2019. “High-
609 Speed Locomotion in the Saharan Silver Ant,.” *The Journal of Experimental Biology* 222 (Pt
610 20). https://doi.org/10.1242/jeb.198705.
611 Quesneville, Hadi, Casey M. Bergman, Olivier Andrieu, Delphine Autard, Danielle Nouaud, Michael
612 Ashburner, and Dominique Anxolabehere. 2005. “Combined Evidence Annotation of
613 Transposable Elements in Genome Sequences.” *PLoS Computational Biology* 1 (2): 166–
614 75.
615 Sanderson, Michael J. 2003. “R8s: Inferring Absolute Rates of Molecular Evolution and Divergence
616 Times in the Absence of a Molecular Clock.” *Bioinformatics* 19 (2): 301–2.
617 Schrader, Lukas, Jay W. Kim, Daniel Ence, Aleksey Zimin, Antonia Klein, Katharina Wyschetzki,
618 Tobias Weichselgartner, et al. 2014. “Transposable Element Islands Facilitate Adaptation to
619 Novel Environments in an Invasive Species.” *Nature Communications*.
620 https://doi.org/10.1038/ncomms6495.
621 Schwander, Tanja, Laurent Keller, and Sara Helms Cahan. 2007. “Two Alternate Mechanisms
622 Contribute to the Persistence of Interdependent Lineages in Pogonomyrmex Harvester
623 Ants.” *Molecular Ecology* 16 (17): 3533–43.
624 Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny
625 M. Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness
626 with Single-Copy Orthologs.” *Bioinformatics* 31 (19): 3210–12.
627 Sommer, Stefan, and Rüdiger Wehner. 2012. “Leg Allometry in Ants: Extreme Long-Leggedness in
628 Thermophilic Species.” *Arthropod Structure & Development* 41 (1): 71–77.

Waterhouse, Robert M., Mathieu Seppey, Felipe A. Simão, Mosè Manni, Panagiotis Ioannidis, Guennadi Klioutchnikov, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2017. “BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics.” *Molecular Biology and Evolution*, December. https://doi.org/10.1093/molbev/msx319.

Wehner, R., A. C. Marsh, and S. Wehner. 1992. “Desert Ants on a Thermal Tightrope.” *Nature* 357 (6379): 586–87.

Willot, Quentin, Cyril Gueydan, and Serge Aron. 2017. “Proteome Stability, Heat Hardening and Heat-Shock Protein Expression Profiles in Desert Ants.” *The Journal of Experimental Biology* 220 (Pt 9): 1721–28.

Xu, Mengyang, Lidong Guo, Shengqiang Gu, Ou Wang, Rui Zhang, Guangyi Fan, Xun Xu, Li Deng, and Xin Liu. 2019. “TGS-GapCloser: Fast and Accurately Passing through the Bermuda in Large Genome Using Error-Prone Third-Generation Long Reads.” *BioRxiv*. bioRxiv. https://doi.org/10.1101/831248.

Yahav, Tal, and Eyal Privman. 2019. “A Comparative Analysis of Methods for de Novo Assembly of Hymenopteran Genomes Using Either Haploid or Diploid Samples.” *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-42795-6.

Yang, Sihai, Long Wang, Ju Huang, Xiaohui Zhang, Yang Yuan, Jian-Qun Chen, Laurence D. Hurst, and Dacheng Tian. 2015. “Parent-Progeny Sequencing Indicates Higher Mutation Rates in Heterozygotes.” *Nature* 523 (7561): 463–67.