

A putatively new ant species from the *Cataglyphis cursor* group displays a low level of polyandry with standard sexual reproduction

Claudie Doums^{1,2}, Pascaline Chifflet-Belle^{1,2}, Thomas Lenormand³, Raphael Boulay⁴, Irène Villalta⁴

¹Institut de Systématique, Évolution, Biodiversité (ISYEB), UMR 7205, MNHN, CNRS, EPHE - PSL, Sorbonne Université, Université des Antilles, 75005 Paris, France

²EPHE - PSL University, 75005 Paris, France

³CEFE, Univ Montpellier, CNRS, Univ Paul Valéry Montpellier 3, EPHE-PSL, IRD, Montpellier, France.

⁴IRBI, UMR 7261, Université de Tours, Tours, France

Corresponding author: Claudie Doums ISYEB, UMR 7205 (CNRS MNHN SU EPHE-PSL), Muséum national d'Histoire naturelle, CP39, Bât. Cryptogamie, 12 rue Buffon, 75005 Paris, France

Email : claudie.doums@ephe.psl.eu

Contents

Title	Page
Table S1 Sampling information	2
Table S2 Characteristics of the microsatellite markers	3
Table S3 Linkage disequilibrium	4
Table S4 Fis values and exact Fisher tests for Hardy Weinberg equilibrium	5
Table S5 Allelic richness	6
Table S6 Gene diversity	7
Figure S1 Distributions of expected relatedness Maximum likelihood model	8
References cited	9
	10

Table S1 Sampling information. The GPS coordinates of each collected colony is given with the number of workers (N workers) and the number of gynes (N gynes) genotyped for each colony in the two populations, Salobralajejo and Castrillo. The reference (see Figures 1 and 2 of main paper) of the mtDNA haplotype found is given for each colony. We also given the code mtDNA haplotype (see Figure 2 of main paper) found for the worker analysed per colony.

Salobralajejo						Castrillo					
Colony	Latitude	Longitude	N_workers	N_gynes	mtDNA	Colony	Latitude	Longitude	N_workers	N_gynes	mtDNA
1	40.571116	-4.899262	1	0	CU01W1	31	41.226865	-5.317609	7	0	CU31W1
2	40.571056	-4.899435	1	0	CU01W1	32	41.226908	-5.317585	1	0	CU31W1
3	40.571086	-4.899504	1	0	CU01W1	33	41.22685	-5.317498	1	0	CU31W1
4	40.571098	-4.899475	12	0	CU01W1	35	41.226854	-5.317452	1	0	CU31W1
5	40.571093	-4.899483	1	0	CU01W1	36	41.226865	-5.31742	1	0	CU31W1
6	40.571067	-4.899782	1	0	CU01W1	37	41.226879	-5.317409	1	0	CU37W1
7	40.571047	-4.899957	11	0	CU01W1	38	41.22689	-5.317395	1	0	CU31W1
8	40.571041	-4.899977	9	3	CU01W1	40	41.226892	-5.317305	1	0	CU31W1
9	40.571041	-4.899979	5	4	CU01W1	41	41.22686	-5.317316	1	0	CU31W1
10	40.571041	-4.899979	1	0	CU01W1	42	41.227017	-5.317313	1	0	CU31W1
11	40.571015	-4.900031	12	0	CU01W1	43	41.227057	-5.317335	1	0	CU37W1
12	40.571011	-4.900047	1	0	CU01W1	45	41.227161	-5.317142	1	0	CU31W1
13	40.571001	-4.900072	12	0	CU01W1	46	41.227216	-5.317126	7	0	CU37W1
14	40.57097	-4.900147	1	0	CU01W1	47	41.227277	-5.317102	1	0	CU31W1
15	40.570945	-4.900184	11	0	CU01W1	50	41.227431	-5.316801	1	0	CU31W1
19	40.570961	-4.900223	4	2	CU01W1	51	41.227493	-5.316679	1	0	CU31W1
20	40.570973	-4.900239	1	0	CU01W1	52	41.227504	-5.316656	1	0	CU31W1
21	40.570952	-4.900294	1	0	CU01W1	53	41.22758	-5.31654	1	0	CU31W1
22	40.570888	-4.900404	1	0	CU01W1	54	41.227768	-5.316442	8	0	CU31W1
23	40.570891	-4.90041	1	0	CU01W1	56	41.227956	-5.316303	1	0	CU31W1
24	40.570891	-4.900475	1	0	CU01W1	57	41.227994	-5.316293	1	0	CU31W1
25	40.570871	-4.900508	1	0	CU01W1	58	41.228143	-5.316188	1	0	CU37W1
26	40.57097	-4.900615	1	0	CU01W1	59	41.228213	-5.316131	6	0	CU31W1
27	40.571088	-4.899455	10	0	CU01W1						
28	40.570837	-4.900681	1	0	CU01W1						
29	40.570801	-4.900682	1	0	CU01W1						
30	40.570883	-4.900674	1	0	CU01W1						

Table S2 Characteristics of the microsatellite markers. For each microsatellite, the mean allelic richness, the mean gene diversity, the mean *Fis* are given. The microsatellite markers designed for *C. piliscapa* (Ccur) were amplified following the protocol described in Chéron et al. (2011). For amplification of the Ant markers (Doums et al. 2013), two mixtures were made, one containing Ant10878, Ant2341, Ant3653 and the other containing Ant575, Ant2794, Ant3993 (Ant 2341, Ant 3653, Ant 575 were not retained as they gave inconsistent amplification results). Each amplification was then performed in 10- μ L volume containing 1 μ L of DNA solution (20–60 ng), 0.15 μ L of dNTP (40 mM), 1 μ L of buffer 10x, 0.75 unit of Taq polymerase and 0.10 μ M of each primer. We amplified individually Ant9218 following the protocol: 10- μ L volume containing 1 μ L of DNA solution (20–60 ng), 0.07 μ L of dNTP (40 mM), 1 μ L Buffer 10x, 0.35 unit of Taq polymerase and 0.10 μ M of each primer. Each PCR was performed with positive and negative controls. The PCR products of the microsatellite loci were sent to INRA-GDEC (platform GENTYANE, Clermont Ferrand, France) for genotyping. Allele scoring was performed using GENEMAPPER v.4 software. The Ccur26 locus was excluded from the analysis as it failed to amplify 9 out of the 50 individuals in our study and was associated with a highly significant heterozygote deficiency at both sites ($P < 0.0001$) suggesting the occurrence of null alleles at high frequency. For each locus, we have also provided the loss of heterozygosity (LOH) at each generation of thelytoky known for each locus based on data from Doums et al. (2013) and on unpublished results. The LOH is estimated for each locus based on heterozygous mothers and their thelytokously produced daughters. The LOH is the ratio of the number of homozygous daughters to the total number of daughters (equivalent to the transition rate to homozygosity of (Pearcy et al. 2004). The number of daughters used to estimate LOH is given in brackets.

Loci	Mean allelic richness	Mean gene diversity	Mean <i>Fis</i>	References	LOH
Ccur11	3.74	0.66	-0.02	(Pearcy et al. 2004)	0.005 (805)
Ccur26*	3.49	0.58	0.13	(Pearcy et al. 2004)	0.00 (532)
Ccur46	3.14	0.55	0.02	(Pearcy et al. 2004)	0.17 (531)
Ccur58	3.48	0.62	-0.01	(Pearcy et al. 2004)	0.29 (378)
Ccur76	---	---	---	(Pearcy et al. 2004)	0.5 (14)
Ccur89	4.02	0.69	0.10	(Pearcy et al. 2004)	0.31 (274)
Ccur99	2.83	0.47	0.02	(Pearcy et al. 2004)	0.34 (454)
Ccur51	3.42	0.56	0.03	(Chéron et al. 2011)	0.02 (626)
Ccur61	5.30	0.81	-0.04	(Chéron et al. 2011)	0.14 (547)
Ccur100	4.43	0.72	-0.02	(Chéron et al. 2011)	0.02 (612)
Ant9218	2.18	0.37	-0.02	(Butler et al. 2014)	0.21 (185)
Anr10878	5.02	0.82	0.04	(Butler et al. 2014)	0.27 (120)
Ant2794	3.56	0.61	0.00	(Butler et al. 2014)	0.27 (22)
Ant3653	---	---	---	(Butler et al. 2014)	0.18 (66)
Ant3993	2.83	0.49	0.03	(Butler et al. 2014)	0.28 (127)

Table S3 Linkage disequilibrium. The P values for exact linkage disequilibrium tests for each microsatellite locus pair across all populations were obtained using the Genepop package in R (R Core Team 2019). P values for the two populations were obtained using the combined Fisher test.

Locus pair	Chi2	df	P-Value	Locus pair	Chi2	df	P-Value	Locus pair	Chi2	df	P-Value
L11&L89	1.02	4.00	0.91	L11&L9218	3.08	4.00	0.55	L11&L3653	2.54	4.00	0.64
L11&L76	5.65	4.00	0.23	L89&L9218	1.37	4.00	0.85	L89&L3653	6.03	4.00	0.20
L89&L76	3.47	4.00	0.48	L76&L9218	3.81	4.00	0.43	L76&L3653	2.88	4.00	0.58
L11&L51	1.79	4.00	0.77	L51&L9218	0.94	4.00	0.92	L51&L3653	9.52	4.00	0.05
L89&L51	3.01	4.00	0.56	L58&L9218	0.00	2.00	1.00	L58&L3653	0.79	2.00	0.67
L76&L51	9.38	4.00	0.05	L61&L9218	0.20	4.00	1.00	L61&L3653	1.14	4.00	0.89
L11&L58	0.12	2.00	0.94	L99&L9218	4.16	4.00	0.38	L99&L3653	2.36	4.00	0.67
L89&L58	1.34	2.00	0.51	L100&L9218	5.17	4.00	0.27	L100&L3653	6.34	4.00	0.17
L76&L58	3.18	2.00	0.20	L11&L10878	0.00	4.00	1.00	L9218&L3653	4.31	4.00	0.37
L51&L58	3.11	2.00	0.21	L89&L10878	0.75	4.00	0.95	L10878&L3653	1.22	4.00	0.87
L11&L61	5.82	4.00	0.21	L76&L10878	2.05	4.00	0.73	L2794&L3653	5.33	4.00	0.25
L89&L61	8.95	4.00	0.06	L51&L10878	2.33	4.00	0.68	L11&L3993	1.03	4.00	0.90
L76&L61	5.47	4.00	0.24	L58&L10878	0.14	2.00	0.93	L89&L3993	0.89	4.00	0.93
L51&L61	9.54	4.00	0.05	L61&L10878	0.00	4.00	1.00	L76&L3993	8.96	4.00	0.06
L58&L61	0.46	2.00	0.79	L99&L10878	0.43	4.00	0.98	L51&L3993	5.73	4.00	0.22
L11&L99	3.57	4.00	0.47	L100&L10878	0.00	4.00	1.00	L58&L3993	2.29	2.00	0.32
L89&L99	7.15	4.00	0.13	L9218&L10878	0.15	4.00	1.00	L61&L3993	4.97	4.00	0.29
L76&L99	7.69	4.00	0.10	L11&L2794	2.78	4.00	0.59	L99&L3993	6.42	4.00	0.17
L51&L99	3.77	4.00	0.44	L89&L2794	3.41	4.00	0.49	L100&L3993	1.53	4.00	0.82
L58&L99	0.00	2.00	1.00	L76&L2794	5.37	4.00	0.25	L9218&L3993	3.14	4.00	0.53
L61&L99	8.42	4.00	0.08	L51&L2794	2.30	4.00	0.68	L10878&L3993	1.38	4.00	0.85
L11&L100	6.25	4.00	0.18	L58&L2794	0.03	2.00	0.99	L2794&L3993	1.99	4.00	0.74
L89&L100	1.56	4.00	0.82	L61&L2794	0.00	4.00	1.00	L3653&L3993	0.65	4.00	0.96
L51&L100	8.59	4.00	0.07	L100&L2794	8.56	4.00	0.07				
L58&L100	0.79	2.00	0.67	L9218&L2794	0.69	4.00	0.95				
L61&L100	0.54	4.00	0.97	L10878&L2794	0.00	4.00	1.00				
L99&L100	8.14	4.00	0.09								

Table S4 Fis values and exact Fisher tests for Hardy Weinberg equilibrium. The Fis value was estimated for each microsatellite locus and locality using the Genepop package (R Core Team 2019). When the exact test for HW equilibrium was significant, values are in red for significance at 0.05. None remained significant after Bonferroni's correction. NA indicates that *Fis* could not be estimated because the locus was not polymorphic in the locality.

Loci	Avila		Castrillo	
	<i>Fis</i>	Exact Fisher test	<i>Fis</i>	Exact Fisher test
L11a	-0.017	0.870	-0.103	0.027
L89a	-0.089	0.078	0.189	0.416
L76a	-0.111	0.631	-0.058	0.645
L51a	-0.037	0.168	-0.073	1
L58a	-0.010	1	NA	NA
L61a	0.0291	0.802	-0.030	0.877
L99a	-0.146	0.581	0.055	1
L100a	-0.142	0.787	0.033	0.384
L9218a	-0.010	1	0.178	0.123
L10878a	-0.029	0.906	0.219	0.037
L2794a	0.137	0.019	0.055	0.531
L3653a	0.213	0.057	-0.058	1
L3993a	-0.232	0.127	0.286	0.011

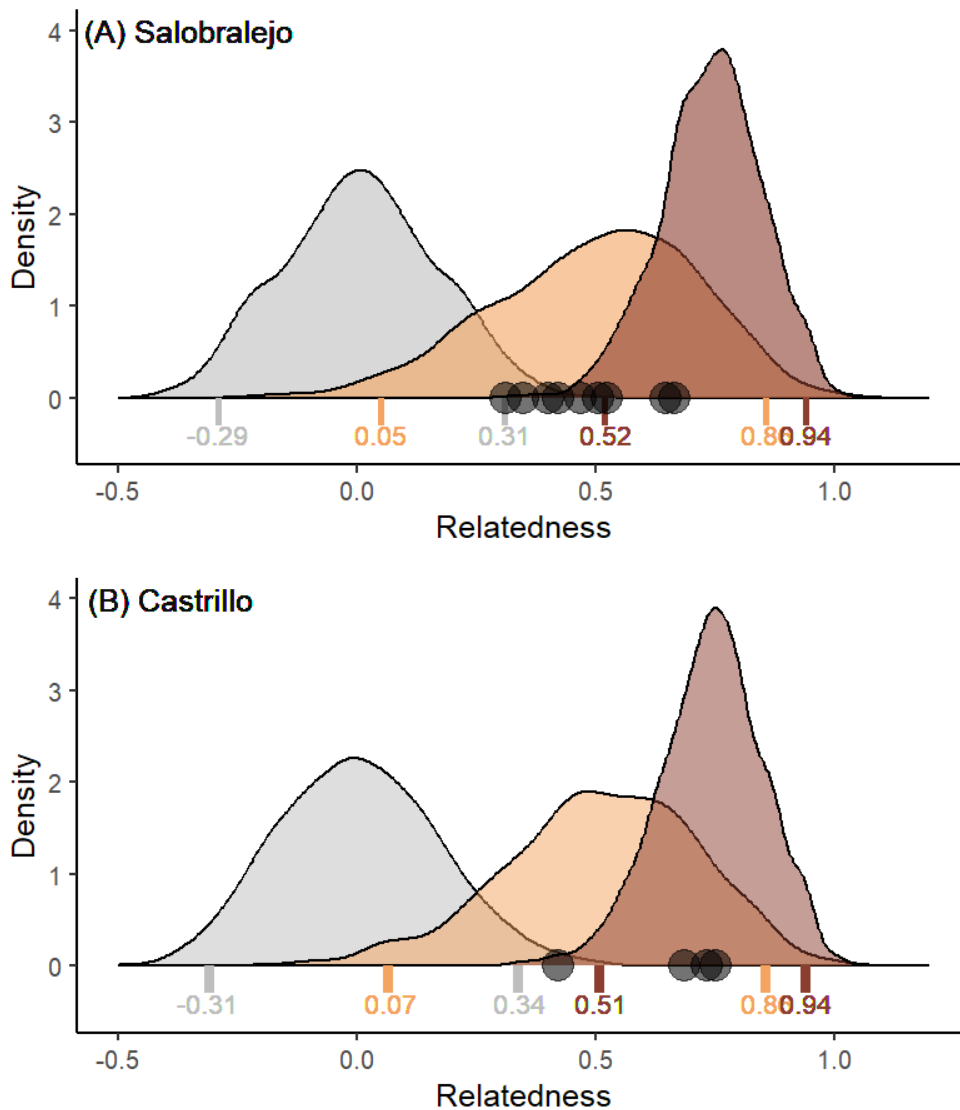
Table S5 Allelic richness. The allelic richness was estimated using the rarefaction method of El Mousadik and Petit (1996) provided by the package PopGenReport for each locality and microsatellite locus, the mean allelic richness between the two localities is also provided. For comparison, we also report the mean allelic richness obtained in *C. cursor* and *C. piliscapa* obtained by Eyer et al. (2022).

	Salobralejo	Castrillo	Mean <i>C. cursor</i> Spain	Mean <i>C. cursor</i>	Mean <i>C. piliscapa</i>
L46	1.00	1.00	1.00	2.60	3.87
L11	4.22	5.12	4.67	3.44	4.14
L89	3.88	1.99	2.94	3.27	5.03
L76	5.38	7.12	6.25	NA	NA
L51	2.90	2.00	2.45	2.35	4.88
L58	1.55	1.00	1.28	2.74	4.50
L61	6.39	6.67	6.53	5.30	5.31
L99	2.98	1.98	2.48	2.61	3.14
L100	4.24	8.84	6.54	4.36	4.53
L9218	1.57	2.92	2.25	1.73	2.79
L10878	8.39	8.97	8.68	4.61	5.57
L2794	10.44	9.16	9.80	3.33	3.88
L3653	6.16	2.40	4.28	NA	NA
L3993	5.72	4.17	4.94	2.13	3.79
Mean	4.63	4.53	4.58	3.23	4.22

Table S6 Gene diversity. The gene diversity (H_e) was estimated for each population and microsatellite locus using the package PopGenReport. The mean gene diversity between the two localities is also reported. For comparison, we also report the mean gene diversity obtained in *C. cursor* and *C. piliscapa* obtained by Eyer et al. (2022).

Locus	Avila	Castrillo	Mean <i>C. cursor</i> Spain	Mean <i>C. cursor</i>	Mean <i>C. piliscapa</i>
L46	0.00	0.00	0.00	0.48	0.66
L11	0.72	0.77	0.74	0.64	0.69
L89	0.70	0.49	0.60	0.61	0.81
L76	0.75	0.84	0.80	NA	NA
L51	0.51	0.48	0.49	0.40	0.78
L58	0.07	0.00	0.04	0.54	0.74
L61	0.70	0.83	0.76	0.82	0.80
L99	0.60	0.31	0.46	0.45	0.50
L100	0.67	0.88	0.77	0.70	0.75
L9218	0.07	0.52	0.30	0.31	0.45
L10878	0.88	0.87	0.88	0.79	0.86
L2794	0.92	0.90	0.91	0.61	0.60
L3653	0.83	0.20	0.51	NA	NA
L3993	0.74	0.71	0.73	0.38	0.64
mean	0.58	0.56	0.57	0.61	0.68

Figure S1 Distributions of expected relatedness. The expected relatedness is based on a simulation of 1000 individuals are shown under the of full-sister ($R = 0.75$ brown), half-sister ($R = 0.5$, light brown) or the unrelated hypothesis ($R = 0$, grey) for both localities (A) Salobrallejo and (B) Castrillo. We also provided the 95% confidence interval for each distribution. The simulation was performed for each locality using COANCESTRY (see M&M for more details). The mean relatedness observed for each colony (values from Table 1) is represented by the black dots. All estimates fall within the hypothesis of $R = 0.5$ for both localities. Three out of four colonies in Castrillo and three out of eight in Salobrallejo could also belong to the $R = 0.75$ distribution.



Maximum Likelihood model

We have three types of data. The first is the number of homozygote A_j and heterozygote a_j individuals observed for each locus j in the population. This dataset allows to estimate heterozygosity observed α_j at this locus. The second is the number of loss-of-heterozygosity (LOH) events L_j and maintenance-of-heterozygosity events l_j observed for each locus j by directly comparing parent and offspring. This dataset allows to estimate the LOH rate λ_j at this locus. The last dataset is the genotype $X_{i,j}$ of each queen i , where for each locus j , we observe whether the queen is heterozygous ($X_{i,j} = 0$) or homozygous ($X_{i,j} = 1$) at this locus. We suppose that a given queen is the product of a given number of generations of thelytoky t_i since the last episode of sexual reproduction. The aim is to estimate this number of generations for each queen using all the information. After one round of sexual reproduction, we assume that, on expectation, the level of heterozygosity returns to the population average α_j , but that each generation of thelytoky reduces heterozygosity at each locus (at a rate λ_j). Hence the chance of being heterozygous at a given locus for a queen after t_i generations of thelytoky is $(1 - \alpha_j)(1 - \lambda_j)^{t_i}$

The joined likelihood for the three dataset is therefore:

$$\sum_{i=1}^{n_Q} \sum_{j=1}^{n_l} ((1 - X_{i,j}) \text{Log}[(1 - \alpha_j)(1 - \lambda_j)^{t_i}] + (X_{i,j}) \text{Log}[1 - (1 - \alpha_j)(1 - \lambda_j)^{t_i}]) \\ + \sum_{j=1}^{n_l} (L_j \text{Log}[\lambda_j] + l_j \text{Log}[1 - \lambda_j]) + \sum_{j=1}^{n_l} (a_j \text{Log}[\alpha_j] + A_j \text{Log}[1 - \alpha_j])$$

where n_l is the number of microsatellite loci and n_Q is the number of genotyped queens. We estimated the parameters of the model (α_j, λ_j, t_i) by maximizing this loglikelihood. We used the FindMaximum function in Mathematica 13 to do so.

References cited

- Butler IA, Siletti K, Oxley PR, Kronauer DJC (2014) Conserved microsatellites in ants enable population genetic and colony pedigree studies across a wide range of species. *PLoS ONE* 9:e107334. <https://doi.org/10.1371/journal.pone.0107334>
- Chéron B, Monnin T, Fédérici P, Doums C (2011) Variation in patriline reproductive success during queen production in orphaned colonies of the thelytokous ant *Cataglyphis cursor*. *Molecular Ecology* 20:2011–2022. <https://doi.org/10.1111/j.1365-294X.2011.05075.x>
- Doums C, Cronin AL, Ruel C, et al (2013) Facultative use of thelytokous parthenogenesis for queen production in the polyandrous ant *Cataglyphis cursor*. *J Evol Biol* 26:1431–1444. <https://doi.org/10.1111/jeb.12142>
- El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoret Appl Genetics* 92:832–839. <https://doi.org/10.1007/BF00221895>
- Eyer PA, Finand B, Mona S, et al (2022) Integrative characterization of genetic and phenotypic differentiation in an ant species complex with strong hierarchical population structure and low dispersal abilities. *Heredity*. <https://doi.org/10.1038/s41437-022-00590-6>
- Pearcy M, Clemencet J, Chameron S, et al (2004) Characterization of nuclear DNA microsatellite markers in the ant *Cataglyphis cursor*. *Mol Ecol Notes* 4:642–644. <https://doi.org/10.1111/j.1471-8286.2004.00759.x>
- R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing,