# Evolutionary history of two cryptic species of Northern African jerboas

**Ana Filipa Moutinho** ( moutinho@evolbio.mpg.de )
Max-Planck-Institut für Evolutionsbiologie https://orcid.org/0000-0002-2838-9113
**Nina Serén**
Universidade do Porto Centro de Investigacao em Biodiversidade e Recursos Geneticos
**Joana Paupério**
Universidade do Porto Centro de Investigacao em Biodiversidade e Recursos Geneticos
**Teresa Luísa Silva**
Universidade do Porto Centro de Investigacao em Biodiversidade e Recursos Geneticos
**Fernando Martínez-Freiría**
Universidade do Porto Centro de Investigacao em Biodiversidade e Recursos Geneticos
**Graciela Sotelo**
Universidade do Porto Centro de Investigacao em Biodiversidade e Recursos Geneticos
**Rui Faria**
Universidade do Porto Centro de Investigacao em Biodiversidade e Recursos Geneticos
**Tapio Mappes**
Jyvaskylan Yliopisto
**Paulo Célio Alves**
Universidade do Porto Faculdade de Ciencias
**José Carlos Brito**
Universidade do Porto Centro de Investigacao em Biodiversidade e Recursos Geneticos
**Zbyszek Boratyński**
Universidade do Porto Centro de Investigacao em Biodiversidade e Recursos Geneticos

**Research article**

**Keywords:** African jerboas, cryptic diversity, demographic history, deserts, Jaculus, local adaptation, phylogenetics, reproductive isolation, Sahara-Sahel, speciation.

**Posted Date:** August 27th, 2019

**DOI:** https://doi.org/10.21203/rs.2.13580/v1

**Version of Record:** A version of this preprint was published on February 13th, 2020. See the published version at https://doi.org/10.1186/s12862-020-1592-z.

# Abstract

Background Complex paleo-climatic and geological events shape species distributions, thus affecting their evolutionary history. In Sahara-Sahel, climatic oscillations shifted the desert extent during the Pliocene-Pleistocene interval, triggering the diversification of several species. Here, we investigated how these biogeographical and ecological events have shaped patterns of genetic diversity and divergence in African Jerboas, desert specialist's species. We focused on two closely-related and cryptic species, Jaculus jaculus and J. hirtipes, where we (1) evaluated their genetic differentiation, (2) reconstructed their evolutionary and demographic history; (3) tested the level of gene flow between them and (4) assessed their ecological niche divergence. Results The analyses based on 321 individuals sampled throughout North Africa, 8 sequence fragments (including one mitochondrial and seven single copy nuclear DNA), 6 microsatellite markers and ecological modelling revealed: (1) a deep genetic divergence between J. jaculus and J. hirtipes, despite their overlapping distribution, (2) low levels of gene flow and strong species delimitation, in agreement with their classification as different taxa, (3) high genetic diversity but no apparent geographic population structure within species, suggesting large-distance migration between remote locations, concordant with putative demographic expansions, and (4) low level of large-scale ecological divergence between the two taxa, suggesting species micro-habitat specialization. Conclusions Overall, our results suggest a speciation event during the Pliocene-Pleistocene transition followed by the colonization of most of the Northern Africa by both evolutionary units. The contemporary distribution of genetic variation might reflect an ongoing expansion, demonstrating the ability of these species for fast and long-range colonization. Despite the largely overlapping distribution at a macrogeographic scale, our genetic results suggest that the two species remain reproductively isolated with only negligible levels gene flow being observed. The overlapping ecological preferences at a macro-geographic scale and the ecological divergence at a micro-habitat scale suggest that local adaptation might have played a crucial role in the speciation process of these species.


# Background

Defining species and understanding the processes behind speciation are subjects under extensive debate in studies of evolutionary ecology (1,2). It is suggested that divergent natural selection in contrasting habitats might trigger reproductive isolation, and thus speciation (3–5). However, divergence between populations may be opposed by gene flow, particularly in the absence of evident barriers to dispersal (6,7). Despite the assumed oversimplification of the traditional categorization of speciation processes into allopatric, parapatric and sympatric, the spatial context and the extent of gene flow during speciation play a key role in determining whether, and how fast, reproductive isolation can evolve (8,9). Thus, the mechanisms of local adaptation and speciation are deeply influenced by the biogeographical and demographic history of populations, and may be enhanced during periods of major ecosystem fluctuations (7,10).

Northern Africa is a region of great interest to study biogeographic processes due to its wide diversity of habitats and heterogeneous landscapes, as well as to its complex paleo-climatic and geological history

(11). Available phylogeographic studies in this region have uncovered substantial taxa diversification induced by the great climate shifts that occurred during the Pliocene-Pleistocene interval (~5 million years ago) and the successive range shifts of the Sahara desert (12–15). The climatic fluctuations caused significant movements of the Sahara-Sahel boundaries, leading to alterations in the ecological compositions of landscapes. Such dynamics resulted in new selective pressures and/or phylogeographic isolation, causing events of genetic diversification, adaptation and eventually speciation (11).

As desert specialists, African Jerboas (*Jaculus* spp., Erxleben 1777, Dipodidae) have drawn attention of researchers due to their vast and broad distribution across the Saharan-Arabian region and their high phenotypic and genetic variation (16,17). Two cryptic species have been recognised within the African Jerboas, presenting a broad and sympatric distribution throughout North Africa with overlapped phenotypic variation but presenting putative divergent ecological preferences: the Lesser Egyptian Jerboa, *Jaculus jaculus* (Linnaeus 1758), with a paler orangish dorsum with whitish-grey vibrissae; and the African Hammada Jerboa, *Jaculus hirtipes* (Lichtenstein 1823), characterized by a darker dorsum with grey vibrissae (18) (Figure S1). However, the characterization of these two evolutionary entities has not been consistent across different studies. Some authors presented them as conspecific populations of the Lesser Egyptian Jerboa, which has been widely acknowledged among taxonomists (19). A more recent study, with new data collected in Israel and Sinai, suggests that the two species have different ecological requisites and gives new insights into their taxonomic status (20). The authors also state that, in the region of Israel and Sinai, these two species can be distinguished in the field according to fur and tail coloration and morphology of male external genitalia. Studies relying on the genetic diversity of mitochondrial (*cytb;* 16,17,19) and nuclear DNA (*WF;* 17), however, agree in distinguishing two divergent lineages within *Jaculus* sp. with a broad and sympatric distribution in North West Africa, and a high environmental and phenotypic overlap, including fur colour (17). Consequently, it was recently proposed that the two lineages present cryptic diversity (Figure S1a), and that they might persist genetically differentiated due to their ecological differences within the complex distribution patterns of sandy (lighter) and rocky (darker) habitats over North Africa (22) (Figure S1b). It is therefore crucial to apply a more comprehensive approach to study this species complex, in order to achieve a better understanding on the evolutionary history of these two forms, namely their level of genetic diversity, divergence, reproductive isolation and ecological diversification.

Here, we assess the evolutionary history of the two putative species of African Jerboas by applying an integrative approach based on multi-locus genetic analyses and ecological niche tests. Our sampling covers all the North African range, focusing on individuals from West African regions, where both lineages present an overlapping range at the macrogeographic scale. Our main aims were: (1) to evaluate the phylogenetic divergence between species by analysing several independent markers (nuclear and mitochondrial) using species delimitation and species tree inference methods; (2) to estimate divergence time and demographic history of the two lineages; (3) to assess the levels of gene flow between lineages through estimations of the current genetic structure and levels of admixture by analysing microsatellite data and isolation-with-migration (IM) models; and finally, (4) give insights into the processes underlying speciation, taking into account niche overlap tests (*i.e.* addressing niche conservatism *vs* divergence),

measures of gene flow and past demography of the species. Our vast sampling and interdisciplinary approach will contribute for a better understanding of the evolutionary history and diversification processes of North African biota.

# Results

*Phylogenetic relationships and species delimitation in Jaculus spp.*

The mtDNA phylogeny was performed by combining data from previous studies (17,21,22; see Methods) in order to have a reference for each evolutionary unit. This analysis recovered two main clades with high support, corresponding to the two putative species: *J. jaculus* and *J. hirtipes* (Figure 1a). The defined species hold a great diversity as evidenced from the great number of haplotypes and high support values for the internal nodes within species (Figure 1a). For both species, there is an Israeli clade being differentiated (Figure 1a), suggesting some geographic isolation and genetic substructure in this region. Further analyses were accomplished by classifying each individual into *J. jaculus* and *J. hirtipes* according to the mtDNA phylogeny. The geographic distributions of the two taxa highly overlap, thus confirming that *J. jaculus* and *J. hirtipes* persist in geographical sympatry at a macrogeographic scale (Figure 1b). The two species are also differentiated at nuclear loci, with nearly absent allele sharing (Figure 2). For *GHR* locus, one individual from Bojador in the Atlantic coast of Morocco is homozygote for one allele that clustered within *J. jaculus.* This individual clustered within *J. hirtipes* at all other loci. In *IRBP* and *Agouti* genes the opposite pattern occurred: one individual from the Inchiri region in Western Mauritania had alleles from *J. hirtipes,* whereas it grouped with *J. jaculus* in the other loci analysed (Figure 2).

Potential recombinants were detected in *DBX5* and *WF* and were removed for subsequent analyses. Bayesian species delimitation consistently supports three species with the maximum posterior probability (speciation probability = 1), corresponding to the outgroup species, *J. orientalis,* and the two species, *J. jaculus* and *J. hirtipes.* Moreover the probability of having three different species was 1 (P[3] = 1), leaving P[2] and P[1] with 0. The species tree inferred by *BEAST recovered two strongly supported speciation events—an ancient split separating *J. orientalis,* and a more recent speciation node delimiting *J. jaculus* and *J. hirtipes* (Figure 3). Calibration of the tree showed that the split between *J. orientalis* and other *Jaculus* species occurred during the Late Miocene-Pliocene transition, around 5.65 Mya (95% highest posterior density (HPD): 4.26–7.27 Mya). The split between *J. jaculus* and *J. hirtipes* is estimated to have occurred through the transition of Pliocene to Pleistocene, about 3.90 Mya (95% HPD: 2.96–4.96).

*Assessing the levels of gene flow*

Estimations of the effective population sizes detected higher values for *J. jaculus* (maximum-likelihood estimates and respective 95% posterior density intervals: 9.99 [5.08–9.99] millions) than *J. hirtipes (*5.45 [3.79–9.99] millions), with an ancestral population size of 9.99 (1.67–9.99) millions. The divergence time between the putative species is estimated to be about 4.73 (0.44–7.65) Myr before present (BP). Population migration rates were found to be significant in LLR tests (23), wherein a higher proportion of

migrants per generation was detected from *J. hirtipes* into *J. jaculus* (0.04 [0.0–7.3], $p < 0.01$; and from *J. jaculus* to *J. hirtipes:* 0.02 [0.0–7.62], $p < 0.05$). Nonetheless, the confidence intervals include 0, suggesting very limited levels of gene flow. The resulting evaluations of the posterior densities for all parameters were consistent across independent runs. Analyses were also performed without the two candidate genes for fur coloration in order to assess potential bias towards selected loci and results showed similar estimates (results not shown).

*Population genetics and demographic history*

Divergence was very high for *cytb* gene between *J. jaculus* and *J. hirtipes* (10.00%), but slightly lower to that observed between both species and the outgroup (*J. orientalis;* 12.00%). The *DBX* intron also reveals a high divergence between *J. jaculus* and *J. hirtipes (*3.00%), being even higher than the genetic diversity separating *J. orientalis* and *J. jaculus* (0.40%), but similar to the genetic variation between *J. hirtipes* and *J. orientalis* (3.30%). Divergence found in the autosomal loci was generally lower, with the exception of the *Agouti and WF* genes (Table 1).

*Table 1.* Average genetic divergence (Dxy) and net nucleotide divergence (Da) between *J. jaculus and J. hirtipes,* between *J. jaculus-J. hirtipes* and *J. orientalis,* and other related rodent species.

| Locus | | *J. jaculus/ J. hirtipes* | *J.orientalis/ J. jaculus* | *J.orientalis/* | *J. hirtipes* | *J.orientalis/ Jaculus spp.* | Other rodents |
|---|---|---|---|---|---|---|---|
| *Cytb* | Dxy | 10.00(1.00) | 12.00(1.00) | 12.00(1.10) | | 12.00(1.00) | 12.50(0.90)[a] 7.80(0.80)[b] |
| | Da | 8.80(0.90) | 10.80(1.00) | 11.40(1.00) | | 9.00(0.80) | - |
| *DBX5* | Dxy | 3.00(0.90) | 0.40(0.30) | 3.30(1.00) | | 2.40(0.70) | 1.10(0.60)[c] |
| | Da | 3.00(0.90) | 0.30(0.30) | 3.30(1.00) | | 1.70(0.50) | - |
| *ADRA2B* | Dxy | 0.50(0.20) | 1.10(0.30) | 0.70(0.20) | | 0.80(0.20) | 2.70(0.50)[d] |
| | Da | 0.30(0.20) | 0.60(0.20) | 0.30(0.20) | | 0.40(0.20) | - |
| *GHR* | Dxy | 0.20(0.10) | 0.80(0.20) | 0.60(0.20) | | 0.70(0.20) | 0.40(0.20)[e] |
| | Da | 0.10(0.10) | 0.50(0.20) | 0.30(0.20) | | 0.30(0.20) | - |
| *IRBP* | Dxy | 0.80(0.20) | 0.80(0.20) | 0.60(0.20) | | 0.70/0.20) | 0.50(0.20)[f] |
| | Da | 0.40(0.10) | 0.60(0.20) | 0.40(0.20) | | 0.40(0.10) | - |
| *WF* | Dxy | 1.30(0.30) | 3.00(0.50) | 3.30(0.50) | | 3.30(0.50) | 1.40(0.30)[g] |
| | Da | 0.80(0.20) | 2.60(0.50) | 2.80(0.50) | | 2.60(0.50) | - |
| *Agouti* | Dxy | 1.90(0.50) | 1.60(0.40) | 1.60(0.50) | | 1.80(0.50) | 0.70(0.40)[h] |
| | Da | 0.90(0.30) | 1.30(0.50) | 1.20(0.40) | | 1.00(0.40) | - |
| *MC1R* | Dxy | 0.80(0.20) | 1.30(0.40) | 1.60(0.40) | | 1.50(0.40) | 0.90(0.30)[i] |
| | Da | 0.50(0.20) | 1.10(0.30) | 1.40(0.40) | | 1.20(0.30) | - |

The standard errors are based on 10,000 replicates (in parenthesis), all estimates are given as percentages.

**Dxy** = Average number of nucleotide substitutions per site between populations (average raw DNA divergence); **Da** = Number of net nucleotide substitutions per site between populations (average net DNA divergence).

[a] *Microtus arvalis* (GQ352469) / *Microtus agrestis* (GQ352470) (24)

[b] *Microtus arvalis* (AY513809) / *Microtus kirgisorum* (AY513809) (26,27; respectively)

[c] *Microtus arvalis* (JX284377) / *Microtus agrestis* (JX284376) (27)

[d] *Acomys russatus* (FM162045) / *Acomys cahirinus* (FN984740) (29,30; respectively)

[e] *Allactaga bullata* (JQ347909) / *Allactaga balikunica* (KM397227) (31,32; respectively)

[f] *Allactaga bullata* (JQ347929) / *Allactaga balikunica* (KM397136) (30,31); respectively)

[g] *Microtus agrestis* (FM200055) / *Microtus socialis* (FM162067) (29,33; respectively)

[h] *Peromyscus polionotus* (DQ482897) / *Peromyscus maniculatus* (DQ482892) (33)

[i] *Peromyscus polionotus* (EU020068) / *Peromyscus maniculatus* (EU020066) (34)

The *cytb* gene displayed the highest intraspecific diversity, with higher values observed in *J. jaculus* than within *J. hirtipes* (Table 2). The *DBX5* intron exhibited the lowest diversity, and the autosomal genes, *IRBP, WF* and *MC1R* had intermediate levels, with the highest diversity values observed for *J. hirtipes,* contrary to that observed in the mtDNA (Table 2). The *Agouti* gene also presented exceptional high levels of nucleotide diversity in *J. hirtipes* but not in *J. jaculus.* Compared with other autosomes, *GHR* recovered the lowest values of genetic diversity (Table 2). Overall, neutrality tests show negative values for almost all loci in the two species for Tajima's D and Fu's Fs statistics (Table 2).

*Table 2.* Diversity estimates within *Jaculus* species.

| us | Species | L | n | S | H | Hd(SD) | π (SD)% | (SD)% | *D* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *tb* | *J. jaculus* | 897 | 96 | 156 | 87 | 0.997(0.002) | 1.61(0.09) | 3.39(0.86) | -1.84** | -89.13*** | 0.04** |
| | *J. hirtipes* | 897 | 137 | 102 | 83 | 0.980(0.005) | 0.58(0.03) | 2.07(0.51) | -2.35*** | -107.96*** | 0.02*** |
| *X5* | *J. jaculus* | 311 | 84 | 3 | 4 | 0.220(0.059) | 0.09(0.03) | 0.20(0.10) | -1.02 | -1.96 | 0.04 |
| | *J. hirtipes* | 306 | 180 | 7 | 8 | 0.208(0.040) | 0.07(0.01) | 0.40(0.20) | -1.77* | -8.94*** | 0.02 |
| *on-rec* | *J. jaculus* | 301 | 84 | 2 | 3 | 0.217(0.057) | 0.07(0.02) | 0.10(0.10) | -0.71 | -0.95 | 0.06 |
| | *J. hirtipes* | 296 | 171 | 6 | 7 | 0.166(0.038) | 0.06(0.01) | 0.40(0.20) | -1.73 | -8.21*** | 0.02 |
| *A2B* | *J. jaculus* | 693 | 72 | 7 | 9 | 0.705(0.031) | 0.30(0.02) | 0.20(0.20) | 0.32 | -0.87 | 0.13 |
| | *J. hirtipes* | 693 | 180 | 11 | 11 | 0.345(0.045) | 0.06(0.01) | 0.30(0.10) | -1.85* | -9.67* | 0.02 |
| *IR* | *J. jaculus* | 798 | 80 | 10 | 11 | 0.378(0.070) | 0.05(0.01) | 0.20(0.10) | -2.09* | -12.09** | 0.03* |
| | *J. hirtipes* | 798 | 183 | 11 | 10 | 0.147(0.036) | 0.03(0.008) | 0.20(0.09) | -2.12* | -12.93*** | 0.02* |
| *3P* | *J. jaculus* | 1058 | 50 | 25 | 19 | 0.905(0.023) | 0.35(0.05) | 0.53(0.20) | -1.07 | -6.48 | 0.07* |
| | *J. hirtipes* | 1058 | 106 | 23 | 35 | 0.948(0.010) | 0.39(0.002) | 0.41(0.13) | -0.19 | -19.37 | 0.09 |
| *7F* | *J. jaculus* | 873 | 40 | 31 | 23 | 0.938(0.026) | 0.57(0.07) | 0.83(0.28) | -1.11 | -10.77 | 0.08* |
| | *J. hirtipes* | 874 | 132 | 33 | 37 | 0.933(0.012) | 0.56(0.02) | 0.69(0.20) | -0.70 | -15.72 | 0.07 |
| *n-rec* | *J. jaculus* | 516 | 26 | 6 | 6 | 0.717(0.079) | 0.38(0.05) | 0.30(0.15) | 0.76 | 0.14 | 0.16 |
| | *J. hirtipes* | 514 | 127 | 18 | 14 | 0.861(0.016) | 0.44(0.27) | 0.65(0.21) | -0.89 | -2.48 | 0.06 |
| *uti* | *J. jaculus* | 453 | 66 | 14 | 12 | 0.746(0.037) | 0.39(0.05) | 0.65(0.24) | -0.06 | -0.06* | 0.10*** |
| | *J. hirtipes* | 461 | 162 | 14 | 15 | 0.765(0.021) | 1.07(0.02) | 0.54(0.18) | -0.09* | -0.36 | 0.08*** |
| *1R* | *J. jaculus* | 894 | 56 | 9 | 11 | 0.890(0.016) | 0.19(0.01) | 0.22(0.09) | -0.06 | -0.08 | 0.11 |
| | *J. hirtipes* | 894 | 110 | 13 | 19 | 0.890(0.019) | 0.25(0.02) | 0.28(0.10) | -0.06 | -0.21* | 0.09 |

In loci with significant levels of recombination, values are shown for the recombinant and non-recombinant datasets (denoted as “none-rec’”).

**L**, number of sites excluding gaps; **n**, number of sequences; **S**, number of segregating sites; **H**, number of haplotypes; **Hd**, haplotype diversity; **π**, nucleotide diversity per site; , computed from the number of segregating sites; **D**, Tajima’s D; **,** Fu’s ; , Ramos-Onsins & Rozas’s . Significant values indicated *(P<0.05), **(P<0.01), ***(P<0.001).

Estimated effective population sizes through time revealed signs of expansion in both *J. jaculus* and *J. hirtipes,* which may have started around 200,000 years BP and that is still ongoing (Figure 4). The analysis suggests that the demographic expansion may have started approximately at the same time in the two species, with a relatively earlier expansion observed in *J. jaculus.* Estimations of contemporary population sizes show that *J. jaculus* has approximately triple the population size as *J. hirtipes* (~12 and ~4 millions in *J. jaculus* and *J. hirtipes* respectively, Figure 4).

*Population structure and admixture*

Six microsatellite loci (Jac04, Jac07, Jac11, Jac12, Jac24, and Jac27) revealed significant deviations from the Hardy-Weinberg equilibrium, presenting heterozygote deficiency (Table S1). Moreover, one locus (Jac01) amplified only samples belonging to *J. jaculus.* After removing these markers, assessments of population structure were performed with the six remaining loci for a total of 132 specimens (40 and 92 for *J. jaculus* and *J. hirtipes* respectively). Structure Harvester results highlighted K = 2 as the most likely number of clusters best explaining the variation in our dataset (both for DeltaK and L(K) methods; results not shown). Structure bar plot exhibited a clear separation between the two species (Figure 5). Additional intraspecific substructure was identified within *J. hirtipes* for *K* = 3 (Figure S3a), where some level of admixture was observed for one individual in the two species, although with no geographical structure (Figure S3b). The Principal Coordinate Analysis based on the genetic distances of microsatellite variation shows that 16.53% and 5.30% explain the diversity between and within species, respectively (Figure S4). Nevertheless, high levels of polymorphism were detected both for the whole dataset (i.e. two species as a single group) and within species, with similar allelic diversity observed between species for all microsatellite markers, varying from 9 to 29 alleles (Table 3). Estimates of the F-statistics show significant values of FST between species, indicating high differentiation between them (Table 3).

*Table 3.* Mean heterozygosity (observed and expected) and F-statistics for *J. jaculus* and *J. hirtipes* based on microsatellite loci.

| Cluster | | N | Na | Ne | Ho | He | uHe | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *J. jaculus* | Mean (SE) | 42 | 12.83 (1.38) | 5.74 (0.81) | 0.80 (0.03) | 0.81 (0.03) | 0.82 (0.03) | - | - | - |
| *J. hirtipes* | Mean (SE) | 90 | 12.83 (2.44) | 5.86 (1.39) | 0.69 (0.13) | 0.70 (0.14) | 0.71 (0.14) | - | - | - |
| Total | Mean (SE) | 132 | 12.83 (1.34) | 5.80 (0.777) | 0.75 (0.07) | 0.76 (0.07) | 0.76 (0.07) | 0.02 (0.01) | 0.18* (0.03) | 0.16* (0.03) |

**SE** – Standard error; **N** - Sample Size; **Na** – Number of Alleles; **Ne** - Number of Effective Alleles; **Ho** - Observed Heterozygosity; **He** - Expected Heterozygosity; **Uhe** - Unbiased Expected Heterozygosity; - Inbreeding Coefficient; – Total Fixation Index; – Fixation Index. Significant values indicated *(P<0.5).

*Niche overlap*

Overall, the observed niche overlap (Schoener's D) is high (D > 0.4) for the 5x5 km scale and for topo-climatic factors at 1x1 km scale (Figure S5). However, niche overlap for habitat measured in the 1x1 km scale was relatively low (D = 0.25). Niches were detected not to be equivalent (*i.e.* niches not constant when randomly reallocating individuals between the two species' ranges) since equivalency tests were significant in all cases ($p < 0.05$) (Figure S5). Similarity tests were also significant ($p < 0.05$) and the value of D (in red, Figure S5) is placed in the second tail of the distribution, therefore, the species tend to have similar patterns of topo-climate and habitat selection, more than expected by chance.

# Discussion

*Two closely-related species: the African Hammada Jerboa and the Lesser Egyptian Jerboa*

Our comprehensive approach clarified the phylogenetic relationship between the two jerboa lineages, with a widespread and overlapping distributions across North Africa (Figure 1). The phylogenetic inferences of mitochondrial DNA revealed two well defined and strongly supported clades (Figure 1a), as shown in previous studies (17,20–22). Moreover, we showed for the first time that the two mtDNA lineages can be further distinguished by seven single copy nuclear markers (Figure 2) and six microsatellite loci (Figure 5). By applying the recently developed coalescent methods of species delimitation and species tree inference (35), two well delimited clades with the maximum support values can be observed (Figure 3). Hence, we have revealed that the loci analysed both at nuclear and mitochondrial DNA agree in the identification of two different species.

Average *cytb* nucleotide divergence (10.0%) was slightly lower to previously documented for these species (10.5% (21); 10.6% (17), but beyond intraspecific variation usually observed in rodents (average 2.1%, up to 6.29 (16,27,36)). Moreover, the observed divergence is slightly above the average genetic distance observed between sister rodent species (average: 9.6%, range 2.7–19.2 (36,37)). In particular, the divergence between the two jerboas were considerably higher than between closely-related *Microtus* species: *M. arvalis* and *M. kirgisorum (*7.8%), but lower than between distant taxa: *M. arvalis* and *M. agrestis* (12.5%; Table 1) (26,29). For nuclear loci, the genetic divergence observed between *J. jaculus* and

*J. hirtipes* in the *IRBP, DBX5* and *Agouti* genes was higher than that observed between other closely-related rodent species, while for *ADRA2B* the values were considerably lower (Table 1). The remaining autosomal genes had similar values of genetic differentiation (Table 1). Altogether, the observed genetic divergences between *J. jaculus* and *J. hirtipes* is compatible with their classification as two different species.

*Insights into the evolutionary history of Jaculus species*

Our species tree inference estimates a divergence time between *J. orientalis* and *J. jaculus-J. hirtipes* during the transition of Late Miocene to Pliocene, around 5.65 (4.26—7.27) Myr BP (Figure 3). The most recent common ancestor (MRCA) of *J. jaculus* and *J. hirtipes* might have colonized a wide region from Central Asia to North Africa (around 5.15 Myr BP; 32), as observed in phylogeographic studies on a wide range of desert organisms, both animals (38) and plants (39). The split between *J. jaculus* and *J. hirtipes* was reported to have been held along the Late Pliocene-Early Pleistocene boundary, about 4.73 (0.44–7.65) Mya according to IM and around 3.90 Mya (2.96—4.96) based on Beast (Figure 3). Although these estimates indicate an older divergence of *Jaculus* species when compared to other rodent species such as *Acomys* (1.25 [0.65–1.94] Myr; (38) or *Mastomys* (2.82 [1.61–4.20] Myr; (40), this should be interpreted with caution due to the lack of accurate substitution rates in these rodent groups, and the unavailability of dated fossil records to time-calibrate the phylogeny. Nevertheless, according to the dated estimates, the divergence between these species could be a consequence of the fluctuations across the Northern African range, probably influenced by the recurrent humid phases, which are believed to counteract expansion events of xeric species, like jerboas (11). Following a period of geographical isolation, perhaps in different refugia, the two lineages might then have spread during the subsequent warmer periods triggered by the Quaternary climatic alterations.

Previous assessments on the historical demography of *Jaculus* species indicated potential signs of expansions, with two different evolutionary histories between species (17). Our results corroborate these suggestions, with detected differences between the species regarding times of expansion and effective population sizes (Figure 4). Neutrality tests and reconstructions of population dynamics in both species rejected a demographic model of population at equilibrium (Table 2), and indicated signs of population expansion (Figure 4). These could have started around 200,000 years BP, coinciding with the major climatic oscillations of the Upper Pleistocene of North Africa that induced critical changes in the genetic signature of several vertebrate species, including other West African rodents (38,41–44). However, we cannot exclude that this pattern of population expansion results from our sampling based on the collection of individuals from different locations rather than entire populations. This could have increased the number of rare alleles, originating a pattern similar to that of a demographic expansion. Future studies focused on the analyses of populations should allow to distinguish between these two different hypotheses.

*Assessing gene flow between J. hirtipes and J. jaculus*

*Jaculus jaculus* and *J. hirtipes,* are often found in sympatry in North Africa, thus increasing the probability of hybridization. Two out of 152 analysed individuals presented alleles at two nuclear markers that are typical from the other species, which could result from incomplete lineage sorting or introgression. However, the IM analysis supported gene flow between the two species in both directions, although higher towards *J. jaculus.* The microsatellite data further suggests potential admixture among species (Figure 5) but the majority of the individuals also revealed a high membership probability to the respective species (Figure 5). Despite being significant, the IM-estimated levels of gene flow had confidence intervals that included 0, suggesting that the level of isolation between species might be very high. Moreover, these estimates were lower than those usually reported between subspecies of mammals, where *2Nm* values can go up to 1.50 (e.g., 46,47). The limited levels of gene flow suggest that *J. jaculus* and *J. hirtipes* remain strongly differentiated genetically, suggesting strong reproductive isolation.

*What drives speciation in this system?*

Population divergence in the presence of gene flow is often referred as evidence that local adaptation is a crucial driver of differentiation between two or more populations (47–49). Persistent habitat-phenotype covariation within jerboas (and other desert rodents) suggests that natural selection may be triggering the phenotypic divergence (22,50). Indeed, previous studies have suggested that, despite the coexistence of two jerboa species in sympatry across much of Sahara-Sahel, they might segregate into distinct micro-habitats, perhaps in response to strong predator driven selection (17,22). Hence, these species might persist in different micro-habitats associated with the admixture of sandy (lighter) and rocky (darker) micro-habitats over North Africa, where *J. jaculus* and *J. hirtipes* are more frequent, respectively (22). A tidier micro-habitat preference was previously suggested for *J. jaculus,* implying that *J. hirtipes* may be competitively excluded to suboptimal micro-habitats, which could explain its observed lower effective population size. We found a strong niche overlap between species and similar patterns of habitat selection (Figure S5). This might explain the observed overlapping distribution in fur colour variation in both species (Figure S1a). However, when tests are performed at a local scale (i.e. 1x1 km), the habitat component of the niches has lower overlap (Figure S5), thus suggesting that the two species might persist in ecological separation at a micro-habitat scale. It is thus possible that the observed divergence between species might have arisen through substantial reproductive isolation, and that natural selection at the micro-scale (lower than 1 km) played a key role in the speciation process, a pattern also observed in other organisms (e.g. marine sea snails, (51). From this, two speciation hypotheses arise: (1) sympatric speciation driven by local adaptation at the micro-habitat scale, allowing the two specimens to persist in reproductive isolation; and (2) allopatric speciation as a consequence of the recurrent climatic oscillations in North Africa followed by geographic isolation and adaptation to the two distinct micro-habitats. The first hypothesis seems more likely by looking at the overlapping geographic distribution of the two species, the similar ecological preferences at a macro-geographical scale and the overlapping range of fur colour variation. However, more studies are needed in order to fully disentangle the two modes of speciation. Moreover, mating preferences experiments are required in order to test if fur colour is a determinant factor for reproduction, which would clarify the drivers of reproductive isolation between the two species.

# Conclusions

Our comprehensive analyses, based on both mitochondrial and nuclear DNA, provide evidence for two different species of African Jerboas that have a similar distribution across North Africa: *J. jaculus* and *J. hirtipes.* Our results propose that these two species might have experienced demographic expansions since the Late Pleistocene period that is still ongoing, with a higher effective population size observed for *J. jaculus.* Despite the detection of gene flow between species, the two species persist strongly differentiated. Moreover, analysis of niche divergence suggests that *J. jaculus* and *J. hirtipes* are ecologically separated at a micro-habitat scale. These findings suggest that natural selection at a micro-scale could have driven the speciation process. However, the relatively deep divergence at multiple loci also suggest that this could have involved some geographic isolation. Further analyses to assess levels of gene flow and loci involved in adaptation across the genome are thus required to fully understand the processes driving the observed diversification of North African jerboas.

# Methods

*Sampling and DNA extraction*

A total of 231 samples distributed throughout North Africa, including 152 tissue samples collected in the field and 79 samples obtained from museum collections, were used in this study (Table S2 and Figure 1). Tissue samples were collected from road-killed and live trapped animals during several field expeditions in North-West Africa or received from collaborators between November 2011 to February 2015 (50,52,53); Table S2). Tissue samples were preserved in 96% ethanol for genetic analyses at the moment of collection. A total of 54 samples were already used in previous studies, for *cytb* (51 samples) and *WF (*21 samples) (17,22); Table S2), but their genomic DNA was re-extracted and analysed for all markers used in this study. Additionally, 10 samples of *J. orientalis* were extracted and included as an outgroup species (Table S2). Extractions of the genomic DNA from tissue samples were performed using EasySpin Kit, following the “Genomic DNA Minipreps Tissue Kit” protocol. Extractions of museum samples were performed in a separate and autonomous facility, under sterile conditions, using the QIAamp® DNA Micro Kit (QIAGEN), following the “Isolation of Total DNA from Nail Clippings and Hair” protocol. Extracted DNA was stored at 20ºC.

*DNA amplification and sequencing*

One mitochondrial locus (cytochrome *b, cytb,* 897 bp) and seven nuclear loci were amplified, including two candidate genes for colour morph variation (the complete coding region of the melanocortin 1 receptor, *MC1R;* and a fragment of the exon 2 of the *Agouti* geneand part of an intron), one X-linked gene (intron 5 from the developing brain, homeobox gene, *DBX,* gene) and four autosomal genes (exon 10 from the growth hormone receptor, *GHR;* exon 1 from alpha−2B adrenergic receptor, *ADRA2B;* exon 1 from the interstitial retinoid binding protein, *IRBP;* and exon 28 from the on Willebrand factor, *WF),,* producing a total of 5,369 bp. Partial amplification of the *cytb* gene (897 bp) was performed for the entire set of

samples (231 samples, contemporaneous and museum) using two primer pairs previously designed for *Jaculus* species (Jac1Fw, Jac1Rv, Jac4Fw, Jac4Rv; 17). The reconstruction of the DNA fragment with the museum samples was done in several steps to produce overlapping sequences in order to obtain the entire fragment. In some cases, only a short fragment (325 bp) of the gene was amplified, which was obtained combining two pairs of primers, Jack4Fw and Jack1Rv (Primers, references and PCR conditions for *cytb* are described in Table S3). The short fragment was only used to confirm the phylogeny with the long fragment. Nuclear loci and microsatellites were amplified only on samples collected during field work (152 samples; Table S3). PCR products of both mitochondrial and nuclear genes were purified with a commercial kit (Qiagen) and both strands were sequenced on an ABI 3130xl Genetic Analyser (AB Applied Biosystems). For the autosomal genes, sequencing of both strands was performed in an external laboratory (Macrogen Inc.). Additionally, available sequence data for the *cytb* gene of our target species (164 sequences) was downloaded from GenBank and included in the analyses (Table S4).

*Sequence alignment and phylogenetic analyses*

Each sequence was verified and manually aligned using SEQSCAPE v2.6 (54). Alignments for each locus were generated with CLUSTAL W (55) implemented in ClustalX v2.0 (56) and edited manually in BIOEDIT v7.1.3 (57) in order to minimize the number of base pairs in the alignment spanned by insertion/deletions (indels). Polymorphic positions for each sequence from nuclear loci were carefully examined to ensure precise and consistent identification of double peaks in heterozygotes. Heterozygous sequences for indels were resolved manually from offset chromatogram peaks (58). Nuclear haplotypes were inferred using PHASE v2.1 (59,60) with three runs performed for each locus with 10,000 burn-in steps and 10,000 interactions. Input files were created in SEQPHASE (61). Phased heterozygotes holding indels were included in SEQPHASE as “known haplotype pairs”. Haplotypes presenting probability phase calls below 80% were discarded from the analysis to ensure that only reliable haplotypes were used in downstream analyses. The indels observed in the *DBX (*21 and 42 bp; Figure S2) and in the partials *Agouti* gene (8 bp) were coded manually and were included in network reconstruction but excluded in further analyses due to their large sizes. Haplotypes for the *cytb* gene were inferred with DnaSP v5 (62).

Phylogenetic analyses were performed for the *cytb* locus. The Akaike information criterion (AIC; (63) was used to select the best-fit model of sequence evolution for each locus alignment among the 88 available in the software jModelTest v2.1.4 (65; Table S5). The phylogenetic relationships between haplotypes were inferred by the Maximum-Likelihood (ML) approach in PHYML v3.0 (65) and the Bayesian phylogenetic inference (BI) implemented in MrBayes v3.2.0 (66). ML analyses were performed with 1000 bootstrap pseudo replicates. Bayesian posterior probabilities were assessed from two runs with four chains of 1 million generations for the nuclear genes 50 million generations for *cytb,* with a sampling frequency that provided a total of 10,000 samples for each run, discarding 25% of them as burn-in. Tracer v1.5 (67) was applied to evaluate the convergence of the ESS values (effective sample size) for each analysis. Resulting trees were drawn with FIGTREE v1.3.1 (68).

Haplotype networks were generated for each nuclear gene individually using parsimony calculations in TCS v1.21 (69) considering gaps as a fifth state. Each indel of the *DBX5* and *Agouti* locus was considered as a single mutational step, regardless of the corresponding size (Figure 2). Analyses were performed for each locus with a connection limit of 95%. *DBX* locus presented disconnected haplotypes and so networks were redrawn with the connection limit fixed at 90% in order to link the more unrelated groups and see the number of mutational steps among them. Networks were edited using tcsBU (70).

*Species delimitation and species tree inference*

Alignments were first tested for the presence of within-locus recombination in SPLITSTREE v4.13.1 (71), and the ones found to be significant were further analysed with IMgc (72) to reduce the dataset to the largest non-recombinant blocks. Moreover, in order to validate the assignment of individuals to the two previously described mitochondrial lineages (16,17,20–22), the program Bayesian Phylogenetics and Phylogeography (BP&P) v3.1 was used to assess the status of species delimitation. Our analyses included all the loci analysed in this study. Due to the large sample size of our dataset, only 30 individuals, chosen randomly, were analysed for each lineage on each locus. The same outgroup sequences of *J. orientalis* were used for this analysis. Population size parameters (θ) and divergence time at the root of the species tree (τ) were estimated with the gamma prior G(2, 1000), while the Dirichlet prior was assigned to all other divergence time parameters. We used “algorithm 0” with the fine-tune parameter set to default. Each species delimitation model was assigned equal prior probability. For the MCMC, samples were collected for 1,000,000 generations, with a sampling interval of 2 and a burn-in of 10%. Each analysis was run 3 times to confirm consistency among runs.

The complete dataset was also used to infer the species tree by applying the multispecies coalescent model that is implemented in *BEAST (35), part of the BEAST v2.3.0 package (73). The input file was produced with the application BEAUti v2.3.0, also including in the BEAST package. Preliminary analyses were carried out to evaluate the clock-like evolution of each locus by inspecting the posterior distribution of the standard deviation of an uncorrelated lognormal relaxed clock. Based on these trial runs the final analysis was accomplished with an uncorrelated lognormal relaxed clock, using the HKY+I+G substitution model for *cytb.* Analyses of the nuclear loci were carried out with an HKY (+I for *WF, ADRA2B, IRBP, MC1R* and *Agouti)* substitution model under a strict molecular clock (Table S5).

Times of divergence were estimated using *cytb* as the reference gene. A fossil-based calibration of substitution rates was not possible due to the poor fossil record of *Jaculus* in North Africa. Similarly, the well-known calibration point Muridae-Rodentia was not used due to the likely saturation effect associated with the ancientness of the divergence between Muridae and Dipodidae. Instead, a universal rodent *cytb* substitution rate was applied after an extensive literature review (21,74,75). Following these assumptions, the prior of the relaxed clock standard deviation was set to a normal distribution with a mean of 0.02 with *sigma* fixed at 0.005, meaning a substitution rate of 2% per million years. This substitution rate was used in all subsequent analyses. The coalescent constant population size was used as tree prior and all the remaining priors were set to defaults. Three independent runs of 300 million generations were

implemented, sampling trees and parameter estimators every 30,000 generations for all loci. The convergence of the runs was verified after the removal of a 10% burn-in using TRACER v1.5. Visual inspection of trace plots indicated a good sampling of all parameters for each *BEAST independent runs, with effective population sizes (ESS) above 1000, suggesting a good convergence of all parameters. The results from all runs were combined with LogCombiner v2.3.0, and the subsequent maximum clade credibility summary trees with posterior probabilities for each node were generated with TreeAnnotater v2.3.0 from the BEAST package. All the trees were visualized and edited with FIGTREE v1.3.1.

*Isolation-with-Migration analyses*

Species tree inferences performed with *BEAST incorporate the uncertainty associated with the coalescent process while estimating the phylogeny. However, it does not assume the possibility of occurrence of gene flow after the initial split. Thus, models of isolation-with-migration (IM) (23) implemented in the IMa2 software (76) were applied to infer whether gene flow has occurred between the two described putative species. This method estimates the multi-locus effective population sizes (for present and ancestral populations), divergence times and migration rates under a model of isolation with migration (IM) (23,77). Analyses were performed with the non-recombining dataset and considering the two *Jaculus* species as populations. After preliminary runs, three independent runs with different starting seeds were performed by sampling 10,000 genealogies per locus with 10% burn-in. We used a geometric model with the first heating term (ha) set to 0.95 and the second (hb) to 0.85 sampling through 150 chains (hn). Priors for population size, splitting times and migration rates were set to 10, 5 and 10 respectively. HKY mutation model was applied to all loci and the same substitution rate for *cytb* (0.02, ranging from 0.01 to 0.03) was used in order to obtain the results in demographic units, considering 1 year of generation time (75). Moreover, the likelihood ratio test (LLR) described by Nielsen and Wakeley (23) was used to assess whether migration rates were significantly different from zero, sampling over 20,000 trees, as implemented in the Load-Genealogy mode (L-mode) of IMa2. In order to get the estimates in demographic units, locus specific mutation rates were estimated from the average number of substitutions per site (Dxy) between *J. jaculus-J. hirtipes* and *J. orientalis,* calculated using Kimura 2-parameter distances in MEGA v5.10 (78), considering a split time of 5.97 Myr (31) and a generation time of 1 year (75). The geometric mean of the locus-specific mutation rates was then used to calculate the effective population sizes and divergence times from the IMa2 highest posterior density of each parameter.

*Population genetics and demographic analyses*

Total (Dxy) and net (Da) divergences between lineages were calculated using p-distance parameter in MEGA v5.1. Additionally, the divergence among several related rodent species, based on published data, was inferred for comparison analysis (24,25,34,26–33). Standard deviations for these divergences were estimated from 10,000 bootstrap replications. Nucleotide diversity ($\pi$), theta computed from the number of segregating sites (W) and haplotype diversity (Hd) were calculated per lineage for each locus analysed. Three test statistics, Tajima's *D* (79), Fu's *Fs* (80) and $R_2$ (81) were performed to investigate deviations

from its neutral expectations, which could imply recent population expansion and/or signs of selection. Significance was evaluated through 10,000 coalescent simulations. These statistics were assessed per locus for each lineage in DnaSP v5. Calculations were made separately for the entire data set and for the non-recombinant portions obtained with IMgc.

The dynamics of effective population sizes through time of the two lineages of *Jaculus* sp. was inferred with Extended Bayesian Skyline Plots (EBSP; 85), using a linear model in BEAST v2.3.0 and inputted through BEAUti v2.3.0. The same non-recombinant dataset used for species tree inference was analysed. The evolutionary models for each locus of each lineage were estimated in jModelTest v2.1.4, which resulted in similar models to the ones previously obtained (Table S4). After preliminary analyses the evolutionary rates of the mitochondrial and nuclear loci were set to a strict molecular clock. The prior for the mean distribution of population sizes was optimized according to the population sizes estimated in preliminary runs and was set with a coalescent prior and a constant population size (82). Remaining priors were set as default. The MCMC parameters were the same as applied in *BEAST analysis. TRACER v1.5 was used to assess the convergence of the independent runs. Results of the separate runs were combined with LogCombiner v2.3.0, part of the BEAST package, after discarding 10% burn-in.

*Microsatellite selection and optimization*

Since there were no specific microsatellite markers available for *Jaculus* spp. or closely related species, a microsatellite library was developed through high-throughput genomic sequencing (454 pyrosequencing) at GenoScreen (http://www.genoscreen.fr/en/) using *J. jaculus* individuals from distinct regions in North Africa. Detailed description of the optimization procedure can be found in supplementary material (supplementary information S1). After optimization we used two multiplexes amplifying seven and four markers each, as well as two additional loci that had to be amplified individually in separate PCR reactions (Table S6).

*Microsatellite genotyping*

A total of 148 contemporary samples were genotyped for 13 microsatellite loci. Multiplex and individual reactions, primers concentrations and amplification conditions are summarized in supplementary information S1. Allele data was obtained using GENEMAPPER v4.0 (Applied Biosystems 2006). Sizing bin windows were created manually and the automated scoring was checked by three independent observers to minimize genotyping errors. In order to assure consistency of results, 30% of the dataset was repeatedly genotyped in three independent runs. Inconsistent genotypes (~2% of all genotypes) were considered as missing data.

*Microsatellite analysis*

As the sampling was continuous across the distribution and it is hard to delimit populations, these analyses were performed considering the two *Jaculus* species as two different populations. MICROCHECKER v2.2.3 (83) was used to assess the presence of genotyping errors due to null alleles and

allele dropout. Linkage disequilibrium (LD) and deviations from Hardy-Weinberg Equilibrium (HWE) were estimated with GENEPOP on the Web (genepop.curtin.edu.au). The significance of the analysis were inferred according to the Bonferroni correction (0.05/[number of populations*number of loci]), and confirmed with three independent analyses. Loci presenting significant deviations from HWE and from LD assumptions and with large amount of missing data (above 40%) were discarded from further analyses. Measures of genetic diversity and differentiation, such as allele frequencies, mean number of alleles sampled per locus and population and the corresponding allelic richness, observed (Ho) and expected (He) heterozygosity, and F-statistics were estimated with FSTAT v1.2 (84). Individual-by-individual genetic distances that were used to compute a Principle Coordinate Analyses (PCoA) were calculated with GENALEX v6.0 (85). The number of clusters and the quantification of admixture between lineages were inferred with the Bayesian Clustering software STRUCTURE v2.3.3 (86). Analyses were accomplished by applying the admixture model with correlated allele frequencies. The software was run for the number of clusters (K) between 1 and 10 with 5 replicates of 1,000,000 MCMC iterations for each K value, following a burn-in period of 100,000 steps. Three independent analyses were performed to ensure similar posterior probabilities between runs. STRUCTURE HARVESTER v0.6.92 (87) was used to determine the probability of each K value. The most likely number of clusters (populations) was assessed using the mean values of likelihood [L(K)] and Delta K (88).

*Niche overlap*

Resemblance of ecological niches between species was tested: for overlap using Schoener's D Index (which ranges from 0, no overlap; to 1, totally overlap), for niche equivalency (i.e. whether the niche overlap is constant when randomly reallocating the occurrences of both entities among the two ranges) and for niche similarity (i.e. whether the environmental niches are more similar than what expected by chance; (89). The PCA-environment ordination approach developed by Broennimann *et al.* (90) was used for analyses. Tests were performed for two regions and scales, for the entire North Africa at ~5x5 km scale and for North-West Africa only (i.e. Mauritania and southern Morocco) at ~1x1 km scale, over two types of background data, composed by topo-climatic or habitat variables. For the North African region, a total of 125 records for *J. jaculus* and 122 records for *J. hirtipes* were included, after reducing spatial clustering by removing records located at lower than ~10 km distance from each other using the "occ.desaggragation" function (90). For the North-West region, a total of 59 records for *J. jaculus* and 97 *J. hirtipes* were retained, using ~1 km as distance threshold to remove records and reduce spatial clustering. In both scales, the background area was delimited accordingly to a minimum convex polygon. Background data consisted of two datasets: (1) topo-climatic, including two topographic, altitude and slope, and 19 bioclimatic variables; and (2) habitat, including six Euclidian distances to habitat categories. Altitude and the 19 bioclimatic variables were downloaded from WorldClim (www.worldclim.org/bioclim). Slope was derived a digital elevation model using the "slope" function from ArcGIS (ESRI 2011). Four of the habitat variables were constructed from land-cover categories for the years 2004–2006, which are likely descriptors of species natural habitats and showed a reasonable spatial representation in both study areas (i.e. sparse vegetation, bare, rocky and sandy areas). The remaining two habitat variables were constructed from spatial representation of water features

(secondary rivers and rock pools) which were digitized from the cartographic maps (91). Distance to these six habitat categories was computed using the “Euclidian distance” function from ArcGIS.

## Declarations

# Ethics approval and consent to participate

Local authorities (the HautCommissariat aux Eaux et Forêts et à la Lutte Contre la Désertificationof Morocco, decisions 20/2013, 41/2014, 42/2014, and the Ministère del’Environnement et du Développement Durable of Mauritania, decision227/ 08.11.2012) approved capturing and handling of animals, and all the procedures adhered to their guidelines and regulations.

# Consent for publication

Not applicable.

# Availability of Data and Materials

DNA sequences and microsatellite genotypes are provided in the supplementary data online with the following temporary link: https://figshare.com/s/1befb0c7c1c4767384aa (it will become public after publication). All supplementary tables and figures referenced in the text are within the file entitled “SupplementaryMaterial” in the supplementary data online (above link).

# Competing interests

The authors declare that they have no competing interests.

# Funding

The study was financially supported by the Academy of Finland to TM (Grant No. 268670) and Portuguese Foundation for Science and Technology to ZB (Grant No. PTDC/BIA-ECO/28158/2017). ZB, FMF and JCB were supported by the Portuguese Foundation for Science and Technology (SFRH/BPD/84822/2012, ICETA/EEC2018/10, SFRH/BD/73680/2010, respectively). Field and museum sampling expeditions were supported by National Geographic Society (grant: GEFNE53–12) and European Commission SYNTHESIS (grants: BE-TAF–1796 and AT-TAF–1665) programs to ZB.

# Author Contributions

AFM and ZB planed, organized and coordinated the work. AFM, ZB, JCB, FMF, TM and NS collected specimens and together with PCA, RF, JP, TLS and GS participated in writing manuscript. FMF, ZB, NS and JCB participated in ecological and AFM, PCA, RF, JP, NS, TLS and GS in molecular analyses.

# Acknowledgments

We acknowledge Pedro Santos Lda (Trimble GPS), Off Road Power Shop, P. N. Banc d'Arguin (Mauritania), Abdeljebbar Qninba and Mohammed Aziz El Agbani (University Mohammed V Agdal, Rabat, Morocco) for logistic support during field expeditions. We acknowledge Carole Paleco (Royal Belgian Institute of Natural Sciences, Brussels), Patricia Mergen and Emmanuel Gilissen (Royal Museum for Central Africa, Tervuren), and Frank Emmanuel Zachos (Museum of Natural History, Vienna) for support during the realization of SYNTHESIS museum grants.

## References

1. Seifert B. Cryptic species in ants (Hymenoptera: Formicidae) revisited: we need a change in the alpha-taxonomic approach. Myrmecological News. 2009;12:149–66.
2. Jowers MJ, Amor F, Ortega P, Lenoir A, Boulay RR, Cerdá X, et al. Recent speciation and secondary contact in endemic ants. Mol Ecol. 2014; 23:2529–2542.
3. Schluter D. Evidence for ecological speciation and its alternative. Science. 2009 Feb 6;323:737–41.
4. Butlin RK, Saura M, Charrier G, Jackson B, André C, Caballero A, et al. Parallel Evolution of Local Adaptation and Reproductive Isolation in the Face of Gene Flow. Evolution. 2014;68:935–49.
5. Rundle HD, Nosil P. Ecological speciation. Ecol Lett. 2005;8(3):336–52.
6. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 1981;17:368–76.
7. Smadja CM, Butlin RK. A framework for comparing processes of speciation in the presence of gene flow. Mol Ecol. 2011;20:5123–40.
8. Butlin RK, Galindo J, Grahame JW. Review. Sympatric, parapatric or allopatric: the most important way to classify speciation? Philos Trans R Soc Lond B Biol Sci. 2008;363:2997–3007.
9. Butlin RK, Debelle A, Kerth C, Snook RR, Beukeboom LW, Castillo Cajas RF, et al. What do we need to know about speciation? Trends Ecol Evol. 2012;27(1):27–39.
10. Pauls SU, Nowak C, Bálint M, Pfenninger M. The impact of global climate change on genetic diversity within populations and species. Mol Ecol. 2013;22:925–46.
11. Brito JC, Godinho R, Martínez-Freiría F, Pleguezuelos JM, Rebelo H, Santos X, et al. Unravelling biodiversity, evolution and threats to conservation in the sahara-sahel. Biol Rev. 2014;89:215–31.
12. Douady CJ, Catzeflis F, Raman J, Springer MS, Stanhope MJ. The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). Proc Natl Acad Sci U S A. 2003;100:8325–30.

13. Carranza S, Arnold EN, Geniez P, Roca J, Mateo JA. Radiation, multiple dispersal and parallelism in the skinks, Chalcides and Sphenops (Squamata: Scincidae), with comments on Scincus and Scincopus and the age of the Sahara Desert. Mol Phylogenet Evol. 2008;46:1071–94.
14. Gonçalves D V, Brito JC, Crochet PA, Geniez P, Padial JM, Harris DJ. Phylogeny of north african agama lizards (reptilia: Agamidae) and the role of the sahara desert in vertebrate speciation. Mol Phylogenet Evol. 2012;64:582–91.
15. Leite JV, Álvares F, Velo-Antón G, Brito JC, Godinho R. Differentiation of North African foxes and population genetic dynamics in the desert—insights into the evolutionary history of two sister taxa, Vulpes rueppellii and Vulpes vulpes. Org Divers Evol [Internet]. 2015; 15:731–745.
16. Ben Faleh A, Cosson JF, Tatard C, Othmen A Ben, Said K, Granjon L. Are there two cryptic species of the lesser Jerboa jaculus jaculus (Rodentia: Dipodidae) in tunisia? evidence from molecular, morphometric, and cytogenetic data. Biol J Linn Soc. 2010. 99(4):673–86.
17. Boratyński Z, Brito JC, Mappes T. The origin of two cryptic species of African desert jerboas (Dipodidae: Jaculus). Biol J Linn Soc. 2012;105:435–45.
18. Ranck GL. The Rodents of Libya: Taxonomy, Ecology, and Zoogeographical Relationships. Bull United States Natl Museum. 1968.
19. Wilson, D. E., & Reeder, D. M. (Eds.). Mammal species of the world: a taxonomic and geographic reference. JHU Press; 2005:2001.
20. Shenbrot G, Feldstein T, Meiri S. Are cryptic species of the Lesser Egyptian Jerboa, Jaculus jaculus (Rodentia, Dipodidae), really cryptic? Re-evaluation of their taxonomic status with new data from Israel and Sinai. J Zool Syst Evol Res. 2016; 54:148:59.
21. Ben Faleh A, Granjon L, Tatard C, Boratyński Z, Cosson JF, Said K. Phylogeography of two cryptic species of African desert jerboas (Dipodidae: Jaculus). Biol J Linn Soc. 2012;107:27–38.
22. Boratyński Z, Brito JC, Campos JC, Karala M, Mappes T. Large spatial scale of the phenotype-environment color matching in two cryptic species of African desert jerboas (Dipodidae: Jaculus). PLoS One. 2014; 9:e94342.
23. Nielsen R, Wakeley J. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics. 2001;158:885–96.
24. Bannikova AA, Lebedev VS, Lissovsky AA, Matrosova V, Abramson NI, Obolenskaya EV, et al. Molecular phylogeny and evolution of the Asian lineage of vole genus Microtus (Rodentia: Arvicolinae) inferred from mitochondrial cytochrome b sequence. Biol J Linn Soc. 2009;99:595–613.
25. Haynes S, Jaarola M, Searle JB. Phylogeography of the common vole (Microtus arvalis) with particular emphasis on the colonization of the Orkney archipelago. Mol Ecol. 2003;12:951–6.
26. Jaarola M, Martinkova N, Gunduz I, Brunhoff C, Zima J, Nadachowski A, et al. Molecular phylogeny of the speciose vole genus Microtus (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol. 2004;33:647–663.
27. Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB. Cryptic speciation in the field vole: A multilocus approach confirms three highly divergent lineages in Eurasia. Mol Ecol.

2012;21:6015–32.

28. Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry RW, Huchon D. Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. BMC Evol Biol. 2009;9:71.
29. Edrey YH, Casper D, Huchon D, Mele J, Gelfond J a., Kristan DM, et al. Sustained high levels of neuregulin–1 in the longest-lived rodents; A key determinant of rodent longevity. Aging Cell. 2012;11:213–22.
30. Lebedev VS, Bannikova AA, Pagès M, Pisano J, Michaux JR, Shenbrot GI. Molecular phylogeny and systematics of Dipodoidea: A test of morphology-based hypotheses. Zool Scr. 2012;42:231–49.
31. Pisano J, Condamine FL, Lebedev V, Bannikova A, Quéré J-P, Shenbrot GI, et al. Out of Himalaya: the impact of past Asian environmental changes on the evolutionary and biogeographical history of Dipodoidea (Rodentia). J Biogeogr. 2015; 42:856–870.
32. Montgelard C, Forty E, Arnal V, Matthee CA. Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. BMC Evol Biol. 2008;8:321.
33. Turner LM, Hoekstra HE. Adaptive evolution of fertilization proteins within a genus: Variation in ZP2 and ZP3 in deer mice (Peromyscus). Mol Biol Evol. 2006;23:1656–69.
34. Steiner CC, Weber JN, Hoekstra HE. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol. 2007;5(9):1880–9.
35. Heled J, Drummond AJ. Bayesian Inference of Species Trees from Multilocus Data. Mol Biol Evol. 2010;27:570–80.
36. Bradley RD, Baker RJ. A test of the genetic species concept: cytochrome-b sequences and mammals. J Mammal. 2001;82:960–73.
37. Baker RJ, Bradley RD. Speciation in mammals and the genetic species concept. J Mammal. 2006;87(4):643–62.
38. Nicolas V, Granjon L, Duplantier JM, Cruaud C, Dobigny G. Phylogeography of spiny mice (genus Acomys, Rodentia: Muridae) from the south-western margin of the sahara with taxonomic implications. Biol J Linn Soc. 2009;98(October 2015):29–46.
39. Besnard G, Rubio De Casas R, Vargas P. Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (Olea europaea). J Biogeogr. 2007;34:736–52.
40. Brouat C, Tatard C, Bã¢ K, Cosson J-F, Dobigny G, Fichet-Calvet E, et al. Phylogeography of the Guinea multimammate mouse (Mastomys erythroleucus): a case study for Sahelian species in West Africa. J Biogeogr. 2009;36:2237–50.
41. Hewitt GM. Some genetic consequences of ice ages, and their role, in divergence and speciation. Biol J Linn Soc. 1996;58:247–76.
42. Guiller A, Coutellec-Vreto MA, Madec L, Deunff J. Evolutionary history of the land snail Helix aspersa in the Western Mediterranean: preliminary results inferred from mitochondrial DNA sequences. Mol

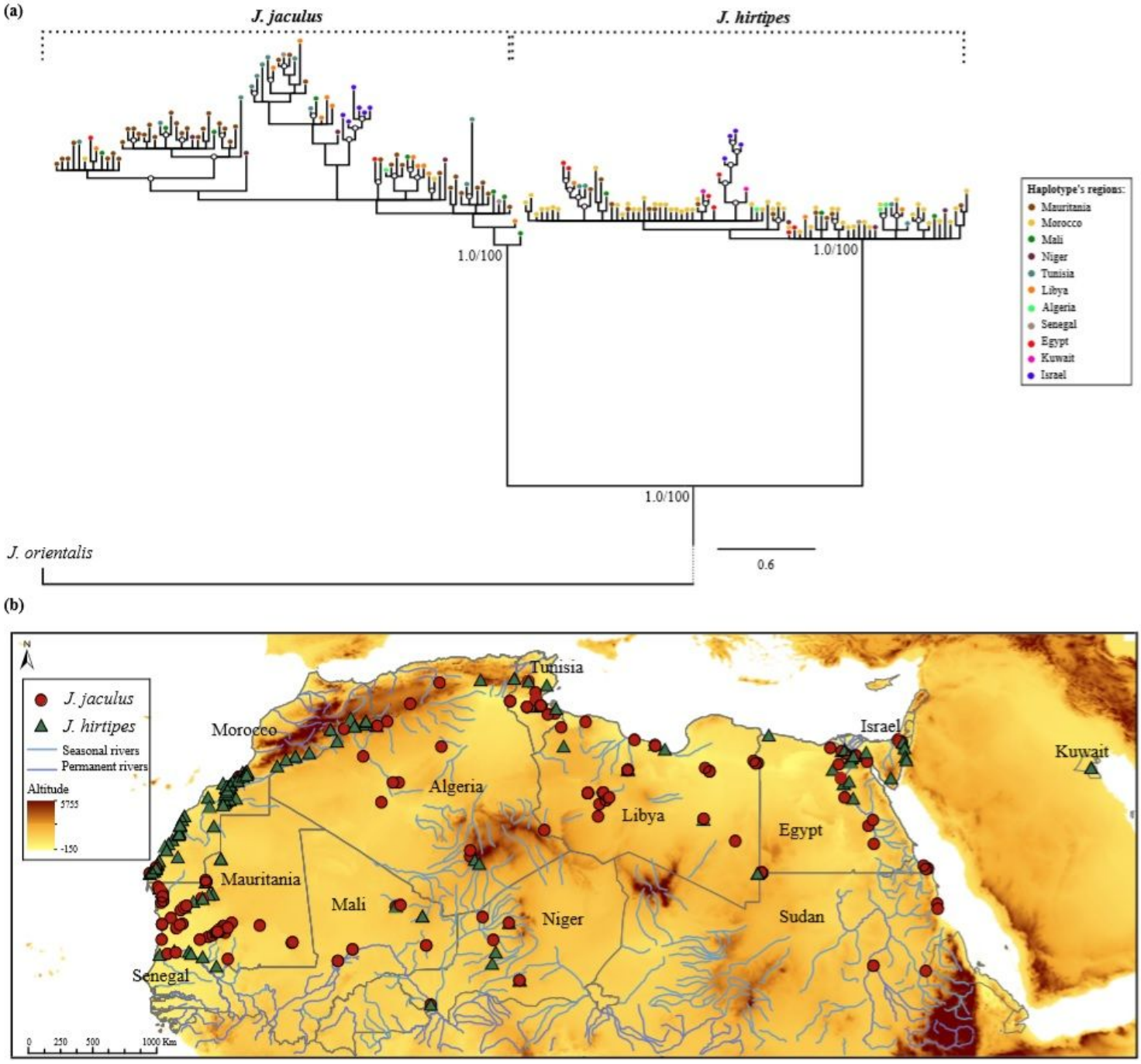
Ecol. 2001;10:81–7.

43. Cosson J-F, Hutterer R, Libois R, Sarà M, Taberlet P, Vogel P. Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, Crocidura russula (Mammalia: Soricidae). Mol Ecol. 2005;14:1151–62.
44. Guiller A, Madec L. Historical biogeography of the land snail Cornu aspersum: a new scenario inferred from haplotype distribution in the Western Mediterranean basin. BMC Evol Biol. 2010;10:18.
45. Carneiro M, Ferrand N, Nachman MW. Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (Oryctolagus cuniculus). Genetics. 2009;181:593–606.
46. Hey J. The Divergence of Chimpanzee Species and Subspecies as Revealed in Multipopulation Isolation-with-Migration Analyses. Mol Biol Evol. 2009;27:921–33.
47. Millicent E, Thoday JM. Gene flow and divergence under disruptive selection. Science. 1960;131:1311–1312.
48. Smith JM. Sympatric speciation. Am Nat. 1966;100:637–650.
49. Kopp M, Servedio MR, Mendelson TC, Safran RJ, Rodríguez RL, Hauber ME, et al. Mechanisms of Assortative Mating in Speciation with Gene Flow: Connecting Theory and Empirical Research. Am Nat. 2018; 19:1–20.
50. Boratyński Z, Brito JC, Campos JC, Cunha JL, Granjon L, Mappes T, et al. Repeated evolution of camouflage in speciose desert rodents. Sci Rep. 2017;7(1):1–10.
51. Rolán-Alvarez E. Sympatric speciation as a by-product of ecological adaptation in the Galician Littorina saxatilis hybrid zone. J Molluscan Stud. 2007;73(1):1–10.
52. Moutinho F, Qninba A, Harrington A, Forbes K. Winter breeding of the Lesser Egyptian Jerboa Jaculus jaculus (Linnaeus, 1758) in Southern Morocco. 2015;24–7.
53. Barros MI, Brito JC, Campos JC, Mappes T, Qninba A, Sousa FV, et al. The effect of rainfall on population dynamics in Sahara-Sahel rodents. Mammal Res. 2018;63(4):485–92.
54. Kosman C, Breu H, Chappell G, Kumar S, Iasnopolski B, Kshirsargar B, et al. Design and Performance Overview of SeqScape TM Software for Comparative Sequencing Analysis and Mutation Detection. Hum Genet. 2001;8–8.
55. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673–80.
56. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947–8.
57. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic Acids Symp Ser. 1999;41:95–98.
58. Flot JF, Tillier A, Samadi S, Tillier S. Phase determination from direct sequencing of length-variable DNA regions. Mol Ecol Notes. 2006;6:627–630.

59. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001;68:978–89.
60. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet. 2005;76:449–62.
61. Flot JF. Seqphase: A web tool for interconverting phase input/output files and fasta sequence alignments. Mol Ecol Resour. 2010;10:162–6.
62. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.
63. Akaike H. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika. 1973; 60:255–265.
64. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Vol. 9, Nature Methods. 2012. 9:772.
65. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.
66. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–42.
67. Drummond AJ, Rambaut A. Bayesian evolutionary analysis by sampling trees. Bayesian Evol Anal with BEAST. 2007;7:214.
68. Rambaut A. FigTree v1.3.1. 2009. Accessed Novemb 29, 2012.
69. Clement M, Posada D, Crandall KA. TCS: A computer program to estimate gene genealogies. Mol Ecol. 2000;9:1657–9.
70. Santos AM, Cabezas MP, Tavares AI, Xavier R, Branco M. TcsBU: A tool to extend TCS network layout and visualization. Bioinformatics. 2015;32:627–8.
71. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. olecular Biology and Evolution. 2006. 23:254–67.
72. Woerner AE, Cox MP, Hammer MF. Recombination-filtered genomic datasets by information maximization. Bioinformatics. 2007;23:1851–3.
73. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol. 2014;10.
74. Triant DA, DeWoody JA. Accelerated molecular evolution in Microtus (Rodentia) as assessed via complete mitochondrial genome sequences. Genetica. 2006;128:95–108.
75. Nabholz B, Glémin S, Galtier N. Strong variations of mitochondrial mutation rate across mammals - The longevity hypothesis. Mol Biol Evol. 2008;25:120–30.
76. Hey J. Isolation with migration models for more than two populations. Mol Biol Evol. 2010;27:905–20.

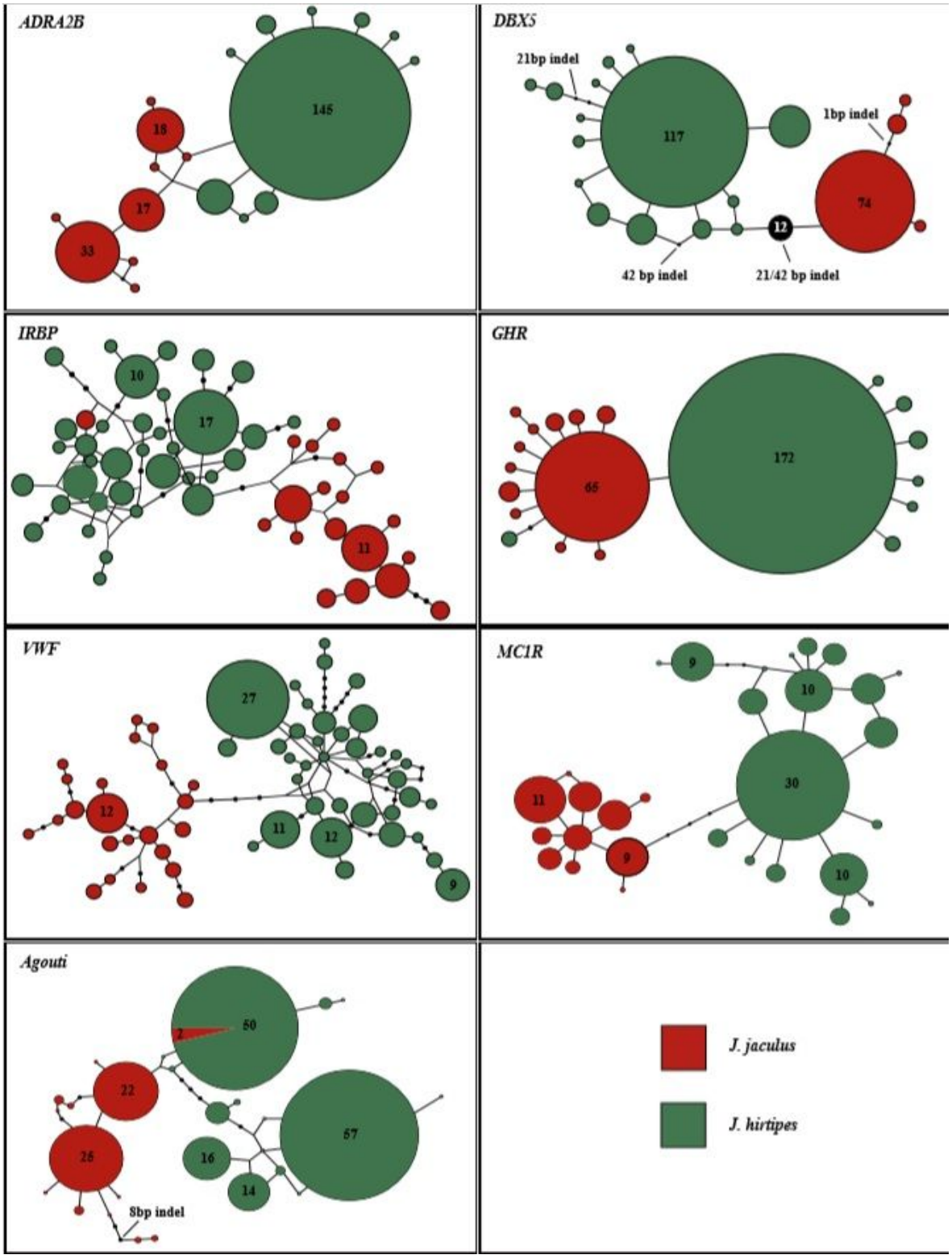
77. Hey J, Nielsen R. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics. 2004;167:747–60.
78. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28:2731–9.
79. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585–95.
80. Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 1997;147:915–25.
81. Ramos-Onsins SE, Rozas J. Statistical properties of new neutrality tests against population growth. Mol Biol Evol. 2002;19:2092–100.
82. Heled J, Drummond AJ. Bayesian inference of population size history from multiple loci. BMC Evol Biol. 2008;8:289.
83. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes. 2004;4:535–8.
84. Goudet J. FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. J Hered. 1995;86:485–6.
85. Peakall R, Smouse PE. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 2006;6:288–95.
86. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.
87. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4:359–61.
88. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol. 2005;14:2611–20.
89. Warren DL, Glor RE, Turelli M. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. Evolution. 2008;62:2868–83.
90. Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, et al. Measuring ecological niche overlap from occurrence and spatial environmental data. Glob Ecol Biogeogr. 2012;21:481–97.
91. Vale CG, Tarroso P, Brito JC. Predicting species distribution at range margins: Testing the effects of study area extent, resolution and threshold selection in the Sahara-Sahel transition zone. Divers Distrib. 2014;20:20–33.

# Figures



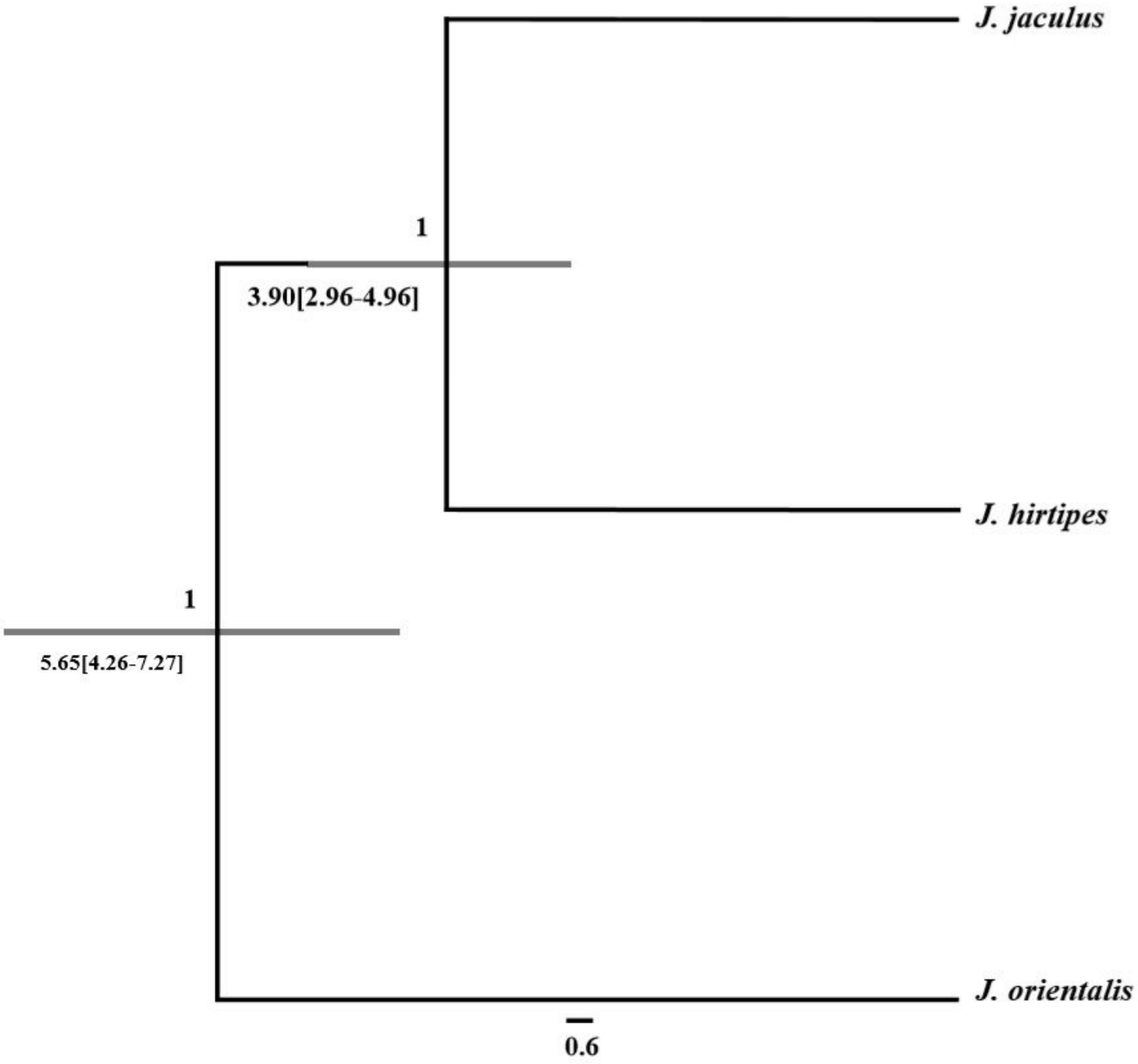

## Figure 1

Map of the distribution of Jaculus individuals across North Africa and their phylogenetic relationship.



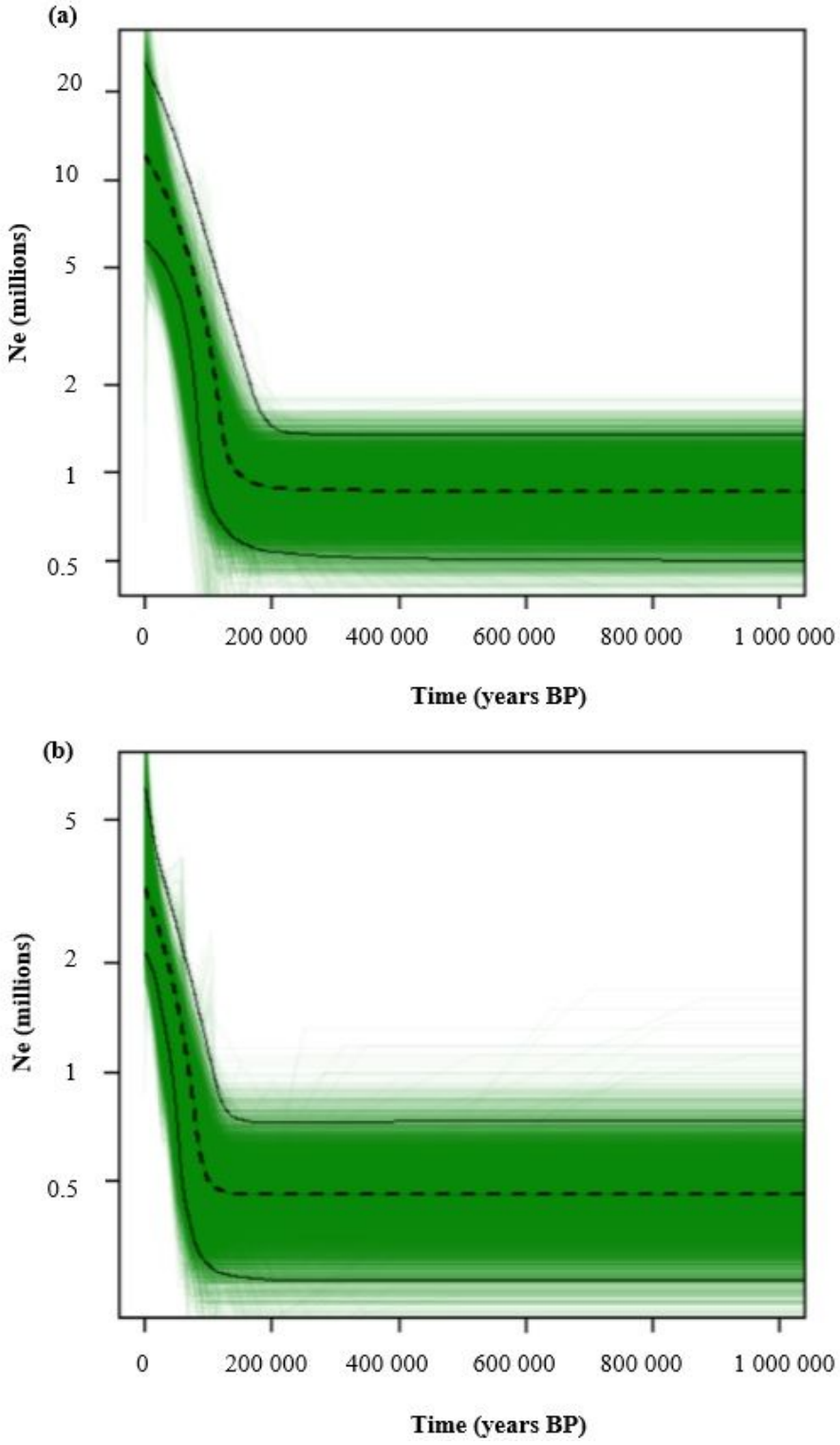

## Figure 2

Statistical parsimony haplotype networks for the single copy nuclear DNA markers.




## Figure 3

Species tree inferred with *BEAST.




## Figure 4

Extended Bayesian Skyline plots (EBSP) of the effective population size through time obtained from the three MCMC simulations

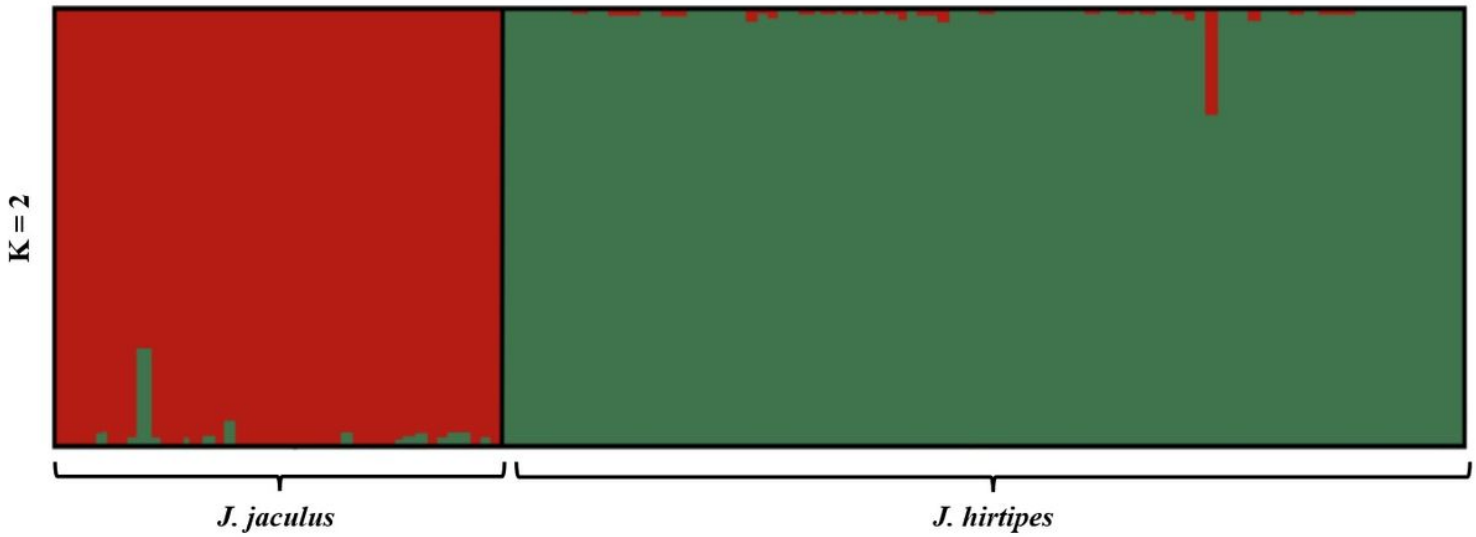



**Figure 5**

Structure bar plot of Bayesian assignments of individual to the respective cluster (K=2).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.pdf