Colonization time of Arabia by the White-tailed Mongoose
*Ichneumia albicauda* as inferred from mitochondrial DNA sequences

(Mammalia: Herpestidae)

Carlos A. Fernandes

Abstract. Given the presence of several species of terrestrial wildlife with an Arabian population separated from their main distribution range, the Arabian Peninsula can be seen as a biogeographic isolate. This arises from its single links with Africa via the Sinai land bridge and the ecological barrier of the Rub’ Al Khali desert to dispersals. In the context of Afro-Arabian biogeography, sub-Saharan and Maghreb populations, for which access to the Sinai land bridge is limited respectively by the eastern Sahara and Libyan deserts, are particularly isolated from their Arabian counterparts. Genetic markers have proved useful in studying the evolutionary history of the Arabian populations. A study of mitochondrial DNA sequence data for the White-tailed Mongoose *Ichneumia albicauda* (an Afrotropical mammal) suggested a single colonization period of the Arabian Peninsula $\approx 32,500$ years ago, making this species a relatively long-term resident and natural colonist of Arabia. Given that colonization of the Arabian Peninsula is estimated to have occurred at a time in which the Red Sea was neither particularly narrow nor shallow, and during a prolonged wet period, the scenario of invasion of Arabia via the Sinai land bridge is perhaps more likely. However, the hypothesis that the Arabian founders derived from a successful landing of a sweepstake-rafting event across the southern Red Sea, difficult to validate or falsify as it is, cannot be categorically rejected. Importation and release of individuals from Africa in the peninsula is thus strongly advised against since it could obliterate a unique evolutionary history.

Key words. *Ichneumia albicauda*, White-tailed Mongoose, Afro-Arabian zoogeography, phyleogeography, mitochondrial DNA, colonization, Arabia

Introduction

Most of the terrestrial mammalian species present in the Arabian Peninsula are, biogeographically, Afrotropical or Saharo-Sindian (DELANY 1989). The former elements tend to be seen as the result of direct dispersals from Eritrea-Somalia into south-western Arabia (DELANY 1989) but, given the evidence for colonization through the Sinai by Afrotropical species from other faunal groups (ROBINSON & MATTHEE 1999, POOK et al. 2009), our view on the route(s) and time(s) of their invasions of Arabia necessarily remains hazy. To my knowledge, only for the Hamadryas Baboon *Papio hamadryas* possible scenarios have been detailed based on phylogeographic genetic data (WILDMAN et al. 2004, WINNEY et al. 2004, FERNANDES 2009).

The White-tailed Mongoose *Ichneumia albicauda* Cuvier, 1829 is a mammalian carnivore of the family Herpestidae widely distributed across sub-Saharan Africa, in a broad range of habitats from woodland to semi-deserts, and also found in southern Arabia (WILSON & MIT-
TERMEIER 2009). A recent study (DEHGhani et al. 2008), using phylogenetic analyses of partial sequences of the mitochondrial DNA (mtDNA) control region, examined the phylogeography of the species and contrasted it with its current morphology-based subspecies taxonomy.

Using this mtDNA dataset and, in the absence of specimens from the eastern Sahel adjacent to the Red Sea in the study by DEHGhani et al. (2008), homologous sequences from White-tailed Mongooses sampled in Eritrea, I investigated the colonization time(s) of Arabia by the species. Depending upon which time intervals these estimates fall within, and their relations with palaeo-environmental evidence, it might be possible to deduce the most feasible scenario for the colonization(s) of Arabia by the White-tailed Mongoose. For instance, if they correspond to periods of a remarkably shallow and narrow southern Red Sea, then direct dispersal from the Horn of Africa into south-western Arabia is a potential explanation (BAILEY 2009), probably involving sweepstake rafting (Fernandes et al. 2006). Conversely, if they match with phases in which the Red Sea was significantly wide and deep and the eastern Sahara could have been substantially reduced, then episodic range continuity or migrations around the Red Sea might be a more likely hypothesis. Finally, if the inferred colonization time is very recent, the possibility of anthropogenic introductions in historic times should not be discounted (Fernandes et al. 2006).

A more general aim of this paper is to contribute much-needed phylogeographic data to the subject of Afro-Arabian biogeography and to increase our knowledge on the evolutionary history and conservation importance of the Arabian populations of wild mammals.

Material and methods

The mtDNA sequences of White-tailed Mongoose obtained by DEHGhani et al. (2008) were downloaded from GenBank (accession numbers EU871433-EU871479). See their Table 1 for voucher information and geographic origin of the used samples.

Total genomic DNA was isolated from muscle tissue samples of two specimens from Eritrea, using the NucleoSpin Tissue Kit (Macherey-Nagel) and following the manufacturer’s protocol. Sequences homologous to EU871433-EU871479 were produced by polymerase chain reaction (PCR) and sequencing with shortened and modified versions of the primers L15926 (Kocher et al. 1989) and H16498 (Shields & Kocher 1991), respectively 5’-CCT TGG TCT TGT AAA CC-3’ and 5’-CCT GAA GTA GGA ACC AGA TG-3’. The PCR reactions were carried out in volumes of 15 µl and run in a thermal cycler using the following protocol: 94°C for 3 min, 10 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 45 s, 25 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 45 s, and a final extension at 72°C for 7 min. PCR products were purified with an Exo-SAP protocol (Rieder et al. 1998) and sequenced at Macrogen Inc.

All sequences were edited, assembled, and aligned using SEQUEENCHER 4.7 (Gene Codes). Alignments were corrected by eye and indels were adjusted to minimize both the total number of substitutions and of transversions. The alignment of 47 sequences of 364 nucleotides in length was analyzed with COLLAPSE 1.2 (Posada & Buckley 2004) to collapse identical individual sequences into representative haplotypes for phylogenetic analysis.

The program Treefinder (Jobb et al. 2004, version October 2008) was used to perform a maximum-likelihood (ML) phylogenetic analysis. The best-fit evolutionary model was selected by the Akaike Information Criterion with correction for small sample size (AICc) (Posada & Buckley 2004), and used for phylogeny reconstruction. Non-parametric bootstrapping (1,000 replicates) was employed to assess the robustness of nodes. Bayesian phylogenetic inference was also carried out, using the software MrBayes 3.1.2 (Ronquist & Huelserbeck 2003). One cold and
Table 1. Haplotype codes, taxa, geographic origins, and GenBank accession numbers. *I.* = *Ichneumia*; *C.* = *Cynictis.*

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Species</th>
<th>Country</th>
<th>GenBank</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Eri</td>
<td><em>I. albicauda</em></td>
<td>Eritrea</td>
<td>HM037041</td>
</tr>
<tr>
<td>H2 Eri</td>
<td><em>I. albicauda</em></td>
<td>Eritrea</td>
<td>HM037042</td>
</tr>
<tr>
<td>H3 Ken</td>
<td><em>I. albicauda</em></td>
<td>Kenya</td>
<td>EU871433</td>
</tr>
<tr>
<td>H4 Eth</td>
<td><em>I. albicauda</em></td>
<td>Ethiopia</td>
<td>EU871434</td>
</tr>
<tr>
<td>H5 Eth</td>
<td><em>I. albicauda</em></td>
<td>Ethiopia</td>
<td>EU871435</td>
</tr>
<tr>
<td>H6 Eth</td>
<td><em>I. albicauda</em></td>
<td>Ethiopia</td>
<td>EU871436</td>
</tr>
<tr>
<td>H7 Eth</td>
<td><em>I. albicauda</em></td>
<td>Ethiopia</td>
<td>EU871437/38</td>
</tr>
<tr>
<td>H8 YemUAE</td>
<td><em>I. albicauda</em></td>
<td>Yemen, U.A.E.</td>
<td>EU871439/40/41</td>
</tr>
<tr>
<td>H9 Yem</td>
<td><em>I. albicauda</em></td>
<td>Yemen</td>
<td>EU871442</td>
</tr>
<tr>
<td>H10 Yem</td>
<td><em>I. albicauda</em></td>
<td>Yemen</td>
<td>EU871443</td>
</tr>
<tr>
<td>H11 Yem</td>
<td><em>I. albicauda</em></td>
<td>Yemen</td>
<td>EU871444</td>
</tr>
<tr>
<td>H12 YemKSA</td>
<td><em>I. albicauda</em></td>
<td>Yemen, Saudi Arabia</td>
<td>EU871445/46/47/48</td>
</tr>
<tr>
<td>H13 CAR</td>
<td><em>I. albicauda</em></td>
<td>Central African Rep.</td>
<td>EU871449</td>
</tr>
<tr>
<td>H14 Uga</td>
<td><em>I. albicauda</em></td>
<td>Uganda</td>
<td>EU871450</td>
</tr>
<tr>
<td>H15 Eth</td>
<td><em>I. albicauda</em></td>
<td>Ethiopia</td>
<td>EU871451/52</td>
</tr>
<tr>
<td>H16 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871453/54</td>
</tr>
<tr>
<td>H17 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871455</td>
</tr>
<tr>
<td>H18 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871456</td>
</tr>
<tr>
<td>H19 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871457</td>
</tr>
<tr>
<td>H20 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871458</td>
</tr>
<tr>
<td>H21 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871459/60</td>
</tr>
<tr>
<td>H22 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871461/62</td>
</tr>
<tr>
<td>H23 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871463/64/65/66/67/68/69/70/71</td>
</tr>
<tr>
<td>H24 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871472</td>
</tr>
<tr>
<td>H25 Gui</td>
<td><em>I. albicauda</em></td>
<td>Guinea</td>
<td>EU871474/75/76</td>
</tr>
<tr>
<td>H26 SA</td>
<td><em>I. albicauda</em></td>
<td>South Africa</td>
<td>EU871477</td>
</tr>
<tr>
<td>H27 SA</td>
<td><em>I. albicauda</em></td>
<td>South Africa</td>
<td>EU871478</td>
</tr>
</tbody>
</table>

*Cynictis* *C. penicillata* South Africa U28429

Three heated Markov chains, applying default heating values (*t*=0.2), were run for 1,000,000 generations. The first 10,000 generations were discarded as burn-in phase and, thereafter, chains were sampled every 50 generations to form the posterior probability distribution of parameters and topologies. A sequence of Yellow Mongoose *Cynictis penicillata* (accession number U28429; Van Vuuren & Robinson 1997) was used as outgroup for phylogenetic analysis. From both the ML bootstrap trees and the Bayesian posterior trees, 50% majority rule consensus trees were computed with TREEFINDER and visualized and edited with FIGTREE (Rambaut 2009; version 1.3.1).

As an approximation to the colonization time of Arabia by the White-tailed Mongoose, the divergence time between the Arabian lineages and their African counterparts in Clade I of the phylogenetic tree shown in Fig. 1 was estimated using the program IM (version December 2009), which implements a coalescent model that can be applied to genetic data drawn from a pair of closely related populations or species (Nielsen & Wakeley 2001, Hey & Nielsen 2004). One of the appealing attributes of IM is that it allows implementation of a version of the basic model in
which the descendant populations, of an ancestral population that has split into two, are free to undergo changes in size. The model has then seven demographic parameters: the splitting time ($t$), the effective population sizes of the ancestral and daughter populations ($N_A$, $N_1$, and $N_2$), the migration rates between the daughter populations ($m_1$ and $m_2$), and the fraction of $N_A$ that forms $N_1$ ($s$). The program performs a Markov chain Monte Carlo (MCMC) simulation with the Metropolis-Hastings algorithm, in which random samples from the posterior probability distribution are generated. Over the course of a sufficiently long run, the distribution of recorded values is expected to approximate the posterior probability density of those values. To evaluate if the program has run for long enough, and the sample has converged to the true posterior probability distribution, a measure of the independence of the recorded values over the course of the run is estimated. This measure, Effective Sample Size (ESS), should have values greater than 200 for any estimated parameter. Results can be trusted if three or more runs show high ESS values and generate similar stationary distributions.

After preliminary runs, to optimize MCMC settings and finding the range of prior distributions that include all or most of the range over which the posterior density is non-trivial, four replicate simulations were run. Estimates were generated under a Hasegawa-Kishino-Yano model (HKY; HASEGAWA et al. 1985), selected by TREEFINDER as best fitting the dataset of 16 I. albicauda sequences with Clade I haplotypes. Simulations used 10 Metropolis-coupled chains, with 45 chain swap attempts per step, and were run for 10 million steps, the first 100,000 steps discarded as “burn-in”. Genealogies were updated ten times per step. Priors for the fraction of the ancestral population that did not colonize Arabia, the parameter $s$ ($0<s<1$) were bounded between 0.5 (assumption of a symmetric split) and 1 (assumption that Arabia was invaded by a minority of the ancestral population). Although the IM model includes the isolation model (WAKELEY & HEY 1997) as a simplified version, by prescribing migration rates of zero, that may be appropriate for studying population divergence under allopatry (WON et al. 2003), I preferred not to make a priori assumptions about the isolation of the Arabian and African populations. I also used IMa (HEY & NIELSEN 2007), which implements the same model as IM but estimates the joint posterior probability density of the model parameters. Parameter ranges for uniform priors were empirically determined in a series of initial runs, and four final runs were conducted. Estimates were generated under a HKY model (HASEGAWA et al. 1985). Simulations used 10 chains, with 45 chain swap attempts per step, and were run for 20 million steps, the first 200,000 steps discarded as “burn-in”. Genealogies were sampled every 100 steps. The four runs resulted in similar parameter estimates and distributions and ESSs exceeding at least 20,000 for each parameter. Saved genealogies from the four runs were combined and the joint marginal distribution of $t$ was estimated from an evenly spaced sample of 80,000 trees.

To convert coalescent times to years before present, estimates of the substitution rate ($\mu$) for the DNA region under study and of the generation time of the species are required. To my knowledge, no substitution rate for the left domain of the mtDNA control region has been estimated in the White-tailed Mongoose or in other herpestid. Given the availability of homologous sequences from the Yellow Mongoose (VAN VUUREN & ROBINSON 1997), and an inferred divergence of the two lineages at 15.8 Ma (PATOU et al. 2009), the substitution rate could in principle be estimated using the formula $d_{xy} = 2\mu T$, where $T$ is the time to the most recent common ancestor and $d_{xy}$ is the genetic distance between species. However, substitution saturation in the fast-evolving left domain of the control region among long-diverged species is known to be high, contributing to an underestimation of the substitution rate. Indeed, using MEGA 4.0 ( TAMURA et al. 2007), which allows for rate heterogeneity among lineages, and the Tamura-Nei evolutionary model ( TAMURA & NEI 1993) with $\Gamma = 0.322$, $d_{xy}$ was estimated as 0.139±0.039, which translates into a substitution rate of $4.4 \times 10^{-9}$ substitutions site$^{-1}$ year$^{-1}$. High substitution saturation was confirmed with tests carried out in DAMBE (XIA & XIE 2001; version 5.2.5). Substitution rates
for the left domain of the mtDNA control region around 1.5 x 10⁻⁷ substitutions site⁻¹ year⁻¹ have been estimated in carnivores (Taberlet & Bouvet 1994, Savolainen et al. 2002, Leonard et al. 2005) and in other mammals (Pesole et al. 1992, Stoneking et al. 1992), and used in studies of species from different carnivore families (McFadden et al. 2008, Aubry et al. 2009). Hence, a substitution rate of 1.5 x 10⁻⁷ substitutions site⁻¹ year⁻¹ was used in this study. A generation time of two years was assumed, based on studies of several small to medium-sized carnivores of different families (Riley et al. 2006, Tchaicka et al. 2007, McFadden et al. 2008).

Divergence time was also estimated using equation $D_A = 2uT$ in which $u$ is the average substitution rate per nucleotide (i.e. the substitution rate per site per year times the length of the analyzed DNA fragment, in this case 359 nucleotides after removing sites with indels), $T$ is the divergence time, and $D_A$ is the net number of nucleotide differences between populations (Nei & Li 1979). The latter was calculated, using the Tamura-Nei model with $\Gamma = 0.1$, with the computer program ARLEQUIN 3.11 (Excoffier et al. 2005). The same software was used to construct the mismatch distribution, the distribution of the observed number of differences between pairs of haplotypes (Rogers & Harpending 1992, Schneider & Excoffier 1999), of the Arabian population to assess its past demography. The distribution is usually unimodal in populations having passed through a recent demographic expansion.

Finally, I estimated divergence time following the coalescent method described by Gaggiotti & Excoffier (2000), also implemented in ARLEQUIN. This method aims to remove the effect of bottlenecks and unequal sizes of the derived populations, which can lead to the overestimation of divergence times from genetic distances.

BEAST (Bayesian Evolutionary Analysis by Sampling Trees) (Drummond & Rambaut 2007), a program for the Bayesian analysis of genetic sequences related by an evolutionary tree, has also Metropolis-Hastings MCMC as its core algorithm for coalescent-based estimation of phylogenetic and population genetics parameters. BEAST (version 1.5.3) was used here to infer the time to the most recent common ancestor (TMRCA) of the Arabian lineages in Clade I. Since BEAST does not correct for ancestral polymorphism, it estimates the TMRCA of lineages/clades and not of the populations they were sampled from. Still, BEAST estimates are useful to gauge the upper bound of population divergence time (Griswold & Baker 2002, Leschen et al. 2008). Estimates were generated under a HKY model with $\Gamma = 0.18$, selected by TREEFINDER as best fitting the dataset of 47 I. albicauda sequences. The starting tree was a random tree reconstructed by the program under the coalescent process. Other priors were left with their default prior distribution. The MCMC operators (the proposal distribution) were auto-optimized by the program during preliminary runs, and non-trivial runs followed its recommendations. For each simulation, the MCMC was run for 30 million steps and sampled every 500 steps, following a discarded burn-in of 150,000 steps. Convergence to the stationary distribution of individual and combined runs was confirmed using the application TRACER 1.5 (Rambaut & Drummond 2007). Estimation of the TMRCA, and associated 95% high posterior density (HPD) intervals, was obtained from the posterior distribution of parameters approximated by the combination of four independent MCMC analyses.

**Results**

A total of 28 haplotypes (Table 1) were found and their phylogenetic relationship, based on a best-fit model of the type HKY with $\Gamma = 0.304$ as selected by TREEFINDER, is presented in Fig. 1. The majority rule consensus tree obtained from the Bayesian analysis (not shown) was nearly identical to the one presented. Also, unsurprisingly, the recovered topology is very similar to the one obtained by Deighani et al. (2008). In particular, concerning East
Africa and Arabia, the same two clades were recovered: one (Clade II in Fig. 1), with the same level of bootstrap support, including the single specimen from Kenya and four out of the five Ethiopian haplotypes, and the other (Clade I in Fig. 1), with a lowered level of bootstrap support due to the addition of the two novel haplotypes from Eritrea, including the single specimens from Uganda and Central African Republic, one Ethiopian haplotype, and all the Arabian haplotypes. The alternative topologies suggested by the 70% and 90% majority-rule consensus trees were evaluated using the likelihood ratio test, and the two alternative topologies (null hypotheses) were rejected ($P = 0.05$). All the Arabian haplotypes were private (i.e. exclusive) to Arabia.

Table 2 presents the results of the IM simulations. The peaks of the posterior distributions were taken as maximum-likelihood estimates of the model parameters. Credibility intervals were recorded as the 90% HPD interval, which represents the shortest span that includes 90% of the probability density of the parameter. The IMa analyses suggest that the Arabian population and the African lineages of Clade I diverged $\approx 32,784$ years ago (90% HPD = 13,736 – ?). The upper bound for the credibility interval is not given because it critically depends on the assumed prior for the maximum value of $t$ when the curve slowly decreases to zero after the mode of $t$ (JOHNSON et al. 2007, GUILLAUMET et al. 2008).
Table 2. IM estimates of the colonization time of Arabia by the White-tailed Mongoose. Time estimates and confidence intervals are given in years, assuming a generation time of two years and a substitution rate for the left domain of the mtDNA control region of $1.5 \times 10^{-7}$ substitutions site$^{-1}$ year$^{-1}$. Effective Sample Size (ESS) for the parameter $t$ is given for each simulation. Also given are the ML estimates of the other demographic parameters in the IM model. $^a$ Estimates of divergence time between Arabian and African lineages in Clade I, assumed to approximate the colonization time of Arabia. $^b$ Confidence intervals are 90% highest posterior density intervals.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Colonization Time Estimate$^a$</th>
<th>Confidence Interval$^b$</th>
<th>ESS</th>
<th>Other Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31,603</td>
<td>8,141-87,115</td>
<td>119,771</td>
<td>$t_A=12$; $t_1=99$; $t_2=14$; $m_1=0.0005$; $m_2=0.0005$</td>
</tr>
<tr>
<td>2</td>
<td>31,731</td>
<td>8,141-87,885</td>
<td>122,539</td>
<td>$t_A=11$; $t_1=118$; $t_2=15$; $m_1=0.0005$; $m_2=0.0005$</td>
</tr>
<tr>
<td>3</td>
<td>31,859</td>
<td>8,269-87,115</td>
<td>46,879</td>
<td>$t_A=12$; $t_1=102$; $t_2=16$; $m_1=0.0005$; $m_2=0.0005$</td>
</tr>
<tr>
<td>4</td>
<td>32,732</td>
<td>8,141-85,962</td>
<td>115,251</td>
<td>$t_A=12$; $t_1=102$; $t_2=16$; $m_1=0.0005$; $m_2=0.0005$</td>
</tr>
</tbody>
</table>

The population pairwise genetic distance $D_A$ between the Arabian population and the African lineages of Clade I was estimated as 4.977 ($P<0.05$), which translates into a population divergence 46,212 years ago. The Arabian population exhibited evidence of expansion, not showing any significant deviation from a unimodal mismatch distribution ($P>0.05$), dated at 63,510 years ago ($\tau = 3.42$), with 95% confidence interval 3,528-119,963 years before present. The raggedness index of the observed distribution (Harpending 1994) was 0.085 ($P=0.05$), also suggesting a recent expansion of the Arabian population.

Following Gaggiotti & Excoffier (2000), divergence between the Arabian population and the African lineages of Clade I was estimated at $\approx 9,192$ years ago ($\tau = 0.99$), with the ancestral female effective population size estimated as 97,400 ($\theta_0 = 10.49$). This method also estimates the relative sizes ($k$ and $1-k$) of the derived populations, and these were equal to 0.94 for the African lineages of Clade I and 0.06 for the Arabian population. This is in agreement with the maximum-likelihood estimate of the parameter $s$ ($= 0.9995$) in the simulations using IM.

The analyses using BEAST allowed the investigation of two different models of rate variation among branches, namely the strict clock and the uncorrelated lognormal-distributed relaxed clock. The data could be treated as compliant with a global clock rate, because the estimate of the $\text{ucld.stdev}$ parameter, as well as of the coefficient of variation parameter, in simulations using the lognormal clock was less than 0.5. Also, model comparison via the Bayes factor (Kass & Raftery 1995), the ratio of the marginal likelihoods of the two models, indicated no evidence against the null model (i.e., the one with lower marginal likelihood). Approximate marginal likelihoods for each model were calculated via importance sampling (1000 bootstrap replicates) using the harmonic mean of the sampled likelihoods with the posterior as the importance distribution (Suchard et al. 2001), as implemented in TRACER 1.5. The Bayes factor was also used to compare coalescent models of effective pop-
ulation size change over time, one assuming a constant population size and other assuming growth (exponential), with the latter being favored. Estimates were, in any case, very similar, respectively 39,734 years (95% HPD 14,535-69,354; ESS = 93,405) and 39,174 years (95% HPD 14,277-68,628; ESS = 81,460).

Discussion
The results of this paper and of Dehghani et al. (2008) suggest a single period for the colonization of Arabia by the White-tailed Mongoose, given that all the Arabian haplotypes cluster in a single clade in the respective phylogenetic reconstructions. Moreover, the results are incongruent with a scenario of introduction of the species in historic times. Assuming that the Arabian lineages share a direct common ancestry with the African ones in Clade I and that their genetic divergence was small prior to colonization, the colonization time of Arabia can be roughly approximated by their divergence time, which was estimated as one order of magnitude older than the time of Red Sea navigation and trade, which could provide an opportunity for the introduction of the species to Arabia. However, sampling in East Africa is still too incomplete to rule out totally the possibility that Clade I Arabian haplotypes will not be found in as yet unsampled populations of the region, for instance from Sudan or northern Ethiopia. So, the hypothesis of a very recent invasion cannot be categorically rejected.

Due to the stochastic nature of mutation and lineage sorting, estimated divergence times from genetic patterns must be considered only as rough approximations of the timing of historical events. Factors like the uncertainty on the substitution rate for the DNA region under study, unsampled populations, and small sample sizes, increase confidence limits on divergence time estimates.

The divergence time was estimated between the Arabian and African lineages in Clade I using different methods, with the aim of drawing more robust conclusions on the timing of the arrival of the White-tailed Mongoose to the Arabian Peninsula. The IM estimate was ≈32,000 years ago, the IMa estimate was ≈33,000 years ago, the estimate derived from DA was ≈46,000 years ago, while the method of Gaggiotti & Excoffier (2000) estimated it as ≈9,000 years ago, with a sudden demographic expansion of the Arabian population inferred from the mismatched distribution at ≈63,500 years ago. It is noteworthy that all point estimates fall within the interval 9,000-64,000 years before present, with a median at 36,500 years ago, and all estimates are within each other’s confidence intervals.

The population pairwise genetic distance DA (Nei & Li 1979) can overestimate divergence time due to the effect of a demographic bottleneck at the divergence time, leading to two derived populations of unequal size and hence unequal genetic drift (Gaggiotti & Excoffier 2000). Evidence for a bottleneck in the Arabian population at the time of its origin comes from the parameter s (= 0.9995) estimated with IM and the parameter k (= 0.94) estimated by the method of Gaggiotti & Excoffier (2000). Support for a sudden demographic expansion in the Arabian population comes from its unimodal mismatch distribution and from a small value of the raggedness index (= 0.085). Given the mtDNA diversity in the Arabian clade, it is feasible that this expansion corresponds to the one following the colonization bottleneck. However, its age is possibly overestimated; the parameter τ from the mismatch distribution can overestimate the time since expansion when it is (very) recent
(SCHNEIDER & EXCOFFIER 1999, LESSIOS et al. 2001, ALMADA et al. 2008). Similarly, the model of GAGGIOTTI & EXCOFFIER (2000) may not perform well when divergence time is recent ($\tau/\theta_0 < 0.5$) and when sizes of the derived populations are very uneven, both of which appear to be the case here.

Coalescent genealogy samplers, like the ones implemented in IM, IMa, and BEAST, are generally more powerful and robust than the above methods to infer population parameters (KUHNER 2009). Therefore, and given that the BEAST estimate for the TMRCA of the Arabian lineages also suggests that the time estimates derived from the $D_A$ and the mismatch distribution are both biased upwards, the divergence time inferred by IM and IMa ($\approx$32,500 years ago) is taken as the most accurate. This is, of course, a working hypothesis that undoubtedly requires further testing, with additional samples and genetic markers. A more numerous and wider sampling from East Africa, particularly from northern Ethiopia, eastern Sudan, and Eritrea-Djibouti, seems crucial, while the analysis of nuclear DNA markers would provide additional useful gene trees.

The colonization time inferred with IM and IMa suggests that the White-tailed Mongoose might have invaded the Arabian Peninsula during a period in which the current record for the depth and width of the southern Red Sea (ROHLING et al. 1998, SIDDALL et al. 2003, FERNANDES et al. 2006, BAILEY 2009) is hard to reconcile with any thoroughfare connecting Africa to Arabia. As an alternative, the hypothesis that the Arabian founders resulted from a successful landing of a sweepstake-rafting event cannot be discounted (FERNANDES et al. 2006). This mode of dispersal is recognized to have played a role in long-distance transmarine migrations by medium to large-sized mammals (ROGERS et al. 2000, VAN DEN BERGH et al. 2001, YODER et al. 2003, DE QUEIROZ 2005). Such a hypothesis is also congruent with the detected signal of a bottleneck at the time of the origin of the Arabian population of White-tailed Mongoose. A difficulty with this type of scenarios is that, unless alternatives are highly unlikely, it is not straightforward to validate (or falsify) them.

The fact that the time estimated by IM and IMa for the origin of the Arabian population matches with a period (35,000-24,000 years before present) in which several wet phases were synchronous in the eastern Sahara and Arabia (YAN & PETIT-MAIRE 1994, WOODS & IMES 1995, GLENNIE & SINGHVI 2002, MOEYERSONS et al. 2002, BRAY & STOKES 2004), lends weight to the hypothesis of colonization of Arabia as a result of episodic range continuity, or long-distance dispersal, around the Red Sea. Past range expansion and long-distance dispersal can also leave strong bottleneck signatures (IBRAHIM et al. 1996, AUSTRALITZ et al. 1997). DELANY (1989) has suggested an infiltration route for faunal invasions of Arabia from the southwestern Levant, along the Red Sea coastal mountains into Asir and Aden. Subsequent cooler and drier periods would have caused range contractions to a distribution similar to what we observe today, in Africa with a northern limit imposed by the Eastern Sahara.

The fossil record of Ichneumia sp. is very limited and cannot provide arguments for or against the two competing hypotheses. Specimens of White-tailed Mongoose have been only reported from East Africa (GERAADS et al. 2004) and, interestingly, no fossils are known from Southern Africa, although this region has been recently suggested to be the geographic origin of the species (DEHGHANI et al. 2008). Absence from deposits does not necessarily mean absence from the fauna, and short-term range shifts and dispersals, during interstadials or even within short climatic optima that peak interglacial periods, might be ephemeral enough to leave no fossil traces (LAHR & FOLEY 1998).
On balance, given that the colonization of Arabia by the White-tailed Mongoose is estimated to have occurred at a time in which the Red Sea was neither particularly narrow nor shallow, and coincided with a prolonged wet period on both sides of the sea, the scenario of expansion/dispersal via the Sinai land bridge is perhaps the more likely option.

To test this biogeographic hypothesis would require numerous latitudinal sampling from both Saudi Arabia-Yemen and Sudan-Somalia, and investigate if populations in the north of the range on either side of the Red Sea are genetically more similar (or not) than those on opposite sides of the Bab-el-Mandeb Strait (e.g. WINNEY et al. 2004). A possible difficulty with this approach is that the original genetic patterns on both sides of the Red Sea might have been severely altered since the time of the invasion of Arabia, due to genetic drift, lineage loss, population range shifts during glacial periods, and local extinction followed by recolonization from neighboring demes.

The conservation status of the Arabian White-tailed Mongoose is raised by the findings of this study, which suggest that the Arabian population are ≈32,500 years old and have been isolated since its foundation. Importation and release of individuals from Africa is thus strongly advised against since it could obliterate a unique evolutionary history, with the Arabian isolate possibly on the path to subspeciation.

Other than the study on the Hamadryas Baboon (WILDMAN et al. 2004, WINNEY et al. 2004, FERNANDES 2009), this is the second study of Afrotropical terrestrial mammals for which dates and routes for their arrival to the Arabian Peninsula have been proposed based on phylogeographic genetic data. It is clear that we need more genetic surveys of Afro-Arabian species to allow informative comparisons of phylogeographic patterns and colonization time estimates. Moreover, genetic assessments of the distinctiveness and uniqueness of the Arabian isolates will likely raise their conservation profile and status.

Acknowledgements. I would like to thank Mr. Abdul Aziz MIDFA (Director General, Sharjah Environment and Protected Areas Authority, United Arab Emirates) and Mr. Paul VERCAMMEN (Breeding Centre for Endangered Arabian Wildlife, Sharjah) for the invitation to participate in the conference on Biodiversity Conservation in the Arabian Peninsula, hosted under the patronage of His Highness Dr Sheikh Sultan bin Mohamed AL QASSIMI, Ruler of Sharjah. I am grateful to Prof. Hidetoshi SHIMODAIRA (Tokyo Institute of Technology) and Prof. Chris FIELD (Dalhousie University) for advice on the testing of alternative topologies. I am deeply grateful to Prof. Jeheskel SHOSHANI for kindly providing the samples from Eritrea when he was based at the University of Asmara. Sadly, he passed away in May 2008. This paper is dedicated to his memory. I acknowledge support from the Fundação para a Ciência e a Tecnologia (FCT-MCTES, Portugal), through the postdoctoral grant BPD/20317/2004 and the Ciência 2007 researcher fellowship C2007-UL-342-CBA1.

References


**Author’s address:** Carlos A. Fernandes, Centro de Biologia Ambiental, Universidade de Lisboa, C2 2.5.41, 1749-016 Campo Grande, Lisboa, Portugal. – Email: CaFernandes@fc.ul.pt.