

# Pleistocene-Holocene Boundary in Southern Arabia From the Perspective of Human mtDNA Variation

Abdulrahim Al-Abri,<sup>1</sup> Eliška Podgorná,<sup>2</sup> Jeffrey I. Rose,<sup>3</sup> Luísa Pereira,<sup>4,5</sup> Connie J. Mulligan,<sup>6</sup> Nuno M. Silva,<sup>4</sup> Riad Bayoumi,<sup>1</sup> Pedro Soares,<sup>4</sup> and Viktor Cerný<sup>7,8\*</sup>

<sup>1</sup>Department of Biochemistry, College of Medicine & Health Sciences, Sultan Qaboos University, Muscat, Oman

<sup>2</sup>Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic

<sup>3</sup>Institute of Archaeology and Antiquity, University of Birmingham, Birmingham, UK

<sup>4</sup>Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto, Portugal

<sup>5</sup>Faculdade de Medicina da Universidade do Porto, Porto, Portugal

<sup>6</sup>Department of Anthropology, University of Florida, Gainesville, FL

<sup>7</sup>Archaeogenetics Laboratory, Institute of Archaeology of the Academy of Sciences of the Czech Republic, Czech Republic

<sup>8</sup>Institute for Advanced Study, Paris, France

**KEY WORDS** mtDNA variation; Arabian Peninsula; migrations

**ABSTRACT** It is now known that several population movements have taken place at different times throughout southern Arabian prehistory. One of the principal questions under debate is if the Early Holocene peopling of southern Arabia was mainly due to input from the Levant during the Pre-Pottery Neolithic B, to the expansion of an autochthonous population, or some combination of these demographic processes. Since previous genetic studies have not been able to include all parts of southern Arabia, we have helped fill this lacuna by collecting new population datasets from Oman (Dhofar) and Yemen (Al-Mahra and Bab el-Mandab). We identified several new haplotypes belonging to haplogroup R2 and generated its whole genome mtDNA tree with age estimates undertaken by different methods. R2, together with other con-

siderably frequent southern Arabian mtDNA haplogroups (R0a, HV1, summing up more than 20% of the South Arabian gene pool) were used to infer the past effective population size through Bayesian skyline plots. These data indicate that the southern Arabian population underwent a large expansion already some 12 ka. A founder analysis of these haplogroups shows that this expansion is largely attributed to demographic input from the Near East. These results support thus the spread of a population coming from the north, but at a significantly earlier date than presently considered by archaeologists. Our data suggest that some of the mtDNA lineages found in southern Arabia have persisted in the region since the end of the Last Ice Age. *Am J Phys Anthropol* 149:291–298, 2012. © 2012 Wiley Periodicals, Inc.

Southern Arabia is a region that is today isolated from the rest of the Arabian Peninsula by the inhospitable Rub' al Khali desert to the north and the Wahiba Sands to the east. Recent archaeological discoveries in Oman (Rose et al., 2011) and Sharjah, United Arab Emirates (Armitage et al., 2011) indicate that at least two distinct, African derived Middle Stone Age populations moved into southern Arabia during Marine Isotope Stage 5 (MIS 5). In the United Arab Emirates, the site of Jebel Faya (Armitage et al., 2011) is thought to represent an early modern human expansion from sub-Saharan Africa at the beginning of the Last Interglacial, during MIS 5e some 130–120 ka (Kukla et al., 2002; Parker, 2009). Evidence for a second, later expansion (Rose et al., 2011) has been identified in southwestern Oman as the Dhofar Nubian Complex, which is linked to a population spread from Northeast Africa during MIS 5c some 110–100 ka. During wet periods, hunter-gatherer occupation penetrated deeply into the Arabian Peninsula, indicated by a stratified archaeological deposit found along a relict lakeshore in the Nefud Desert, dated to 75 ka (Petraglia et al., 2011). Which human species produced these Middle Stone Age tools in Arabia is still a matter of question, as no skeletal remains have yet been found with any of the Arabian assemblages. At the site of Taramsa 1 in Egypt, a modern human child was found associated with a Nubian Complex assemblage, suggesting that at

least the Dhofar Nubian Complex toolmakers were anatomically modern humans.

However, analyses of mtDNA variation do not link these MIS 5 human populations with the contemporary

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Ministry of Education of the Czech Republic; Grant number: KONTAKT ME 917 (to V. Č.). Grant sponsor: Council of American Overseas Research Centers, the American Institute for Yemeni Studies (to V. Č.), and FCT (the Portuguese Foundation for Science and Technology); Grant numbers: PTDC/CS-ANT/113832/2009, SFRH/BD/69119/2010 (to N.M.S.), SFRH/BPD/64233/2009 (to P.S.); Grant sponsor: UK Arts & Humanities Research Council; Grant number: AH/G012733/1 (to J.I.R.). Grant sponsor: United States National Science Foundation; Grant number: BCS-0518530 (to C.J.M.).

\*Correspondence to: Viktor Černý, Archaeogenetics Laboratory, Institute of Archaeology in Prague, Academy of Sciences of the Czech Republic, Letenská 4, 118 01 Prague 1, Czech Republic. E-mail: cerny@arup.cas.cz

Received 9 February 2012; accepted 23 July 2012

DOI 10.1002/ajpa.22131

Published online 24 August 2012 in Wiley Online Library (wileyonlinelibrary.com).

southern Arabian gene pool. The most ancient genetic signal of the Out of Africa expansion only reaches as far back as ~70 ka (MIS 4), when the first non-African mtDNA branches (M and N) diverged from East African haplogroup L3 (Behar et al., 2008b; Cabrera et al., 2009; Soares et al., 2012). Still today, several relict mtDNA lineages within N\* survive in Arabia and coalesce to the most ancient non-African mtDNA lineage ~60 ka ago (Fernandes et al., 2012). At the same time, there is virtually unanimous agreement among archaeologists working in the Arabian Peninsula that stone tool assemblages after MIS 5 are significantly different from coeval assemblages found in Africa; hence, indicating that there was no demographic movement from Africa into Arabia after 70 ka (Armitage et al., 2011; Bailey, 2009; Crasard, 2009; Marks, 2009; Rose and Petraglia, 2009; Rose, 2010; Rose et al., 2011). On the contrary, the Mudayyan Industry of Dhofar is cited as evidence for the survival of an indigenous population within the Dhofar refugium after MIS 5 (Usik et al., in press). Some have even raised the possibility of back migrations from southern Arabia into Africa during MIS 5a and/or MIS 3 (Rose and Usik, 2009).

Several mtDNA clades presently found in Arabia and neighboring regions show that its southern part played an important role in the Pleistocene-Holocene boundary. The frequent and widespread R0a haplogroup has provided an excellent example how the paleoenvironmental reconstructions of the Arabian refugia can be matched with phylogenetic and phylogeographic inferences (Černý et al., 2011). In fact, this haplogroup showed that a substantial part of the contemporary Yemeni population is biologically related to one or more demographic expansion events that have taken place over the last 20 ka, especially after 16 ka. Similarly, whole genome analyzes of HV1 clades detected close population contacts between both coasts of the Red Sea in the Holocene (Musilová et al., 2011).

The modern southern Arabian archaeology is contributing interesting insights into the important question of the population history around the Pleistocene-Holocene boundary in Arabia. Between approximately 9 and 8 ka, there was a sudden influx of archaeological sites appearing throughout the interior of Arabia, bearing a specialized toolkit comprised of low mass, hafted projectile points produced via a single platform blade technology. In its earliest manifestation, such tools are referred to as "Fasad points," typified by points manufactured on simple flakes and blades with a characteristic tang at the bottom for mounting onto a shaft. Some scholars have made comparisons between Fasad points and Pre-Pottery Neolithic B (PPNB) "Byblos points" found in the Levant, citing the presence of Fasad points throughout Arabia as representing an expansion of PPNB hunter-herders from the Levant (Dreschler, 2007; Uerpman et al., 2009). This scenario is contested by researchers

working in Dhofar, who report single platform blade technologies and Fasad points from archaeological contexts pre-dating the PPNB (Rose and Usik, 2009), instead arguing for an indigenous origin of the Early Holocene population in Arabia.

To more comprehensively address this question from a genetic perspective, we decided to explore subsequent points: fill lacunas of the previous sampling of the region, obtain information for salient mtDNA haplogroups through complete sequencing and apply modern statistical analyses allowing to infer population expansions and migrations in the past. We sampled three regions in southern Arabia, said to be placed in two post-glacial refugia: Al-Mahra and Dhofar in the border between Yemen and Oman, respectively; and Bab el-Mandab, the southwestern most tip in Yemen. The new samplings led to the identification of a quite under-represented haplogroup R2, for which we performed complete mtDNA sequencing in this work. For a statistical evaluation, we used Bayesian Skyline Plots (BSP) for this haplogroup together with other southern Arabian mtDNA haplogroups (R0a, HV1) for which we have obtained complete sequences before (Černý et al., 2011; Musilová et al., 2011). Based on the reconstructed phylogenies this method allowed us to reliably infer the past population dynamics (Drummond et al., 2005). Although R0a, HV1, and R2 cannot be identified with the southern Arabian population as a whole (they represent approximately one quarter of total haplogroup distribution), the signal associated with these haplogroups might nevertheless mirror demographic processes in the ancestral population as suggested previously (Atkinson et al., 2009). However, the BSPs can display changes of effective population size resulting either from expansion of local lineages or input from migrant lineages. To distinguish between them, we applied the founder analysis allowing to identify and date migrations between postulated source and sink populations (Richards et al., 2000). Both methodologies are based on reliable phylogenies and dating of lineages, which were implemented through the use of several phylogenetic techniques [maximum parsimony, maximum likelihood (ML) and Bayesian] and mtDNA mutation rates corrected for purifying selection (Soares et al., 2009).

## MATERIALS AND METHODS

### Population samples

The samples were collected from regions that were not yet screened in previous genetic studies of southern Arabian demographic prehistory (Richards et al., 2003; Kivisild et al., 2004; Černý et al., 2008; Černý et al., 2009). We secured new samples from Dhofar (southern Oman), Al-Mahra (eastern Yemen) and Bab al-Mandab (southern Yemen). For comparative purposes, we retrieved also a large number of published mtDNA sequences from neighboring populations. In total, we gathered 7,538 hypervariable segment I (HVS-I) mtDNA sequences from 91 populations (1,274 sequences from Arabian Peninsula, 1,357 from Near East and Iran, 591 from Middle East, 948 from northern India, 1,323 from southern India, 542 from Upper Nile Valley, and 1,503 from eastern Africa; a few additional eastern African mtDNA sequences (the Somali and Turkana) were further generated and merged to the published datasets; see Supporting Information Table S1 for further details and references).

---

#### Abbreviations

|      |                                      |
|------|--------------------------------------|
| BSP  | Bayesian Skyline Plots               |
| rCRS | Revised Cambridge Reference Sequence |
| HVS  | Hypervariable segment                |
| IDW  | Inverse Distance Weighted            |
| MDS  | Multidimensional scaling             |
| MIS  | Marine Isotope Stage 5               |
| ML   | Maximum likelihood                   |
| PPNB | Pre-Pottery Neolithic B              |

### Laboratory methods

We generated in total 329 new mtDNA HVS-I sequences (196 from Dhofar, 70 from Mahra, 30 from Bab al-Mandab, 24 from the Turkana, and 9 from the Somali). HVS-I sequences were generated by the primer P23 (Černý et al., 2011) allowing to acquire the variation from the segment between the nucleotide positions (np) 15,855–16,569 in majority of cases. PCR products were sequenced with the forward primers and in case of poly-C stretch reverse complement was produced as well. The sequences were compared with the revised Cambridge Reference Sequence (rCRS) and with the help of mtDNA-GeneSyn (Pereira et al., 2009) the mutated variants were identified. Haplogroup affiliations were made by help of the algorithm provided by HaploGrep software (Kloss-Brandstatter et al., 2011), which is linked with whole genome mtDNA sequences deposited in PhyloTree (van Oven and Kayser, 2009). Mutated variants of the sequences presented in this work together with analyzed ranges of control region variation and haplogroup affiliations are reported in Supporting Information Table S2.

For the whole genome mtDNA sequencing, we focused on haplogroup R2, whose internal variation was until now almost unknown. Up to date there were only six known complete sequences belonging to this haplogroup and reported in PhyloTree (van Oven and Kayser 2009) or in GenBank. We selected six R2 samples from our new dataset (marked red in the Supporting Information Table S2) and added one R2 sample from the dataset published by Černý et al., 2008 and two R2 samples from the dataset of Connie Mulligan (unpublished), therefore in total nine new R2 samples. We followed the same whole mtDNA genome sequencing methodology as reported in Černý et al., 2011. The complete mtDNA sequences of R2 haplogroup were than submitted to GenBank (accession numbers JX155264–JX155272).

### Statistical analyses

For comparative reasons we used 340 bp long HVS-I segment (from np 16,030 to 16,370). Among 91 analyzed populations, we calculated  $F_{ST}$  genetic distances as suggested by Reynolds et al., (1983) with help of Arlequin software ver. 3.5.1.2. (Excoffier et al., 2005) using 1,000 permutations and visualized them in two dimensions by multidimensional scaling (MDS) analysis. The geographical distribution of R2 haplogroup frequencies was visualized by construction of the interpolation maps using the “Spatial Analyst Extension” of ArcView version 3.2 ([www.esri.com/software/arcview/](http://www.esri.com/software/arcview/)). The “Inverse Distance Weighted” (IDW) option with power of two for the interpolation of the surface was used. See our previous studies for further details (Pereira et al., 2010a; Černý et al., 2011; Soares et al., 2012). For specific spatial analyses, such as the comparison between genetic and geographic matrices, Mantel test (Mantel, 1967) was used. In correlogram analysis Moran’s I metric was applied (Moran, 1950). Both previously mentioned analyses were assayed in PASSaGE software (Rosenberg, 2001).

Reduced-median network of R2 complete sequences led to suggested branching order for the R2 tree, which was then constructed most parsimoniously by hand. Age estimates were made with help of  $\rho$  statistic and ML. We used  $\rho$  with a mutation rate estimate for the complete mtDNA sequence of one substitution in every 3,624 years, which is corrected for purifying selection (Soares

et al., 2009), and a synonymous mutation rate of one substitution in every 7,884 years. The calculator provided in Soares et al., 2009 was used for these calculations. Standard errors were estimated as in Saillard et al., 2000. We have also obtained ML estimates of branch lengths using PAML 3.1361, assuming the HKY85 mutation model with gamma-distributed rates. We converted mutational distance in ML to time using the same complete mtDNA genome clock corrected for purifying selection.

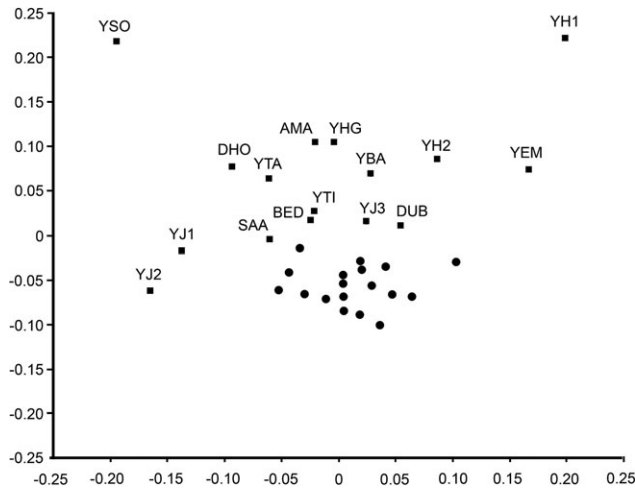
Thanks to the robust phylogeny reconstruction described above, we were able to apply BSPs. The signals of past population growth were estimated by help of BEAST 1.4.663 software (Drummond et al., 2005). In this analysis, we implemented not only R2 but also other southern Arabian haplogroups such as R0a and HV1 published before (Černý et al., 2011; Musilová et al., 2011). We have used generation time of 25 years, and BSPs were generated for the 142 mtDNA whole genome sequences when whole dataset was considered and 91 sequences when only southern Arabian samples were taken in consideration. Plots were generated using the mutation rates for the complete mtDNA genome obtained using internal calibration points in a study of haplogroup U5/U6 (Pereira et al., 2010b) of  $2.514 \times 10^{-8}$  substitutions/site/year. Analysis was run for a Markov chain length of 50,000,000 states. BSPs were further visualized with Tracer v1.3.

For the founder analysis, we built networks of HVS-I variation for R0a, HV1 and R2 for the Arabian Peninsula, Near East and East Africa. We estimated founder ages for each of the haplogroups assuming an  $f_1$  criterion (Richards et al., 2000), meaning that only shared sequences that presented variation in the hypothetical source would be considered as founders. An effective number of samples for each founder was calculated as previously (Soares et al., 2012) and founder ages were statistically distributed across 200-year intervals. In the case of R0a, HV1 and R2 lineages, it is not straightforward to postulate Near East and Arabian Peninsula as source and sink population, respectively, or vice-versa; but the method of founder analysis allows us to test and compare both these hypotheses. Near East and Arabian Peninsula were hypothetically considered source and sink populations to one another in two independent analyses. Mutation rate employed was of one mutation every 16,677 years (Soares et al., 2009).

## RESULTS

### Population and spatial analyses

$F_{ST}$  genetic distances based on mtDNA HVS-I sequences among the 16 Arabian population samples studied here show that even after Bonferroni’s correction around half values are still significant (see Supporting Information Table S3). When all the analyzed populations are pulled together in a MDS analysis (stress value 0.102) they cluster well according to their geography, suggesting that the analyzed dataset is not influenced by selection (Supporting Information Fig. S1). Interestingly, the eastern African populations are concentrated horizontally and occupy approximately the same amount of space in the graph as all analyzed non-Africans whose dispersion follows rather vertical pattern. Arabian populations are situated at the lower part of the graph and India at the upper part of this vertical dispersion. When the  $F_{ST}$  genetic distances for Arabian and Near Eastern



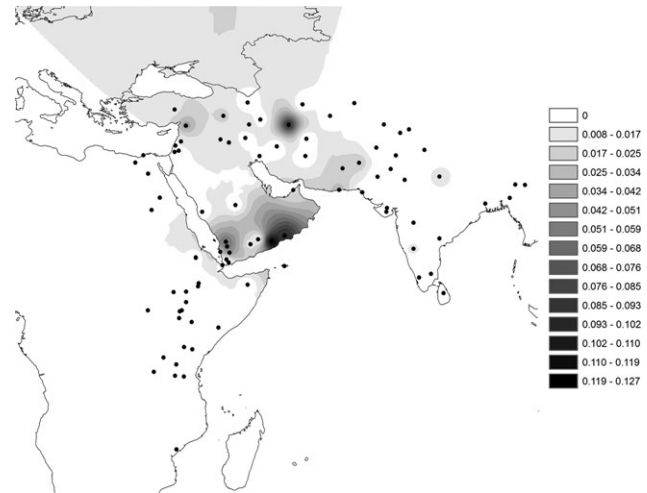
**Fig. 1.** The MDS of the population groups from Arabian Peninsula and Near East based on  $F_{ST}$  matrix distances. The Arabian samples are marked by squares, the Near Eastern by circles; see Supporting Information Table S1 for abbreviations of the Arabian populations.

populations are pulled together in a MDS analysis (stress value 0.128) a larger space is occupied by the Arabian groups than by the Near Eastern ones (Fig. 1). This is, however, due mainly to the outlier positions of some of its populations such as the Yemeni Jews (YJ1, YJ2), Soqatra (YSO), and Hadramawt (YH1), the rest of the analyzed Arabian groups being more or less close together.

The HVS-I mtDNA sequences obtained from Dhofar and Al-Mahra revealed relatively high frequency of relatively rare haplogroup R2. The visualization of the geographic distribution of R2 frequencies within the analyzed population samples confirmed its high occurrence in southern Arabia, especially in Dhofar (10%) and Al-Mahra (12%) but revealed also its rather discontinuous presence in same places of the Near and Middle East (currently between 1 and 4% with exception of the Mazandarian sample where it surprisingly occurs in more than 9% (Fig. 2). Relatively stable frequency of R2 (5–6%) is encountered also in all three population samples of the Yemeni Jews. Owing to this patchy distribution no intelligible geographic pattern was observed in spatial analyses using different combinations of distance classes in Moran's correlograms.

### Tree R2 and the phylogeographic inferences

The tree of fifteen R2 sequences (nine of them generated in this study and six from the published datasets; see Supporting Information Table S4) is presented in Figure 3. It is composed from ten southern Arabian (one is from a Yemeni Jew—EF556167), one Pakistani (EU597551), one Indian (AY714007), one Mongolian (HM030516), one Turkish (HM852870) sequences, and one unpublished sequence from an Italian lab (JN581649). We confirmed that R2 is distinguished from its ancestor R by several coding region mutations (4,216; 4,769; 7,657; 8,473; 9,932; 10,685; 12,654; 13,500; and 14,305) and two control region mutations (152, 16071). R2 as a younger sister of JT is dated to Late Glacial/post-Last Glacial Maximum around 18–21 ka and can be divided into two sub-haplogroups, which we call here R2a and R2b. Both these subgroups are defined by very



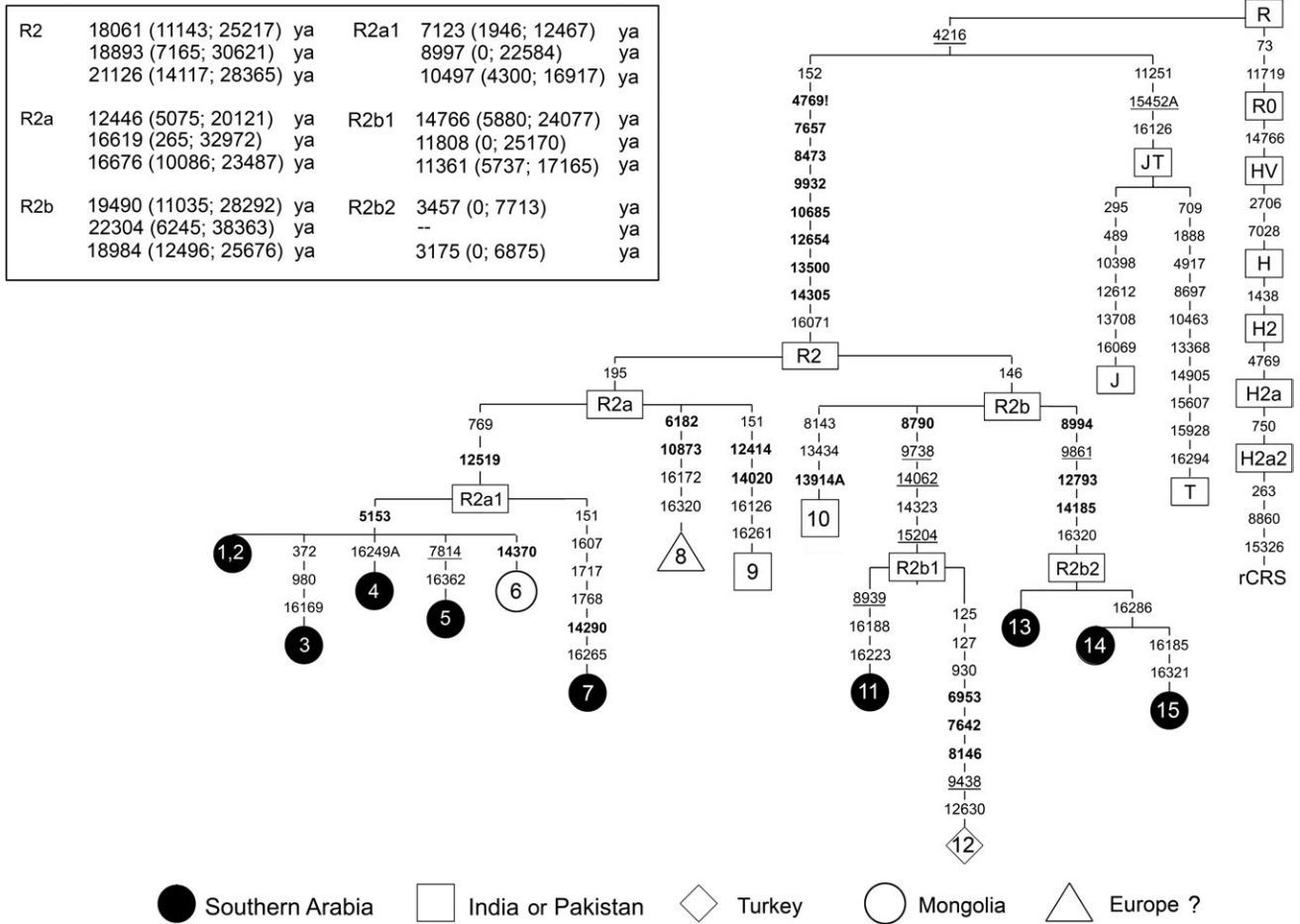
**Fig. 2.** Interpolation map for R2 haplogroup.

fast mutations, 195 and 146, respectively. Their phylogenetic status can thus be taken as hypothetical, although it does not change drastically the inferences obtained. R2a has three branches, first is mainly from southern Arabia (R2a1) except one sample from Mongolia but the other two are from different regions (India and probably Europe) gravitationally indicating here the unrepresented Near East as the most probable point of origin for R2a. R2b also presents three branches with similar distributions—one was only detected in southern Arabia (R2b2), one was represented by a sample from Pakistan and the third one (R2b1) was detected in Turkey and Yemen, although the last was a Yemeni Jew with some possibility of a Near Eastern origin (Behar et al., 2008a). Given this pattern, the unrepresented Near East can again be considered as a possible origin for R2b. Although R2 is a very rare haplogroup and much of its internal variation is still undetected, the phylogeographic analysis indicates that the Near East is the likely original region for this haplogroup, an inference that is supported by the HVS-I diversity, as we will discuss below.

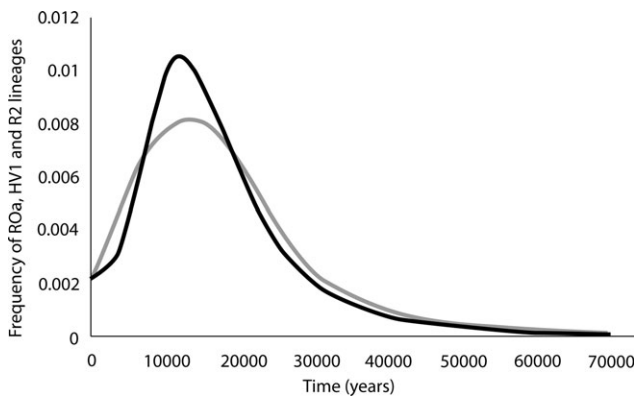
### Founder analysis and past population growth

We scanned possible migration times of haplogroups R0a, HV1 and R2 hypothetically from Near East to Arabian Peninsula and in the opposite direction using the founder analysis approach on the HVS-I data (Richards et al., 2000; Soares et al., 2012). Both scanners provided a peak around 12 ka, just slightly older but not as well defined if we consider the hypothetical migration from Arabia into the Near East (Fig. 4). The similarity of the age of the peaks relates to the fact that both mainly correspond to data from the shared haplotypes that are also the most common on the network (Supporting Information Table S5). This suggests that the clades in the analyses mostly expanded ~12 ka, an expansion time that matches the population increase observed in the BSP, as shown below.

It is much more difficult to assess if these clades moved from the Near East into the Arabia or vice-versa. We checked the overall diversity of each clade in both regions. Haplogroup R2 has substantially higher diversity in the Near East compared with the Arabian Penin-



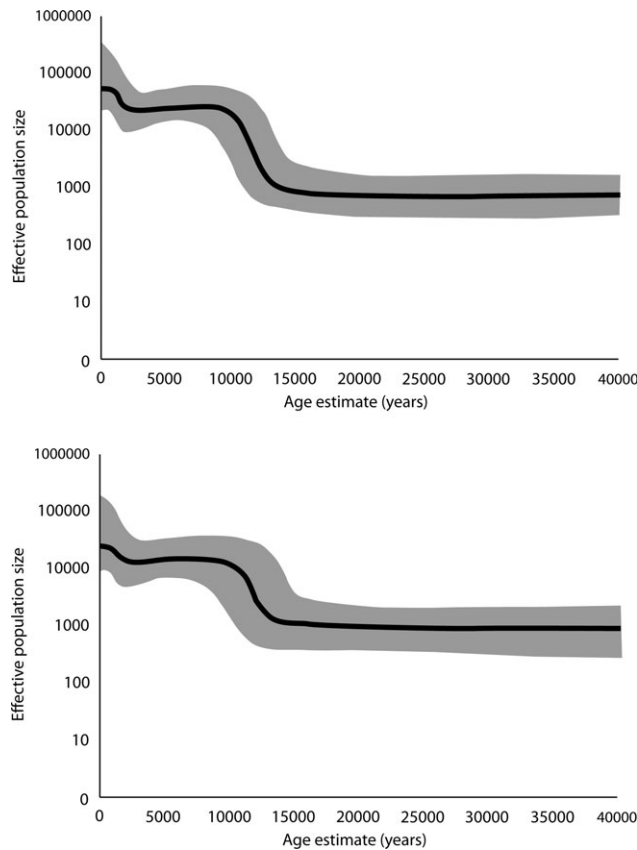
**Fig. 3.** Tree of the R2 haplogroup. Integers represent transition, only suffix A, G, C, or T indicates a transversion. Integers in bold indicate synonymous substitutions, underlined non-synonymous. Age estimates for R2 and R2a and R2b are based on diversity for the complete molecule (first lines), only synonymous polymorphisms (second lines) and ML estimates (third lines); rCRS—revised Cambridge Reference Sequence; the insertion 3151.1C was not considered. Provenience of the samples: 1—Tihama (JX155266), 2—Dhofar (JX155267), 3—Dhofar (JX155264), 4—Dhofar (JX155265), 5—Al-Mahra (JX155270), 6—inner Mongolia (HM030516), 7—Al-Mahra (JX155268), 8—Italy ? (JN581649), 9—India (AY714007), 10—Pakistan (EU597551), 11—Yemeni Yew (EF556167), 12—Turkey (HM852870), 13—Soqotra (JX155271), 14—Al-Mahra (JX155269), and 15—Dhamar (JX155272). For the main features of the samples see also Supporting Information Table S4.



**Fig. 4.** Probabilistic distribution of R2, R0a, and HV1 founder clusters across migration times scanned at 200 years intervals from 0–70 ka, using the f1 criterion, considering an origin of the haplogroups in the Near East and consequent migration into the Arabian Peninsula (black line) and vice-versa (grey line).

sula, when considering mean number of pairwise differences  $\pi$  (2.364 against 1.371) and gene diversity (or heterozygosity)  $H$  (0.860 against 0.687). This suggests that R2 probably had an origin in the Near East, but further complete sequences from that region are needed to more properly address that question. The very frequent R0a also showed higher diversity in the Near East for  $\pi$  and  $H$  (3.083 and 0.944, respectively) when compared with the Arabia Peninsula (2.253 and 0.896). HV1 showed an ambiguous result with higher diversity for  $\pi$  in the Arabian Peninsula (1.237 against 1.116) but higher diversity in the Near East in terms of  $H$  (0.674 against 0.456). More samples from the Near East are required to draw more detailed conclusions as to the geographic origin of these clades, but the overall pattern shown already here suggests that they had a more likely origin in the Near East and moved into the Arabian Peninsula ~12 ka.

The BSPs based on complete mtDNA sequences of the southern Arabian haplogroups R0a, HV1 and R2 whose main features are presented in Supporting Information Table S4 (irrespective if taken from southern Arabia



**Fig. 5.** BSPs indicating hypothetical effective population size changes through time based on data from the entire dataset (above) and only southern Arabia (below). For the main features of the complete sequences analyzed here see Supporting Information Table S4.

only or from overall datasets) point to two main episodes of population growth (Fig. 5). The first is steepest at  $\sim 13$  ka, followed by stagnation from about 8 ka. A comparatively slighter increase is further visible in a very recent period, some 2 ka. Peaks of increment of the effective population size as obtained from the BSPs are included in Supporting Information Table S6.

## DISCUSSION

The question about continuity or discontinuity of the settlement in the Arabian Peninsula after the first colonization by modern humans is an important issue (Maher, 2009). In this study, we have analyzed three clades (R0a, HV1, and R2) representing  $\sim 22\%$  of the total mtDNA gene pool of the southern Arabia with some branches specific for this regions. We thus contend they should be linked with important aspects of the demographic history of the region. We also significantly increased the sampling within southern Arabia by characterizing at the mtDNA level three previously unsampled locations such as Dhofar in Oman and Al-Mahra and Bab el-Mandab in Yemen.

The increased sampling and the finer molecular characterization of haplogroup R2, together with the application of sophisticated statistical methodologies (BSP and founder analysis) applied to two other clades we analyzed previously without applying these methods,

allowed us to infer the role of these high frequent clades in the population history of southern Arabia.

Both the BSP and the founder analysis indicated that these clades spread into the southern portion of the Arabian Peninsula around 13–12 ka, leading to population growth in the region. Most probably, these clades had an origin in the Near East and expanded southward. Thus, these genetic data indicate a great influx of lineages into the area at the end of the Pleistocene/beginning of the Holocene. It can be suggested that the peoples bearing the ancestors of these haplogroups were settled in an Arabian refugium, possibly the Persian Gulf oasis (Fernandes et al., 2012), and only later spread to the Near East and consequently to Southern Arabia.

On the other hand, such results are difficult to reconcile with the proposed Neolithic expansion of PPNB hunter-herders from the Levant (Dreschler, 2007; Uerpman et al., 2009), which is predicated on the tenuous link between tanged arrowheads of the Fasad or Qatar B-Type with Byblos points typical of the Levantine PPNB (Kapel, 1967). Given the timeframe of the PPNB in the Levant between 10 and 8 ka, this expansion is thought to have either occurred early in the PPNB during the Holocene climatic optimum, or toward the end of the PPNB; in the latter case, demographic movement is attributed to environmental desiccation in the Levant that resulted in population pressure.

Recently, pre-Neolithic sites dating between 13 and 8 ka have been found in the Yemeni Highlands (Fedele, 2009), Dhofar (Rose and Usik, 2009), and United Arab Emirates (Uerpman et al., 2009), all of which bear lithic technologies that are clearly unrelated to the Levantine PPNB. In Dhofar and Yemen, these findings have been interpreted as suggesting some degree of indigenous human occupation in southern Arabia reaching at least back to the Terminal Pleistocene.

It is noteworthy, however, that these pre-Neolithic sites do not bear any technological traits analogous to Terminal Pleistocene (Epipalaeolithic) assemblages found in the Near East. The only germane possibility of a connection between Arabia and the Near East during this period comes from the Faw Well site at the western edge of the Rub' Al Khali (Edens, 2001). Although undated, the Faw Well lithic assemblage bears a close resemblance to the Late Ahmarian of the Levant (20–17 ka). Perhaps it was this, or a subsequent pulse from the Levant, that provided the demographic input expressed by the genetic lineages documented in this article.

The results from the three analyzed southern Arabian clades do not support population continuity from the first occupants more than 50 ka ago (Fernandes et al., 2012) but do suggest some continuity across the Pleistocene-Holocene boundary. Our analysis indicates that the observed population expansion 13–12 ka is probably the result of genetic input from the Near East a few thousand years before the (debated) arrival of the PPNB culture in Arabia. If, however, there was a population expansion southward through Arabia some 13–12 ka, we have not yet found its archaeological signatures. Both regions exhibit stone tool technologies with some overlapping features, so it is warranted to suppose that we may one day locate a firm link between southern Arabia and the Near East sometime during the Late Pleistocene. Given the vast amount of unexplored territory in Arabia and paucity of archaeological sites with numerical ages, future investigations (both archaeogenetical

and archaeological) throughout the Peninsula will undoubtedly serve to shed more light on this question.

### ACKNOWLEDGMENTS

IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by FCT.

### LITERATURE CITED

- Armitage SJ, Jasim SA, Marks AE, Parker AG, Usik VI, Uerpmann HP. 2011. The southern route "out of Africa": evidence for an early expansion of modern humans into Arabia. *Science* 331:453–456.
- Atkinson QD, Gray RD, Drummond AJ. 2009. Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc R Soc B* 276:367–373.
- Bailey G. 2009. The Red Sea, coastal landscapes, and Hominin dispersals. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 15–38.
- Behar DM, Metspalu E, Kivisild T, Rosset S, Tzur S, Hadid Y, Yudkovsky G, Rosengarten D, Pereira L, Amorim A, Kutuev I, Gurwitz D, Bonne-Tamir B, Villems R, Skorecki K. 2008a. Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. *PLoS One* 3:e2062.
- Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith Ch, Wells RS, Rosset S, and The Genographic Consortium. 2008b. The dawn of human matrilineal diversity. *Am J Hum Genet* 82:1130–1140.
- Cabrera VM, Abu-Amero KK, Larruga JM, González AM. 2009. The Arabian Peninsula: gate for human migrations out of Africa or cul-de-sac? A mitochondrial DNA phylogeographic perspective. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 79–87.
- Crassard R. 2009. The Middle Paleolithic of Arabia: the view from the Hadramawt Region, Yemen. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 151–168.
- Černý V, Mulligan CJ, Rídl J, Žaloudková M, Edens CM, Hájek M, Pereira L. 2008. Regional differences in the distribution of the sub-Saharan, West Eurasian, and South Asian mtDNA lineages in Yemen. *Am J Phys Anthropol* 136:128–137.
- Černý V, Pereira L, Kujanová M, Vašíková A, Hájek M, Morris M, Mulligan CJ. 2009. Out of Arabia—the settlement of island Soqatra as revealed by mitochondrial and Y chromosome genetic diversity. *Am J Phys Anthropol* 138:439–447.
- Černý V, Mulligan CJ, Fernandes V, Silva NM, Alshamali F, Non A, Harich N, Cherni L, El Gaaied AB, Al-Meerri A, Pereira L. 2011. Internal diversification of mitochondrial haplogroup R0a reveals post-last glacial maximum demographic expansions in South Arabia. *Mol Biol Evol* 28:71–78.
- Dreschler P. 2007. The Neolithic dispersal into Arabia. *Proc Sem Arab Stud* 37:93–109.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22:1185–1192.
- Edens C. 2001. A bladelet industry in southwestern Saudi Arabia. *Arab Archaeol Epigr* 12:137–142.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.
- Fedele FG. 2009. Early Holocene in the highlands: data on the peopling of the Eastern Yemen Plateau, with a note on the Pleistocene evidence. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 215–236.
- Fernandes V, Alshamali F, Alves M, Costa MD, Pereira JB, Silva NM, Cherni L, Harich N, Černý V, Soares P, Richards MB, Pereira L. 2012. The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. *Am J Hum Genet* 90:347–355.
- Kapel H. 1967. *Atlas of the stone-age cultures of Qatar*. Aarhus: Aarhus University Press.
- Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet* 75:752–770.
- Kloss-Brandstaetter A, Pacher D, Schoenherr S, Weissensteiner H, Binna R, Specht G, Kronenberg F. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Available at: <http://www.haplogrep.uibk.ac.at>.
- Kukla GJ, Bender ML, de Beaulieu J-L, Bond G, Broecker WS, Cleveringa P, Gavin JE, Herbert TD, Imbrie J, Jouzel J, Keigwin LD, Knudsen KL, McManus JF, Merkt J, Muhs DR, Müller H, Poore RZ, Porter SC, Seret G, Shackleton NJ, Turner Ch, Tzedakis PC, Winograd IJ. 2002. Last interglacial climates. *Quat Res* 58:2–13.
- Maher LA. 2009. The Late Pleistocene of Arabia in relation to the Levant. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 187–204.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220.
- Marks AE. 2009. The Paleolithic of Arabia in an inter-regional context. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 295–308.
- Moran PA. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37:17–23.
- Musilová E, Fernandes V, Silva NM, Soares P, Alshamali F, Harich N, Cherni L, Gaaied AB, Al-Meerri A, Pereira L, Černý V. 2011. Population history of the Red Sea: genetic exchanges between the Arabian Peninsula and East Africa signaled in the mitochondrial DNA HV1 haplogroup. *Am J Phys Anthropol* 145:592–598.
- Parker AG. 2009. Pleistocene climate change in Arabia: developing a framework for Hominin dispersal over the last 350 ka. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 39–50.
- Pereira L, Černý V, Cerezo M, Silva NM, Hájek M, Vašíková A, Kujanová M, Brdicka R, Salas A. 2010a. Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. *Eur J Hum Genet* 18:915–923.
- Pereira L, Freitas F, Fernandes V, Pereira JB, Costa MD, Costa S, Maximo V, Macaulay V, Rocha R, Samuels DC. 2009. The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet* 84:628–640.
- Pereira L, Silva NM, Franco-Duarte R, Fernandes V, Pereira JB, Costa MD, Martins H, Soares P, Behar DM, Richards MB, Macaulay V. 2010b. Population expansion in the North African late Pleistocene signalled by mitochondrial DNA haplogroup U6. *BMC Evol Biol* 10:390.
- Petraglia MD, Alsharekh AM, Crassard R, Drake NA, Groucutt H, Parker AG, Roberts RG. 2011. Middle Paleolithic occupation on a Marine Isotope Stage 5 lakeshore in the Nefud Desert, Saudi Arabia. *Quat Sci Rev* 30:1555–1559.
- Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the co-ancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767–779.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Norby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A,

- Bandelt HJ. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251–1276.
- Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V, Torroni A. 2003. Extensive female-mediated gene flow from sub-Saharan Africa into Near Eastern Arab populations. *Am J Hum Genet* 72:1058–1064.
- Rose JI. 2010. New light on human prehistory in the Persian Gulf oasis. *Curr Anthropol* 51:849–883.
- Rose JI, Petraglia MD. 2009. Tracking the origin and evolution of human populations in Arabia. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 1–12.
- Rose JI, Usik VI. 2009. The “Upper Paleolithic” of South Arabia. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 169–185.
- Rose JI, Usik VI, Marks AE, Hilbert YH, Galletti CS, Parton A, Geiling JM, Černý V, Morley MW, Roberts RG. 2011. The Nubian complex of Dhofar, Oman: an African Middle Stone Age industry in southern Arabia. *PLoS One* 6:e28239.
- Rosenberg M, editor. 2001. *PASSAGE, Pattern Analysis, Spatial Statistic, and Geographic Exegesis*. Version I.I. Tempe, AZ: Department of Biology, Arizona State University.
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, Norby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67:718–726.
- Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilova E, Macaulay V, Richards MB, Černý V, Pereira L. 2012. The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol Biol Evol* 29:915–927.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84:740–759.
- Uerpmann H-P, Potts DT, Uerpmann M. 2009. Holocene (re-) occupation of Eastern Arabia. In: Petraglia MD, Rose JI, editors. *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics*. New York: Springer. p 205–210.
- Usik VI, Rose JI, Hilbert YH, Van Peer P, Marks AE. In press. Nubian Complex reduction strategies in Dhofar, southern Oman. *Quat Int*.
- van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386–E394.