3-1-2014

# Lactase Persistence Variants in Arabia and in the African Arabs

Edita Priehodova
*Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic*

Abdelhay Abdelsawy
*The National Corporation for Antiquities and Museums (NCAM), Khartoum, Soudan*

Evelyne Heyer
*National Museum of Natural History, Paris, France*

Viktor Cerny
*Institute of Archaeology, Academy of Science, Prague, Czech Republic*, cerny@arup.cas.cz

## Recommended Citation

Priehodova, Edita; Abdelsawy, Abdelhay; Heyer, Evelyne; and Cerny, Viktor, "Lactase Persistence Variants in Arabia and in the African Arabs" (2014). *Human Biology Open Access Pre-Prints.* Paper 48.
http://digitalcommons.wayne.edu/humbiol_preprints/48

# Lactase Persistence Variants in Arabia and in the African Arabs

Edita Priehodova,[1] Abdelhay Abdelsawy,[2] Evelyne Heyer,[3] and Viktor Cerny[4]*

[1]Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic.

[2]The National Corporation for Antiquities and Museums (NCAM), Khartoum, Soudan.

[3]National Museum of Natural History, Paris, France.

[4]Institute of Archaeology, Academy of Science, Prague, Czech Republic.

*Correspondence to: Victor Cerny. E-mail: cerny@arup.cas.cz.

**Key words: Arabia, Arica, lactase persistence, pastoralism.**

**Short Title: Lactase Persistence Variants in Arabia and in the African Arabs**

**Abstract.** Lactase persistence (LP), the state enabling the digestion of milk sugar in adulthood occurs only in some human populations. The convergent and independent origin of this physiological ability in Europe and Africa is

linked with animal domestication that had either started in both places independently or had spread from the Near East by acculturation. However, it has recently been shown that at least in its southern parts, the population of Arabia not only has a different LP-associated mutation profile than the rest of Africa and Europe but had also experienced an independent demographic expansion occurring before the Neolithic around the Pleistocene – Holocene boundary. In Arabia, LP is associated with mutation -13,915*G and not, as in Europe, with -13,910*T, nor, as in Africa, with -13,907*G and -14,010*C. We show here that, in Arabia, -13,915*G frequency conforms to a partial clinal pattern and that this specific mutation has likely been spread from Arabia to Africa only recently from the 6th century AD onward by nomadic Arabs (Bedouins) looking for new pastures. Arabic populations in Africa that still maintain a nomadic way of life also have more -13,915*G variants and fewer sub-Saharan L-type mtDNA haplogroups; this observation matches archaeological and historical records suggesting that the migration of Arabic pastoralists was accompanied by gradual sedentarization that allowed for admixture with the local African population.

Food-production in the Neolithic stimulated significant demographic growth of several human populations (Bocquet-Appel 2011) and facilitated the spread of certain traits having selective advantages. One such genetically determined trait, lactase persistence (LP), allows for the digestion of milk sugar (lactose)--an ability common in all mammals during nursing--to continue into adulthood rather than being limited to children (Swallow 2003). Since the milking of domestic animals became a common practice (Bersaglieri et al. 2004; Hollox et al. 2001) at this time, having the lactase persistence trait conferred advantage on its carriers  by allowing them to consume fresh milk as a quickly available and nutritious food source.

Lactase persistence has also been theorized to be particularly advantageous in populations exposed to lower levels of UV radiation that as a result suffer a concomitant lower production of vitamin D. Since vitamin D acts as a hormone that controls calcium levels in the blood, higher consumption of milk in these areas provides higher intake of calcium, preventing rickets. This calcium assimilation hypothesis, along with the lighter pigmentation supporting the production of vitamin D found in northern populations (Gerbault et al. 2009), is consistent with the fact that high frequencies of LP individuals are found in Europe in general, and that they are highest in northern European latitudes where UV radiation is weakest (Gerbault et al. 2009).

Higher frequency of LP has, however, also been reported in Africa and the Middle East, where UV radiation is sufficiently high that rickets do not pose a threat. The fact that the prevalence of the trait in these areas is chiefly tied to nomadic pastoralists with a strong dependence on milk consumption has inspired the hypothesis of co-evolution between genes and culture to be put forward (Holden and Mace 2002).

At the molecular level, there is evidence that LP can be caused independently by any of several specific mutations in an enhancer element regulating the expression of the lactase gene LCT. This element is located in an intron of the neighboring gene MCM6 (Enattah et al. 2002). The most frequent variant overall is -13,910*T, thought to have originated somewhere between central Europe and the Balkans ~ 7.5 ka (Bersaglieri et al. 2004; Itan et al. 2009) and to have subsequently been spread northward (Malmström et al. 2010). This “European” LP mutation also occurs in India (where it follows a general northwest to southeast declining pattern, (Gallego Romero et al. 2011), in central Asia (where no association with pastoralism has been identified (Heyer et al. 2011) and, last but not least, in the West African Sahel among the Fulani nomads (Lokki et al. 2011).

The -13,910*T variant, while being the most frequently occurring, is not the only lactase-regulating one documented (Ingram et al. 2007; Mulcare et al. 2004; Tishkoff et al. 2007). In Arabia, 13,915*G has arisen in parallel, possibly as an adaptation to consumption of camel milk (Enattah et al. 2008; Imtiaz et al. 2007). Africa has an even more complex repertoire of LP variants as well as a complex population history reflecting several genetic inputs from Eurasia evidenced by Eurasian uniparental markers present today especially among nomadic groups (Cruciani et al. 2010; Černý et al. 2011b; Olivieri et al. 2006; Pereira et al. 2010). Lactase persistence in Africa is associated with at least three native variants, including -13,907*G, -14,009*G, and -14,010*C, originating on the continent independently of each other (Ingram et al. 2007;

Jones et al. 2013; Tishkoff et al. 2007). In addition, however, Africa also harbors both the “European” -13,910*T and “Arabic” -13,915*G variants, both of which must have been introduced to the area in a similar manner to the above-mentioned Eurasian uniparental markers.

It has been suggested that the Neolithic in Arabia, at least in its southern part, evolved independently of that in the Near East (Rose and Usik 2009). Although the subsistence pattern of the first food-producing populations in Arabia is still not understood in full detail, it can be suggested that the domestication of animals (pastoralism linked with milking) came before the cultivation of domestic plants (Fedele 2009; McCorriston and Martin 2009). Using large mitochondrial DNA (mtDNA) datasets, we have recently proposed that the population of southern Arabia underwent expansion as early as 12 ka (Al-Abri et al. 2012), i.e. before the putative Near Eastern Neolithic input (Dreschler 2007). It has also been revealed that some mtDNA haplogroups such as R0a and HV1 were in fact recently introduced to Africa from Arabia (Černý et al. 2011a; Musilová et al. 2011), confirming contact between these two areas.

In this study we formulated several research questions and hypotheses to clarify the spread of the “Arabic” LP variant -13,915*G to Africa. First, if the variant arrived to Africa only recently with nomadic pastoralists, then higher frequencies should be found today in those Arabic groups leading a pastoralist lifestyle. Second, since the mating pattern in Arabic tribes is not always random as many social anthropologists maintain (Bonte 1994) and some genetic studies indeed determined certain degree of heterozygote deficiency possibly resulting from consanguinity (Cadenas et al. 2008), we have examined whether -13,915*G follows the Hardy-Weinberg equilibrium in the populations here studied. Last but not least, we analyzed 13,915*G frequencies across Arabia and Northeastern Africa for evidence of a cline. For the purposes of our analysis we present new LP data from five sedentary Arabian and five African Arabic populations. Of these 10 analyzed groups, while all from Arabia lead sedentary lifestyle, the Arabic-speaking tribes in Africa rely on differing degrees of nomadic pastoralism. We present here the data from 920 chromosomes and discuss our genetic results in a wider archaeological and historical framework.

## Material and Methods

**Population samples.** We have scanned for LP variants in five population samples from Arabia (Yemen) and five Arabic speaking populations in Africa. All but two (the Rashaayda from Sudan and the Sudanese Arabs) of the examined populations correspond with samples whose mtDNA data have been previously published (Černý et al. 2008; Černý et al. 2009; Černý et al. 2007; Kujanová et al. 2009). The results and geographical locations of all the populations examined here are presented in Table 1. While all Yemeni groups and two Arabic populations in Africa maintain a sedentary lifestyle, the ancestors of the three remaining Arabic speaking African populations sampled in Sudan, Chad and Nigeria entered Africa as nomadic pastoralists, and their present-day descendants still continue in this tradition. However, while the Rashaayda in Sudan and Baggara in Chad continue to rely on nomadism fully, the Shuwa in Nigeria today do already lead a more sedentary semi-nomadic lifestyle. These Shuwa Arabs are also, according to our previous mtDNA analyses, more admixed with sub-Saharan populations (Černý et al. 2007) . For some comparative purposes, previously published datasets characterizing other populations from the Arabian Peninsula and its neighborhood were included in some of the analyses; see the list in Appendix Table 1.

**Laboratory and statistical analyzes.** In total, 920 chromosomes (460 individuals) were screened for the variants associated with LP. We sequenced the 359bp fragment located in intron 13 of the MCM6 gene where the four main LP-associated mutations can be detected. We used the same primers as reported in a previous study (Coelho et al. 2009). PCR products were sequenced with forward primers and in case of ambiguities reverse complements were generated. The variants were identified by means of BioEdit software (Hall 1999) and allele frequencies then calculated. Chi-square tests for the Hardy-Weinberg equilibrium, heterozygosity, fixation indexes (F), analysis of molecular variance (AMOVA) and principal coordinate analysis (PCoA) were undertaken by means of the GenAlEx statistical package (Peakall and Smouse 2012). The spatial frequency distributions of the main “Arabian” LP variant -13,915*G found in Arabia and neighboring regions was visualized by constructing interpolation maps using the “Spatial Analyst Extension” of ArcView version 3.2 (www.esri.com/software/arcview/). The clinal pattern of -13,915-G within the Arabian Peninsula was further tested by correlogram autocorrelation analysis (Moran 1950) as implemented in PASSaGE software (Rosenberg 2001); these spatial analyses also included other previously published datasets (Appendix Table 1).

## Results

Genotypes and allelic frequencies of -13,915*G, the most frequently occurring LP variant in our populations, are reported in Table 2, showing that it appears -albeit in various frequencies - in all analyzed Arabian (Arabian Peninsula) and Arabic speaking (Africa) groups. While the Yemeni populations display more or less similar frequencies (except for the one from Hadramawt), Arabic speaking populations from Africa fall into one of two distinct groups – the first one having a high and the second one having a low population frequency of this variant. Interestingly, this division correlates with lifestyle, supporting the hypothesis of the recent introduction of this variant from Arabia; both nomadic groups from Sudan (the Rashaayda) and Chad (the Baggara) attain relatively high -13,915*G frequencies, while the semi-nomadic Shuwa from Nigeria and sedentary Arabs from Sudan and Egypt show a much lower frequency of this variant than do the Yemenis and the Arabian nomads from Africa.

AMOVA provides an inter-population insight into the structure of the LP variants in our samples; both -13,915*G as well as other more frequently occurring variants such as -13,910*T and -14,009*G were considered in this analysis. The percentage of whole molecular variance for all variants combined was 60% within populations and 40% among populations. When considering the -13,915*G variant alone, 56% of the molecular variance was within populations and 44% among populations; for -13,910*T alone, 93% of the molecular variance was within populations and 7% among populations; and for -14,009*G alone, 87% of the molecular variance was within populations and 13% among populations. Most of the overall variance between the combined populations is therefore explained by variant -13,915*G.

The construction of the PCoA plots was carried out on the PhiPT matrix values calculated by the above executed AMOVA. This analysis clearly separates the analyzed samples into two different groups (see Figure 1) – the first one composed of populations with higher frequencies of the 13,915*G variant as seen on the left of the figure, and the second group, located on the right of the figure, composed of populations with lower frequencies of this variant. The populations of the Hadramawt (YHA) and the Baggara Arabs (ACH), both having intermediate distributions of -13,915*G, lie between these groups

The Chi-square test for the Hardy-Weinberg equilibrium shows that the frequency of 13,915T/G genotypes is in equilibrium in all 10 analyzed populations (Table 2). This result is confirmed by comparisons of observed vs. expected heterozygosities (Ho vs. He) measured by fixation indexes (FI); only the population from Soqotra Island (YSO) and the Shuwa Arabs (ASW) from Nigeria are characterized by higher positive values indicating a very slight deviation toward an excess of homozygotes, even if levels of significance were not exceeded.

The geographical distribution of -13,915*G is presented in Figure 2, where its geographical dominance in Arabia is clearly visible. Only the Hadramawt (YHA) and most of the northern Omani populations (ODA, ODH, OMU and OBA; Appendix Table 1) have a lower frequency of this mutation. The diagram clearly shows that there is a virtual absence of -13,915*G eastward of Arabia in the Middle East and India. On the other hand, Africa has a much higher frequency of this variant. It occurs in the Bagarra and especially in the Rashaayda, both of which, as evidenced by the mtDNA data, have a relatively low African admixture; both groups are also highly dependent on nomadic pastoralism and milking. The correlogram analyses of the -13,915*G population frequencies (Indian, Chad and Nigerian data excluded) show that a clinal pattern cannot be rejected. In fact, there is a depression signal of a cline that does not affect the entire study area but only a part of it; see significant autocorrelation coefficients present for both the highest and lowest Moran's I values (Figure 3).

Aside from the above reported -13,915*G variant we have detected other mutations within the analyzed segment of the MCM6 gene (see Table 3). The "European" -13,910*T was found in two eastern Yemeni samples in Hadramawt and Soqotra, and in the Baggara and Shuwa. On the other hand, in

Yemeni samples the “African” LP-associated variants -13,907*G and -14,010*C  were revealed to come from the western part of the country, which is geographically close to Africa. In Africa, -13,907*G was found only in sedentary Arabs from Sudan. The Shuwa Arabs show two additional mutations: 13,965*G and -14,107*A. Moreover, -14,009*G was observed quite frequently in all analyzed samples of the African Arabs (except in the Rashaayda), and in one case from the western part of Yemen. Last but not least, one Egyptian from el-Hayz carries an additional -14,042*G mutation. All these mutations were detected in heterozygote state only (Table 3).

Interestingly, some of the variants described in the above paragraph were present together in the same individual, showing three combinations in total. The combination of -13,915*G and -14,009*G was seen twice in the population of the Baggara Arabs and once in the population of the Sudanese Arabs. A combination of the -13,910*T and -13,915*G variants was seen in three populations – once in the sample of Baggara Arabs, once in the Shuwa Arabs and once in the Yemeni 4 (YHA) group. One individual from Yemeni 1 (YAC) population also revealed a case of the combination of -14,010*C and -13,915*G.

## Discussion

It has been believed that the ancestors of the Yemeni people were (in contrast to the Bedouins of Saudi Arabia) sedentary cultivators (Chelhod 1984). However, we have shown here that LP, which both enables fresh milk consumption in larger quantity and supports the hypothesis of the co-evolution of genes and culture (Gerbault et al. 2009; Holden and Mace 2002), occurs with a relatively high frequency throughout Yemeni territory (perhaps except in the Hadramawt region). Although the original food-production system in southern Arabia is not yet understood in full detail, there are clear archaeological indications of the existence of cattle keepers in Yemen at the beginning of the Neolithic. In fact, a cattle-keeping population has been discovered (Fedele 2009) at the Early Neolithic Yemeni settlement at Chawlan, in the at-Tiyal area to the east of Sana'a (de Maigret 2003).

In the eastern part of Yemen (Hadramawt), the first traces of pastoralism are even more pronounced, as documented by 6,400-year-old cattle sacrifices found in Kheshiya (McCorriston et al. 2012). Such findings underline the economic and social importance of cattle, and likely also of milking practices. Archaeological evidence also reveals that, due to climatic deterioration in the Middle Holocene and its associated degradation of natural pastures, the original Hadramawt herders later started to concentrate in a more limited area and established the first irrigation systems to maintain pastures, enabling and with time leading to the cultivation of some plants (Harrower 2008). These changes subsequently led to a division within the southern Arabian population into nomadic pastoralists on one hand and settled farmers on the other. In fact, such archaeological findings match our genetic data showing that LP-associated variants are found throughout all Arabian populations, with the only exception being those found in the Hadramawt.

This discrepancy can be explained by the repeated waves of out-migration and back immigration experienced by this specific region of southern Arabia during the last 500 years (Manger 2010), also making it not surprising that the region today harbors an uncommonly high level of sub-Saharan mtDNA haplogroups (Černý et al. 2008).

The demographic history of the Arabian pastoralists is closely linked with the rise of South Arabian Caravan kingdoms and the flourishing trade with Mediterranean civilizations. Pastoralists took an active part in this business as domestication of the camel (Breton 1999; Retsö 1991) allowed them to export frankincense and other items across the barren desert of central Arabia. After the collapse of South Arabian civilization in 5th century AD some of the Arabian herders lost their jobs, however, and were forced look for new opportunities outside of Arabia. Their dispersal to Africa is linked with the spread of Islam and is carefully recorded in archaeological and historical sources (Levy and Holl 2002; Zeltner 2002).

The most important migration of the Arabian pastoralist tribes to Africa was associated with the conquest of Egypt in 7th century AD (Zeltner 2002), but deeper penetration into the continent was prevented for several centuries by the Christian kingdoms of Nubia. It was only after the Mamluks had conquered the Nubian Dongola in the beginning of 14th century that the Arabian pastoralists could continue with their further migration into the African interior. By the beginning of the 16th century they had dispersed themselves along the Blue and White Nile as well as westward to the Lake

Chad Basin. It is also known that Arabic tribes appeared between Lake Fitri in Chad and Bahr el-Ghazal in Sudan also in that century, in the area once called Shuwa (Seignobos 2000). Some tribes reached Baguirmi in the mid-16th century and the eastern shores of Lake Chad at the beginning of the 18th century, crossing the Shari River shortly thereafter. Immigration of Arabian pastoralists to Africa was, however, likely a continuous process, as is evidenced by the Rashaayda Bedouins who entered Africa only very recently in the 1860s (Young 1996).

The above-described wandering of the Arabic tribes through Africa has been revealed by both archaeological findings and historical documents, and is consistent with our observed frequencies of -13,915*G: its prevalence and partly clinal pattern within the Arabian Peninsula and closer-lying parts of Africa on one hand, and patchy distribution within broader Africa on the other. Interestingly, while the recent Arabic immigrants (the Rashaayda) bear only the -13,915*G variant, groups arriving earlier have gained a more varied repertoire of LP-associated mutations which might have been introduced into their population by other African nomads. This is especially the case for the -14,009*G variant, which might have been introduced into the Arabic pastoralist population in Africa by Somali camel herders (Ingram et al. 2009). The Shuwa, who reached as far as the Lake Chad Basin, bear -13,910*T; this variant might have been introduced into their population by possible contacts with the Fulani, the only African population where this “European” variant has been detected (Lokki et al. 2011).

In conclusion, our data contribute to the hypothesis that -13,915*G's potential place of origin is located in Arabia (Enattah et al. 2008). The high frequency of lactase persistency in Arabia, and even in the southern locations where plant cultivation is more significant than pastoralism, is consistent with archeological records describing a Yemeni Early Neolithic mode of subsistence (Fedele 2009; McCorriston and Martin 2009). We show that the non-Arabic speaking people on Soqotra bear a similar frequency of the -13,915*G variant to that found on the Arabian mainland which in fact suggests that this island had likely been colonized by people already bearing this variant; indeed, a relatively recent colonization of this specific island around 6 ka has been revealed by mtDNA analyses (Černý et al. 2009). Furthermore, the decreasing gradient of the frequency of -13,915*G among Arabic speaking groups in Africa supports a model of southward migration, sedentarization and admixture with local African populations as is evidenced in the Shuwa Arabs (Levy and Holl 2002).

Our study also presents the highest frequency of the -13,915*G variant so far recorded in the genetic literature, attaining 76.9% in the Rashaayda Bedouins who, while today living in Eastern Sudan, have descended from Arabian ancestors that arrived to Africa only some 150 years ago. The ancestors of the African Rashaayda originally probably lived in the region of Hejaz lying in western Saudi Arabia close to Mecca. In fact, there are many tribes in Arabia bearing the name Rashaayda, but all claim to have originated in Hejaz. Interestingly, further support of the Hejaz origin of the African

Rashaayda is provided by the similar features of their female costumes (Young 2000).

**Acknowledgements.** The authors are grateful to all project participants who provided samples to this study and two anonymous reviewers for their comments on an earlier version of the manuscript. The project was supported by the Grant Agency of the Czech Republic (Grant no. 13-37998S-P505) and Grant Agency of Charles University (Grant no 651112).

## Literature Cited

Al-Abri, A. R., E. Podgorná, J. I. Rose et al. 2012. Pleistocene-Holocene boundary in Southern Arabia from the perspective of human mtDNA variation. *Am. J. Phys. Anthropol.* 149:291–298.

Bersaglieri, T., P. C. Sabeti, N. Patterson et al. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* 74:1,111–1,120.

Bocquet-Appel, J. P. 2011. When the world's population took off: The springboard of the Neolithic Demographic Transition. *Science* 333:560–561.

Bonte, P. 1994. *Épouser au plus proche. Inceste, prohibitions et stratégies matrimoniales autour de la Méditerranée*. Paris: Éditions de l'École des Hautes, Études en Sciences Sociales, coll. Civilisations et Sociétés 89.

Breton, J.-F. 1999. *Arabia Felix from the Time of the Queen of Sheba: Eighth Century B.C. to First Century A.D.* Notre Dame: University of Notre Dame Press.

Cadenas, A. M., L. A. Zhivotovsky, L. L. Cavalli-Sforza et al. 2008. Y-chromosome diversity characterizes the Gulf of Oman. *Eur. J. Hum. Genet.* 16:374–386.

Coelho, M., F. Sequeira, D. Luiselli et al. 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol. Biol.* 9:80.

Cruciani, F., B. Trombetta, D. Sellitto et al. 2010. Human Y chromosome haplogroup R-V88: A paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *Eur. J. Hum. Genet.* 18:800–807.

Černý, V., C. J. Mulligan, V. Fernandes et al. 2011a. Internal diversification of mitochondrial haplogroup R0a reveals post-last glacial maximum demographic expansions in South arabia. *Mol. Biol. Evol.* 28:71–78.

Černý, V., C. J. Mulligan, J. Rídl et al. 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., L. Pereira, M. Kujanová et al. 2009. Out of Arabia-the settlement of island Soqotra as revealed by mitochondrial and Y chromosome genetic diversity. *Am. J. Phys. Anthropol.* 138:439–447.

Černý, V., L. Pereira, E. Musilová et al. 2011b. Genetic structure of pastoral and farmer populations in the African Sahel. *Mol. Biol. Evol.* 28:2,491–2,500.

Černý, V., A. Salas, M. Hájek et al. 2007. A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome. *Ann. Hum. Genet.* 71:433–452.

de Maigret, A. 2003. *Arabia Felix. An Exploration of the Archaeological History of Yemen*. London: Stacey International.

Dreschler P. 2007. The Neolithic dispersal into Arabia. *Proceedings of the Seminar for Arabian Studies* 37:93–109.

Enattah, N. S., T. G. Jensen, M. Nielsen et al. 2008. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am. J. Hum. Genet.* 82:57–72.

Enattah, N. S., T. Sahi, E. Savilahti et al. 2002. Identification of a variant associated with adult-type hypolactasia. *Nature Genet.* 30:233–237.

Fedele, F. G. 2009. Early Holocene in the Highlands: Data on the Peopling of the Eastern Yemen Plateau, with a Note on the Pleistocene Evidence. In *The Evolution of Human Populations in Arabia Paleoenvironments, Prehistory and Genetics*, M. D. Petraglia and J. I. Rose, eds. New York: Springer.

Gallego Romero, I., C. Basu Mallick, A. Liebert et al. 2012. Herders of Indian and European cattle share their predominant allele for lactase persistence. *Mol. Biol. Evol.* 29:249–260.

Gerbault, P., C. Moret, M. Currat et al. 2009. Impact of selection and demography on the diffusion of lactase persistence. *PLoS ONE* 4:e6369.

Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic. Acids. Symp. Ser.* 41:95–98.

Harrower, M. J. 2008. Hydrology, ideology and the origins of irrigation in ancient southwest Arabia. *Curr. Anthropol.* 49:497–510.

Heyer, E., L. Brazier, L. Segurel et al. 2011. Lactase persistence in central Asia: Phenotype, genotype, and evolution. *Hum. Biol.* 83:379–392.

Holden, C., and R. Mace 2002. Pastoralism and the evolution of lactase persistence. In *Human Biology and Pastoral Populations*, W. R. Leonard and H. Crawford, eds. Cambridge: Cambridge University Press.

Hollox, E. J., M. Poulter, M. Zvarik et al. 2001. Lactase haplotype diversity in the Old World. *Am. J. Hum. Genet.* 68:160–172.

Chelhod, J. 1984. *Le peuple yéménite et ses racines*. Paris: G.-P. Maisonneuve et Larose.

Imtiaz, F., E. Savilahti, A. Sarnesto et al. 2007. The T/G 13915 variant upstream of the lactase gene (LCT) is the founder allele of lactase persistence in an urban Saudi population. *J. Med. Genet.* 44:e89.

Ingram, C. J., M. F. Elamin, C. A. Mulcare et al. 2007. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum. Genet.* 120:779–788.

Ingram, C. J., T. O. Raga, A. Tarekegn et al. 2009. Multiple rare variants as a cause of a common phenotype: Several different lactase persistence associated alleles in a single ethnic group. *J. Mol. Evol.* 69:579–588.

Itan, Y., A. Powell, M. A. Beaumont et al. 2009. The origins of lactase persistence in Europe. *PLoS Comput. Biol.* 5:e1000491.

Jones, B. L., T. O. Raga, A. Liebert et al. 2013. Diversity of lactase persistence alleles in Ethiopia: Signature of a soft selective sweep. *Am. J. Hum. Genet.* 93:538–544.

Kujanová, M., L. Pereira, V. Fernandes et al. 2009. Near eastern neolithic genetic input in a small oasis of the Egyptian Western Desert. *Am. J. Phys. Anthropol.* 140:336–346.

Levy, T. E., and A. F. C. Holl. 2002. Migrations, ethnogenesis and settlement dynamics: Israelites in Iron Age Canaan and Shuwa-Arabs in the Chad Basin. *J. Anthropol. Archaeol.* 21:83–118.

Lokki, A. I., I. Jarvela, E. Israelsson et al. 2011. Lactase persistence genotypes and malaria susceptibility in Fulani of Mali. *Malar. J.* 10:9.

Malmström, H., A. Linderholm, K. Liden et al. 2010. High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evol. Biol.* 10:89.

Manger L. 2010. *The Hadrami Diaspora: Community-building on the Indian Ocean Rim*. New York: Berghahn Books.

McCorriston, J., M. Harrower, L. Martin et al. 2012. Cattle cults of the Arabian Neolithic and early territorial societies. *Am. Anthropol.* 114:45–63.

McCorriston, J., and L. Martin 2009. Southern Arabia's early pastoral population history: Some recent evidence. In *The Evolution of Human Populations in Arabia Paleoenvironments, Prehistory and Genetics,* M. D. Petraglia and J.I . Rose, eds. New York: Springer.

Moran, P. A. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37:17–23.

Mulcare, C. A., M. E. Weale, A. L. Jones et al. 2004. The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am. J. Hum. Genet.* 74:1,102–1,110.

Musilová, E., V. Fernandes, N. M. Silva et al. 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.

Olivieri, A., A. Achilli, M. Pala et al. 2006. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* 314:1,767–1,770.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--An update. *Bioinformatics* 28:2,537–2,539.

Pereira, L., V. Černý, M. Cerezo et al. 2010. 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.

Retsö, J. 1991. The domestication of the camel and the establisment of the frankincence road from South Arabia. *Orientalia Suecana* 40:187–219.

Rose, J. I., and V. I. Usik 2009. The "Upper Paleolithic" of South Arabia. In *The Evolution of Human Populations in Arabia Paleoenvironments, Prehistory and Genetics*, M.D. Petraglia and J. I. Rose, eds. New York: Springer.

Rosenberg, M. S., and C. D. Anderson 2011. PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods Ecol. Evol.* 2:229–232.

Seignobos, C. 2000. Mise en place du peuplement et répartition ethnique. In *Atlas de la province extreme-nord Cameroun*, C. Seignobos and O. Iyébi-Mandjek, eds. Paris: IRD MINREST/INC.

Swallow, D. M. 2003. Genetics of lactase persistence and lactose intolerance. *Annu. Rev. Genet.* 37:197–219.

Tishkoff, S. A., F. A. Reed, A. Ranciaro et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39:31–40.

Young, W. C. 1996. *The Rashaayda Bedouin: Arab Pastoralists of Eastern Sudan*. Fort Worth: Harcourt Brace College Publishers.

Zeltner, J. C. 2002. *Histoire des Arabes sur les rives du lac Tchad*. Paris: Karthala.

Table 1. Population Samples Used in this Study; Geographical Coordinates Are Approximate Estimates of the Center of the Sampled Sites

| **population** | **abbrev** | **n** | **sampling location** | **country** | **longitude** | **latitude** | **lifestyle** |
|---|---|---|---|---|---|---|---|
| Yemeni 1 | YAC | 43 | al-Akhkum | Yemen | 44.18 | 13.32 | sedentary |
| Yemeni 2 | YTI | 66 | around Hudeida | Yemen | 43.00 | 14.82 | sedentary |
| Yemeni 3 | YHG | 34 | around Hajja | Yemen | 43.60 | 15.70 | sedentary |
| Yemeni 4 | YHA | 40 | Wadi Hadramawt | Yemen | 48.74 | 15.93 | sedentary |
| Yemeni 5 | YSO | 65 | Soqotra | Yemen | 53.85 | 12.50 | sedentary |
| Arabs | | | | | | | |
| Rashaayda | RAS | 52 | Abu Talha | Sudan | 36.30 | 15.35 | nomadic |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
| Arabs Sudan | ARS | 46 | along the Nile | Sudan | 30.55 | 19.88 | sedentary |
| Arabs Baggara | ACH | 27 | around Mao | Chad | 15.31 | 14.12 | nomadic |
| Arabs Shuwa | ASW | 53 | Ngala and around | Nigeria | 14.19 | 12.34 | semi-nomadic |
| Arabs Egypt | ELH | 34 | el Hayz | Egypt | 28.66 | 28.02 | sedentary |

Table 2. Intra-population Variation of -13,915*G in the Yemeni Populations and African Arabs

| Population | TT | TG | GG | T | G | Chi sq | p | sig | Ho | He | FI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| YAC | 6 | 18 | 19 | 0.349 | 0.651 | 0.265 | 0.606 | ns | 0.419 | 0.454 | 0.079 |
| YTI | 14 | 26 | 26 | 0.409 | 0.591 | 2.263 | 0.132 | ns | 0.394 | 0.483 | 0.185 |
| YHG | 5 | 16 | 13 | 0.382 | 0.618 | 0.000 | 0.983 | ns | 0.471 | 0.472 | 0.004 |
| YHA | 20 | 19 | 1 | 0.738 | 0.263 | 2.057 | 0.151 | ns | 0.475 | 0.387 | -0.227 |
| YSO | 15 | 24 | 26 | 0.415 | 0.585 | 3.737 | 0.053 | ns | 0.369 | 0.486 | 0.240 |
| RAS | 4 | 16 | 32 | 0.231 | 0.769 | 0.924 | 0.336 | ns | 0.308 | 0.355 | 0.133 |
| ARS | 39 | 7 | 0 | 0.924 | 0.076 | 0.312 | 0.576 | ns | 0.152 | 0.141 | -0.082 |
| ACH | 8 | 13 | 6 | 0.537 | 0.463 | 0.027 | 0.869 | ns | 0.481 | 0.497 | 0.032 |
| ASW | 47 | 5 | 1 | 0.934 | 0.066 | 2.932 | 0.087 | ns | 0.094 | 0.123 | 0.235 |
| ELH | 33 | 1 | 0 | 0.985 | 0.015 | 0.008 | 0.931 | ns | 0.029 | 0.029 | -0.015 |

Chi sq – Chi-square tests for Hardy-Weinberg equilibrium; p – probability of the Chi-square tests (ns – non significant, ** P<0.01); Ho – observed heterozygosity; He – expected heterozygosity; FI – Fixation Index

Table 3. Counts in Parentheses and Frequencies of Mutations Found in the Analyzed Segment of MCM6 Gene

| **Population** | **-13,915*G** | **-13,907*G** | **-13,910*T** | **-13,965*G** | **-14,009*G** | **-14,010*C** | **-14,042*G** | **-14,107*A** |
|---|---|---|---|---|---|---|---|---|
| YAC | (56) 0.651 | | | | (1) 0.012 | (1) 0.012 | | |
| YTI | (78) 0.591 | (1) 0.015 | | | | (1) 0.008 | | |
| YHG | (42) 0.618 | | | | | | | |
| YHA | (21) 0.263 | | (3) 0.038 | | | | | |
| YSO | (76) 0.585 | | (5) 0.046 | | | | | |
| RAS | (80) 0.769 | | | | | | | |
| ARS | (7) 0.076 | (1) 0.022 | | | (11) 0.130 | | | |
| ACH | (25) 0.463 | | (1) 0.019 | | (2) 0.037 | | | |
| ASW | (7) 0.066 | | (9) 0.085 | (1) 0.009 | (1) 0.009 | | | (1) 0.009 |
| ELH | (1) 0.015 | | | | (4) 0.059 | | (1) 0.015 | |

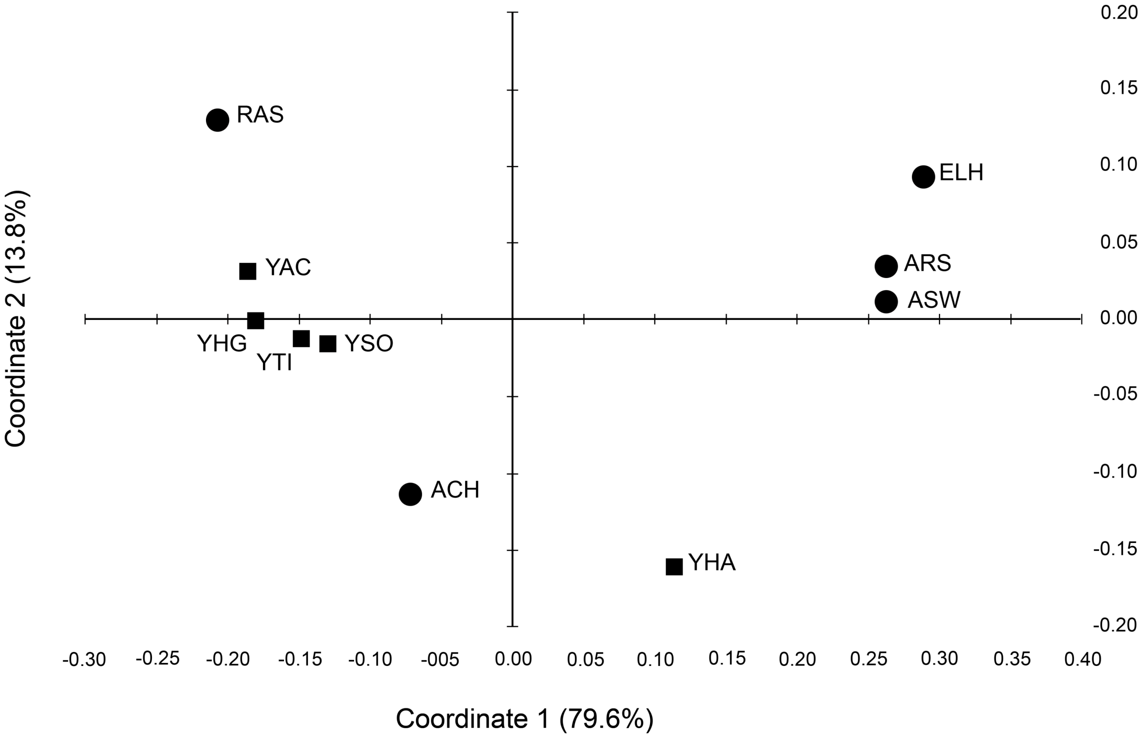

Figure 1. PCoA plot based on the frequencies of -13,915*G, -13,910*T and -14,009*G. Arabian samples are marked by squares, African samples by circles. Population abbreviations can be found in Table 1.

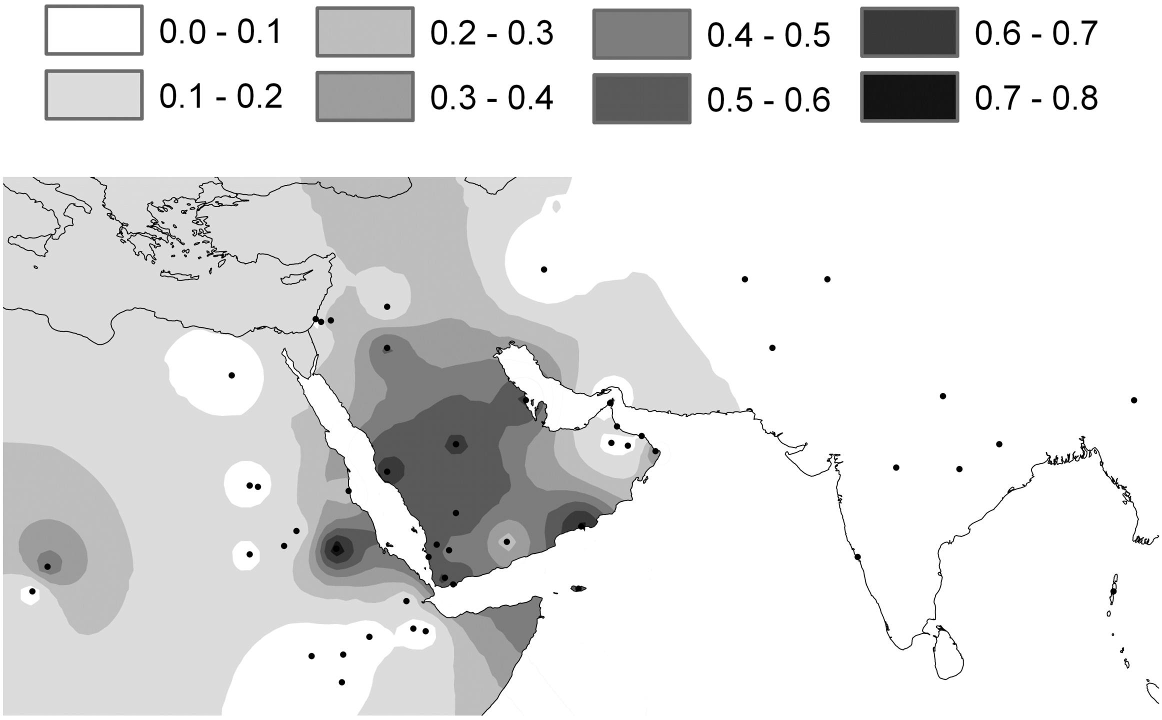

Figure 2. Interpolation map for -13,915*G frequencies in Arabia and neighboring regions. Population names, geographical coordinates and frequencies of the allele can be found in Appendix, Table 1.

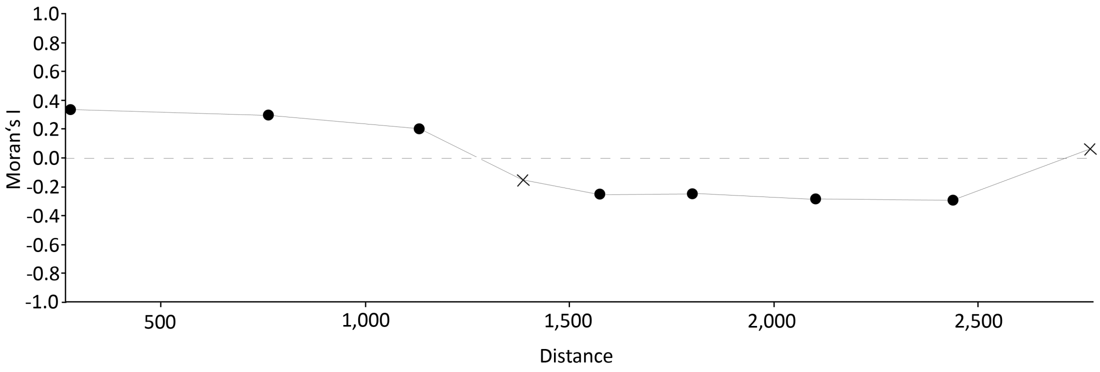

Figure 3. Spatial autocorrelation analyses for -13,915*G; black, solid points are statistically significant at the 5% level, crosslets statistically non-significant at the 5% level.

## Appendix

Appendix Table 1

| **Country** | **Abbreviation** | **Population/Region** | **Longitude** | **Latitude** | **N** | **-13915G** | **Reference** |
|---|---|---|---|---|---|---|---|
| Oman | ODA | Ad Dakhliyah | 57.50 | 22.90 | 76 | 0.06 | Al-Abri et al. 2012 |
| Oman | ODH | Al Dhahira | 56.30 | 23.10 | 47 | 0.06 | Al-Abri et al. 2012 |
| Oman | OMU | Muscat | 58.50 | 23.60 | 205 | 0.09 | Al-Abri et al. 2012 |
| Oman | OSH | Ash Sharqiyah | 59.50 | 22.50 | 28 | 0.41 | Al-Abri et al. 2012 |
| Oman | OBA | Al Batnah | 56.70 | 24.30 | 70 | 0.14 | Al-Abri et al. 2012 |
| Oman | DFR | Dhofar | 54.10 | 17.00 | 210 | 0.72 | Al-Abri et al. 2012 |
| Oman | OMS | Musandam | 56.20 | 26.00 | 2 | 0.00 | Al-Abri et al. 2012 |
| Yemen | YAD | Adan | 44.80 | 12.80 | 99 | 0.56 | Al-Abri et al. 2012 |
| Yemen | YSA | San‘a’ | 44.50 | 15.30 | 93 | 0.40 | Al-Abri et al. 2012 |
| Pakistan | PGI | Gilgit | 68.00 | 30.00 | 73 | 0.00 | Enattah et al 2008 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Pakistan | PKP | Kyber-Pakhtunkhwa | 72.00 | 35.00 | 58 | 0.00 | Enattah et al 2008 |
| Pakistan | PQU | Quetta | 66.00 | 35.00 | 92 | 0.00 | Enattah et al 2008 |
| Iran | IRA | Iranians | 51.40 | 35.70 | 42 | 0.00 | Enattah et al. 2008 |
| Jordan | JOR | Jordanian | 35.90 | 32.00 | 112 | 0.05 | Enattah et al. 2008 |
| Saudi Arabia | SAR | Arabs | 45.00 | 23.00 | 248 | 0.57 | Enattah et al. 2008 |
| Sudan | SMA | Mahas | 32.50 | 15.60 | 30 | 0.17 | Enattah et al. 2008 |
| Syria, Iraq, Lebanon | AR2 | Arabs | 40.00 | 33.00 | 40 | 0.11 | Enattah et al. 2008 |
| India | INE | East | 84.50 | 23.00 | 310 | 0.00 | Gallego Romero et al. 2012 |
| India | INW | West | 76.97 | 21.29 | 468 | 0.00 | Gallego Romero et al. 2012 |
| India | INN | North | 80.35 | 26.47 | 290 | 0.00 | Gallego Romero et al. 2012 |
| India | INC | Central | 81.63 | 21.23 | 179 | 0.00 | Gallego Romero et al. 2012 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| India | INO | North East | 94.30 | 26.15 | 139 | 0.00 | Gallego Romero et al. 2012 |
| India | INS | South | 74.18 | 14.80 | 864 | 0.00 | Gallego Romero et al. 2012 |
| India | ANN | Andaman and Nicobar Islands | 92.78 | 12.26 | 34 | 0.00 | Gallego Romero et al. 2012 |
| Saudi Arabia | SNO | Northern | 40.00 | 30.00 | 164 | 0.52 | Imtiaz et al. 2007 |
| Saudi Arabia | SSO | Southern | 45.00 | 18.00 | 184 | 0.58 | Imtiaz et al. 2007 |
| Saudi Arabia | SCE | Central | 45.00 | 23.00 | 180 | 0.61 | Imtiaz et al. 2007 |
| Saudi Arabia | SEA | Eastern | 50.10 | 26.20 | 164 | 0.62 | Imtiaz et al. 2007 |
| Saudi Arabia | SWE | Western | 40.00 | 21.00 | 172 | 0.65 | Imtiaz et al. 2007 |
| Ethiopia | SO1 | Somali | 42.80 | 9.40 | 74 | 0.04 | Ingram et al. 2009 |
| Ethiopia | SO2 | Somali | 41.90 | 9.60 | 218 | 0.05 | Ingram et al. 2009 |
| Ethiopia | AFA | Afar | 41.40 | 11.60 | 74 | 0.12 | Ingram et al. 2009 |
| Ethiopia | AMH | Amharic | 38.70 | 9.00 | 38 | 0.13 | Ingram et al. 2009 |
| Israel | AR1 | Non-Bedouin Arabs | 34.80 | 32.10 | 160 | 0.05 | Ingram et al. 2009 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Israel | DRU | Druze | 34.80 | 32.10 | 28 | 0.11 | Ingram et al. 2009 |
| Israel | BED | Bedouin | 34.80 | 32.10 | 38 | 0.13 | Ingram et al. 2009 |
| Israel/PAA | PAL | Palestinian Arabs | 35.20 | 31.90 | 36 | 0.00 | Ingram et al. 2009 |
| Jordan | JOB | Jordanian Bedouin | 35.90 | 32.00 | 46 | 0.35 | Ingram et al. 2009 |
| Sudan | SSA | Shaigi | 30.00 | 20.00 | 18 | 0.06 | Ingram et al. 2009 |
| Sudan | SJA | Jaali | 33.40 | 16.70 | 172 | 0.13 | Ingram et al. 2009 |
| Saudi Arabia | SBE | Bedouin | 45.00 | 23.00 | 94 | 0.48 | Ingram et al. 2009 |
| Sudan | SB2 | Beni Amer | 37.20 | 19.60 | 162 | 0.25 | Ingram et al. 2009 |
| Ethiopia | ET4 | Ethiopian | 36.80 | 7.70 | 146 | 0.08 | Itan et al. 2009 |
| Ethiopia | ET5 | Ethiopian | 41.40 | 11.60 | 148 | 0.19 | Itan et al. 2009 |
| Ethiopia | ET1 | Ethiopian | 34.50 | 7.60 | 120 | 0.00 | Itan et al. 2010 |
| Ethiopia | ET2 | Ethiopian | 38.70 | 9.00 | 130 | 0.02 | Itan et al. 2010 |
| Ethiopia | ET3 | Ethiopian | 36.70 | 5.70 | 132 | 0.05 | Itan et al. 2010 |
| Sudan | SSU | Sudanese | 30.00 | 15.00 | 60 | 0.07 | Itan et al. 2010 |
| Nigeria | ASW | Ngala and around | 14.19 | 12.34 | 34 | 0.03 | Present study |
| Chad | ACH | around Mao | 15.31 | 14.12 | 27 | 0.46 | Present study |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Sudan | RAS | Abu Talha | 36.30 | 15.35 | 52 | 0.77 | Present study |
| Sudan | ARS | along the Nile | 30.55 | 19.88 | 46 | 0.076 | Present study |
| Egypt | ELH | el Hayz | 28.66 | 28.02 | 34 | 0.015 | Present study |
| Yemen | YAC | al-Akhkum | 44.18 | 13.32 | 36 | 0.62 | Present study |
| Yemen | YTI | around Hudeida | 43.00 | 14.82 | 66 | 0.59 | Present study |
| Yemen | YHA | in Wadi Hadramawt | 48.74 | 15.93 | 39 | 0.26 | Present study |
| Yemen | YHG | around Hajja | 43.60 | 15.70 | 30 | 0.65 | Present study |
| Yemen | YSO | in Soqotra | 53.85 | 12.50 | 57 | 0.59 | Present study |
| Sudan | SSH | Shilook | 30.00 | 20.00 | 16 | 0.00 | Tishkoff et al. 2007 |
| Sudan | SBH | Beja (Hadandawa) | 30.00 | 20.00 | 22 | 0.09 | Tishkoff et al. 2007 |
| Sudan | SB1 | Beja (Banuamir) | 30.00 | 20.00 | 12 | 0.17 | Tishkoff et al. 2007 |