A history of male migration in and out of the Green Sahara

Yali Xue

Abstract
The Sahara was once fertile; a recent study identifies human Y-chromosomal lineages that flourished in this Green Sahara and their relation to present-day populations.

Introduction
Modern humans originated in Africa at least 300 thousand years ago (kya). The migration out of Africa (~ 60 kya) was central to human history, but the subsequent migrations and demographic events within Africa, stimulated by climate changes, developments in technology and other historical factors remain under-appreciated and under-studied. A recent study [1], focusing on the consequences of climatic changes in the Saharan region over the past 12 thousand years, fills one of the gaps in our knowledge of human migration.

Y-chromosomal lineages from the Green Sahara
The transformation of the Sahara over the past 12 thousand years has been profound: a once hospitable environment with lakes, savannahs and animals 5–12 kya (referred to as the Green Sahara or the Holocene Wet Phase) has become a hostile environment that forms the largest non-Polar desert on earth. This transition is expected to have had major consequences for the human inhabitants of the region and may have left imprints in the genomes of nearby present-day populations, but can these traces be identified?

D’Atanasio et al. [1] hypothesized that, because the Sahara has been such a strong barrier to human migration for the past 5000 years, haplogroups that are found in the current populations both to the north and to the south of the Sahara must have spread during the green phase. Previous work had identified four such haplogroups: A3-M13, E-M2, E-M78 and R-V88. (The letter prefix specifies the major Y lineage, and the suffix the particular Y-SNP that defines the lineage.) Targeted sequencing of the Y chromosomes (~ 3.3 Mb) of a set of 104 men, most of whom carried one of these four lineages, allowed D’Atanasio and colleagues to construct a dated phylogenetic tree. This tree, shown in a simplified form in Fig. 1a, revealed the coalescent time of these haplogroups: that is, the period of time that has elapsed since the most recent common ancestor of the haplogroups lived. These time points mostly lay within the 5–12 kya green period, although E-M78 showed a more complex history with some older and more diverse lineages.

Correspondence: ylx@sanger.ac.uk
Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK

© The Author(s). 2018 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
shows an additional striking feature: a very rapid expansion (visualized as lineages in the phylogeny branching so closely together that no mutations occurred between the origin of the different branches) at 5.73 (SD 0.49) kya. The authors extended their findings by genotyping 142 informative SNPs (many newly discovered in this study) in 7955 additional men from 145 African, Eurasian and African-American populations. This enabled them to better understand both the likely origins of the four haplogroups earlier than 12 kya (Fig. 1b) and the locations of the present-day descendants (Fig. 1c). E-M78 and R-V88 [3] probably originated from the north because the most closely related lineages were found there, and A-M13 and E-M2 from the south.

There are some caveats to the conclusions; for example, the authors used their own estimate of the mutation rate, which happened to be slightly slower than the commonly used one, leading to a slight difference in time estimations that need to be taken into account when comparing with other estimates. Nevertheless, the lineage distributions and their fit with the time of the Green Sahara are very robust. How do these conclusions fit within a broader context?

**Insights from the Green Saharan Y-chromosomal findings**

It is widely accepted that sub-Saharan Y chromosomes are dominated by E-M2 lineages carried by Bantu-speaking farmers as they expanded from West Africa starting < 5 kya, reaching South Africa within recent centuries [4]. The E-M2-Bantu lineages lie phylogenetically within the E-M2-Green Sahara lineage and show at least three explosive lineage expansions beginning 4.9–5.3 kya [5] (Fig. 1a). These events of E-M2-Bantu expansion are slightly later than the R-V88 expansion, and highlight the range of male
demographic changes in the mid-Holocene. North of the Sahara, in addition to the four trans-Saharan haplogroups, haplogroup E-M81 (which diverged from E-M78 ~ 13 kya) became very common in present-day populations as a result of another massive expansion ~ 2 kya [6] (Fig. 1a).

Although Y chromosomes exist within populations and so share and reflect the general history of those populations, they can sometimes show some departures from other parts of the genome that result from differences in male and female behaviors. D’Atanasio et al. [1] highlight one such contrast in their study. Present-day North African populations show substantial sub-Saharan autosomal and mtDNA genetic components ascribed to the Roman and Arab slave trades 1–2 kya [7], but carry few sub-Saharan Y lineages from this source, probably reflecting the smaller numbers of male slaves and their reduced reproductive opportunities when compared to those of female slaves. The sub-Saharan Y chromosomes in these North African populations thus originate predominantly from the earlier Green Sahara period.

In this part of Africa, the indigenous languages that are spoken belong to three of the four African linguistic families (Afro-Asiatic, Nilo-Saharan and Niger-Congo). Interestingly, these languages show non-random associations with Y lineages. For example, Chadic languages within the Afro-Asiatic family are associated with haplogroup R-V88, whereas Nilo-Saharan languages are associated with specific sublineages within A3-M13 and E-M78, further illustrating the complex human history of the region.

Perspectives
Ancient DNA studies usually show that the human past was much more complicated than can be inferred from studies of the DNA in present-day populations [8]. With the demonstration of successful recovery and analysis of ancient African DNA as old as ~ 8 kya from Malawi [9] and > 2 kya from Egypt [10], it seems that ancient DNA may be recoverable more widely from Africa, and thus could inform us more directly about Y-chromosomal history and population movements during the Green Sahara period. The current study raises many questions that remain to be addressed. For example, what are the reasons for the very rapid R-V88 expansion 5–6 kya [1] and E-M81 expansion ~ 2 kya [6], and how do these expansions fit within general worldwide patterns of male-specific expansions, which in other cases have been linked to cultural and technological changes [5]?

Abbreviation
kya: Thousand years ago

Funding
YX was supported by a grant from the Wellcome Trust (098051).