

## ARTICLE

# The emergence of Y-chromosome haplogroup J1e among Arabic-speaking populations

Jacques Chiaroni<sup>\*1</sup>, Roy J King<sup>2</sup>, Natalie M Myres<sup>3</sup>, Brenna M Henn<sup>4</sup>, Axel Ducourneau<sup>1</sup>, Michael J Mitchell<sup>5</sup>, Gilles Boetsch<sup>1</sup>, Issa Sheikha<sup>6</sup>, Alice A Lin<sup>2</sup>, Mahnoosh Nik-Ahd<sup>2</sup>, Jabeen Ahmad<sup>2</sup>, Francesca Lattanzi<sup>7</sup>, Rene J Herrera<sup>8</sup>, Muntaser E Ibrahim<sup>9</sup>, Aaron Brody<sup>10</sup>, Ornella Semino<sup>11</sup>, Toomas Kivisild<sup>12</sup> and Peter A Underhill<sup>2</sup>

**Haplogroup J1 is a prevalent Y-chromosome lineage within the Near East. We report the frequency and YSTR diversity data for its major sub-clade (J1e). The overall expansion time estimated from 453 chromosomes is 10 000 years. Moreover, the previously described J1 (DYS388=13) chromosomes, frequently found in the Caucasus and eastern Anatolian populations, were ancestral to J1e and displayed an expansion time of 9000 years. For J1e, the Zagros/Taurus mountain region displays the highest haplotype diversity, although the J1e frequency increases toward the peripheral Arabian Peninsula. The southerly pattern of decreasing expansion time estimates is consistent with the serial drift and founder effect processes. The first such migration is predicted to have occurred at the onset of the Neolithic, and accordingly J1e parallels the establishment of rain-fed agriculture and semi-nomadic herders throughout the Fertile Crescent. Subsequently, J1e lineages might have been involved in episodes of the expansion of pastoralists into arid habitats coinciding with the spread of Arabic and other Semitic-speaking populations.**

*European Journal of Human Genetics* (2010) **18**, 348–353; doi:10.1038/ejhg.2009.166; published online 14 October 2009

**Keywords:** Y-chromosome haplogroup J1e; Neolithic; Arabic languages; pastoralism

## INTRODUCTION

Tracing the origin and expansion of pastoral nomadism in the Middle East has widespread significance for understanding the development of the civilizations of the ancient Near East and the spread of the Semitic languages throughout the Levant, the Arabian Peninsula and Mesopotamia. Y-chromosome analyses of modern populations of the Middle East can contribute to the delineation of the demographic and migration processes in this region. The predominant categories of Y chromosomes in this region are varieties associated with haplogroup J-M304. This haplogroup essentially bifurcates into two main sub-clades, J1-M267 and J2-M172.<sup>1</sup>

Previous studies of J1-M267<sup>2–7</sup> have found it to occur at high frequencies among the Arabic-speaking populations of the Middle East, conventionally interpreted as reflecting the spread of Islam in the first millennium CE.<sup>8</sup> However, before the middle first millennium CE, a variety of Semitic languages were spoken throughout the Middle East. Recently, historical linguists<sup>9</sup> have constructed novel classification trees of the Semitic languages in which the first split from the root of Proto-Semitic separated into East Semitic (Akkadian, Assyrian, Babylonian and Eblaite) and West Semitic. West Semitic then partitions into Ethiopic, Modern South Arabian (spoken in areas of Oman and Yemen) and the core cluster of Central Semitic. Central Semitic would then include the languages of Yemen (Old South Arabian),

Arabic and the Northwest Semitic languages of the Levant – Ugaritic, Hebrew, Phoenician and Aramaic.<sup>9–12</sup> Not only have linguists reconstructed the phylogeny of Semitic languages, but also they have dated Proto-Semitic's age to the Chalcolithic Era, circa 5500–3500 BCE.<sup>13</sup> In addition to the common Semitic language substrate found throughout the Levant and Arabian Peninsula, recent archeological studies have shown an early presence (ca. 6000–7000 BCE) of domesticated herding in the arid steppe desert regions.<sup>14</sup>

We recently showed an inverse correlation between J1-M267 frequency and mean annual rainfall in the Middle East populations.<sup>15</sup> This finding was interpreted as a founder effect associated with small groups of Neolithic herder–hunters moving into the arid regions of the Arabian Peninsula with a pastoral economy, whereas another ancestral population with a closely associated sister clade, J2a-M410, remained mainly in the regions of the Fertile Crescent that had sufficient rainfall to support a Neolithic farming economy. Although humidity levels fluctuated during the Holocene, the present climatic regime in Arabia was established ~5000 years ago.<sup>16</sup> Marginal habitats such as desert regions that were plausibly colonized by a few founders result not only in reduced genetic diversity but may also reduce linguistic diversity as evidenced by the broad geographical footprint of the Arabic language in the arid regions of the Middle East.

<sup>1</sup>UMR6578 (Université de la Méditerranée, CNRS, EFS) Anthropologie bio-culturelle, Faculté de médecine de Marseille, Marseille, France; <sup>2</sup>Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA; <sup>3</sup>Sorenson Molecular Genealogy Foundation, Salt Lake City, UT, USA; <sup>4</sup>23andMe, Inc., Mountain View, CA, USA; <sup>5</sup>Génétique moléculaire de la spermatogénèse, Inserm UMR 910, Faculté de médecine, Marseille, France; <sup>6</sup>Blood Center, University Hospital of Damas, Damascus, Syria; <sup>7</sup>Department of Anthropology, University of California, Berkeley, CA, USA; <sup>8</sup>Department of Human and Molecular Genetics, College of Medicine, Florida International University, Miami, FL, USA; <sup>9</sup>Department of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, Sudan; <sup>10</sup>Pacific School of Religion, Berkeley, CA, USA; <sup>11</sup>Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy; <sup>12</sup>Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Cambridge, UK

\*Correspondence: Dr J Chiaroni, UMR6578 (Université de la Méditerranée CNRS, EFS) Anthropologie bio-culturelle, Faculté de médecine de Marseille, France.

Tel: +33 4911 89581; Fax: +33 4911 89598;

E-mail: jacques.chiaroni@efs.sante.fr

Received 7 May 2009; revised 21 July 2009; accepted 2 August 2009; published online 14 October 2009

Although considerable sub-haplogroup diversification has been previously described within the J2-M172 clade,<sup>17</sup> the occurrence of J1-M267 affiliated subtypes at frequencies exceeding a few percentage has not yet been reported.<sup>18</sup> Here, we present the phylogeographical and haplotype diversity data from a major sub-clade of J1-M267 that is defined by the J1e-Page08 (aka P58) SNP.<sup>19,20</sup>

Hereafter, we shall refer to this major sub-clade as simply J1e.

## MATERIALS AND METHODS

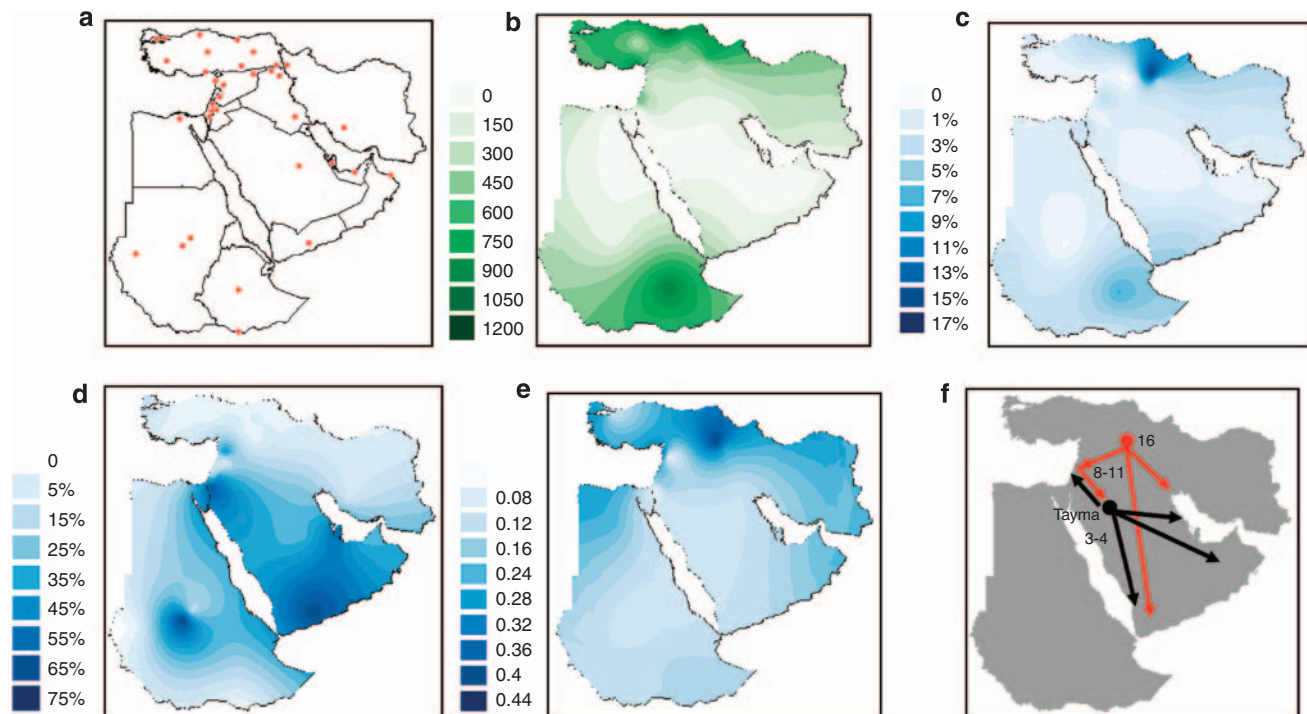
The nomenclature used for haplogroup labeling is in agreement with YCC conventions and a recent update.<sup>16</sup> All samples designated as haplogroup J1 were determined to be derived at M267. Chromosomes labeled as J1\* are J1(xJ1e). Our study involves a total of 553 haplogroup J1 samples involving 38 populations (Supplementary Table 1). These distribute to 494 J1e-derived and 59 J1\* samples. The majority of the samples were experimentally analyzed for the haplogroup J1e-defining SNP by either RFLP or DHPLC methodology, except for 55 reported as being of J1 membership from the Sudanese from Khartoum; Amhara from Addis Ababa, Ethiopia; and Iraqis from Nassiriya.<sup>18</sup> These were inferred to belong to J1e based on companion YSTR haplotype data. The criteria to deduce J1e status involved the filter of  $DYS388 \geq 15$  repeats and YCAII A, B allele sizes of either 19, 22 or 22, 22. The haplotype data used in our analyses are given in Supplementary Table 2. The following eight loci, *DYS19*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, *DYS393* and *DYS439*, were used to estimate expansion times using the methodology described by Zhivotovsky *et al.*<sup>21</sup> as modified according to Sengupta *et al.*<sup>17</sup> A microsatellite evolutionary effective mutation rate of  $6.9 \times 10^{-4}$  was used. Networks were constructed by the median joining method using Network 4.5.0.2, where  $\epsilon=0$  and microsatellite loci were weighted proportionally to the inverse of the repeat variance observed in each haplogroup.<sup>22</sup> For J1\* chromosomes, the network included the following nine loci: *DYS19*, *DYS388*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, *DYS393* and *DYS439*. With the exclusion

of *DYS388*, the same eight STR loci were also used to construct a network for haplogroup J1e-affiliated chromosomes.

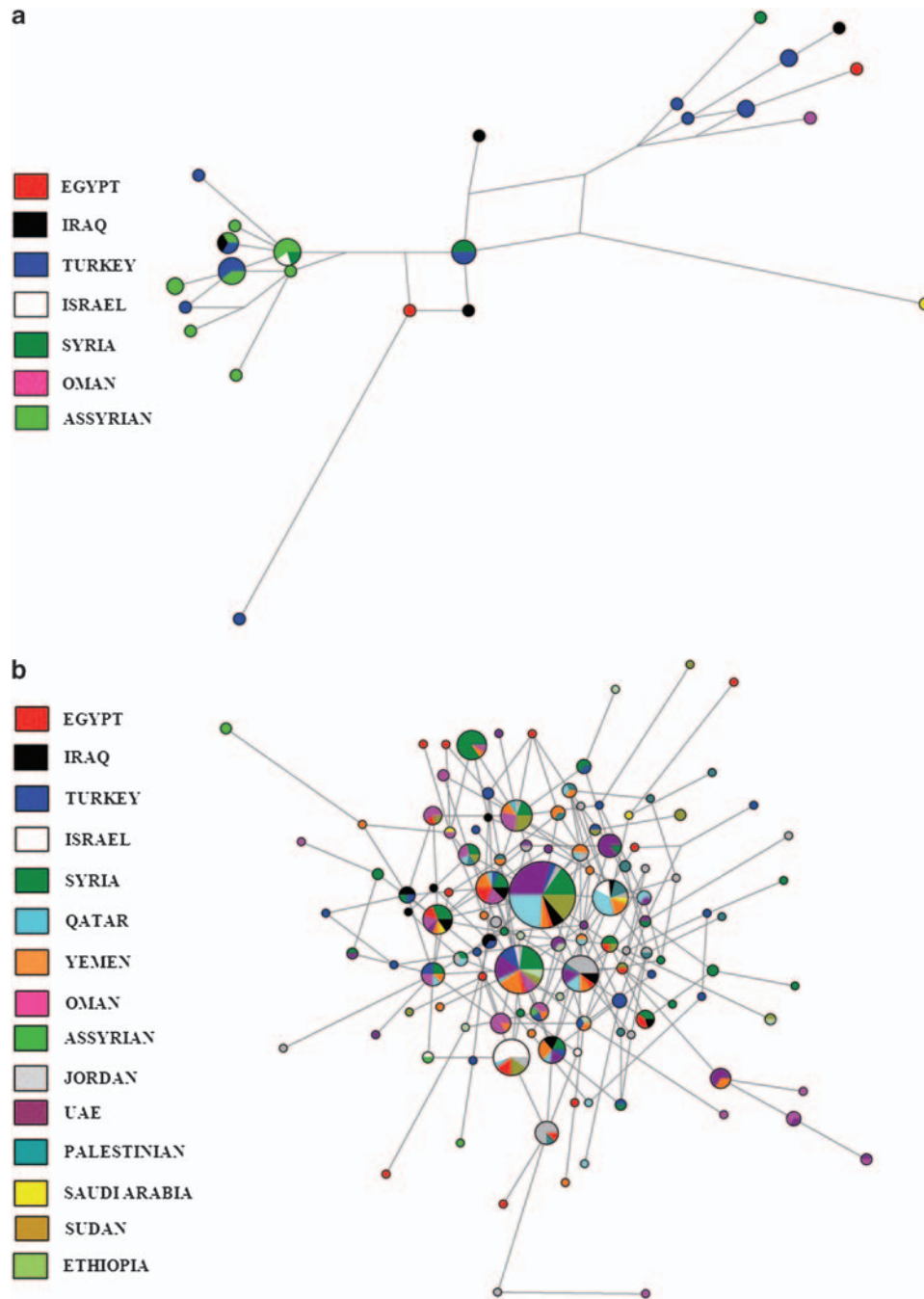
## RESULTS AND DISCUSSION

Figure 1a shows the geographical location of populations included in this study. J1\* chromosomes have their maximal frequency in the Taurus and Zagros mountain regions of Eastern Anatolia, Northern Iraq and Western Iran (Figure 1c). It is noted that the J1\* chromosomes frequently appear in combination with the 12 or 13 repeat pattern at *DYS388*, whereas the J1e chromosomes almost always display 15 or more repeats. Therefore, the J1e SNP information supports the previous inference that J1 chromosomes linked with *DYS388*=13 repeats share a common ancestry.<sup>1</sup> Network analysis of J1\* chromosomes (Figure 2a) show a bifurcating substructure. One cluster is associated with *DYS388*=15 and *DYS390* > 23 repeats and the other cluster with *DYS388*=13 repeats. The locale of highest J1\* frequency occurs in the vicinity of eastern Anatolia (Figure 1c). Both J1\* and J1e occur in Sudan and Ethiopia (Supplementary Table 1). Our data show that the YCAII 22-22 allele state is closely associated with J1e (Supplementary Table 2). Interestingly, in Ethiopia, all Cushitic Oromo and ~29% of Semitic Amharic J1 chromosomes are J1\*.

Table 1 shows the average variance and expansion times of J1e with their linguistic and archeological correlates from those populations with five or more samples; the Assyrians of Syria, Iraq, Turkey and Iran were amalgamated into one group and the Arab populations of Qatar, UAE and Saudi Arabia were also combined. The mean variance across the 19 populations in Table 1 correlates significantly with latitude ( $r=0.36$ ,  $P<0.035$ , two-tailed Kendall's  $\tau$ )



**Figure 1** (a) Red symbols indicate the geographical locations of 36 populations analyzed. (b) Interpolated spatial contours of annual precipitation (mm) distribution. (c) Interpolated J1\* frequency spatial distribution. (d) Interpolated J1e frequency spatial distribution. (e) Interpolated J1e mean haplotype variance spatial distribution. (f) Construed trajectories of J1e lineage spread episodes. In red are delineated the initial Holocene migrations from the Taurus/Zagros Mountains to the Arabian Peninsula. Shown with black arrows are the subsequent expansions of Arabic populations in Arabia beginning in the Bronze Age.



**Figure 2** (a) Median-joining network for J1\* using the nine-locus Y-STR haplotypes. Networks were weighted according to Qamar *et al.*<sup>22</sup> Loci analyzed included DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and DYS439. (b) Median-joining network for eight-locus (excluding DYS388) Y-STR haplotypes for J1e.

and nonsignificantly with longitude ( $r=0.02$ , NS). This result supports the hypothesis that the origin of J1e is likely in the more northerly populations in Table 1 and spreads southward into the Arabian Peninsula (Figure 1f). The high YSTR variance of J1e in Turks and Syrians (Table 1, Figure 1e) supports the inference of an origin of J1e in nearby eastern Anatolia. Moreover, the network analysis of J1e haplotypes (Figure 2b) shows that some of the populations with low diversity, such as Bedouins from Israel, Qatar, Sudan and UAE, are tightly clustered near high-frequency haplotypes suggesting founder effects with star burst expansion in the Arabian Desert.

The series of expansion times (Table 1) is also consistent with a subsequent Neolithic range expansion of J1e from a geographical zone, including northeast Syria, northern Iraq and eastern Turkey toward Mediterranean Anatolia, Ismaili from southern Syria, Jordan, Palestine and northern Egypt. Although there is a trend between the mean variances and the expansion time estimates, the latter do not uniformly increase with variance (Table 1) as some populations likely have more than one J1e founder. Support for this explanation involves cases in which there is the presence of two distinct varieties of YCAII chromosomes, namely, 19, 22 and 22, 22, whereas those with low

**Table 1** J1e and J1\* expansion time, mean YSTR variance based on DYS19, DYS390, DYS391, DYS392, DYS393, DYS389I, DYS389II and DYS439, linguistic and archeological correlates by population

| Population                | N   | Geography                | Annual precipitation (mm) | Mean YSTR variance | Expansion time kya | SD kya | Archeological horizon                | Language current | Language early Iron Age | Language family Iron Age  | Source  |
|---------------------------|-----|--------------------------|---------------------------|--------------------|--------------------|--------|--------------------------------------|------------------|-------------------------|---------------------------|---|
| Sunni Hama                | 16  | North Inner Syria        | 300                       | 0.1                | 4                  | 2.2    | Middle Bronze Age                    | Arabic           | Aramaic                 | NW Semitic                | This study  |
| Negev Bedouin             | 18  | Negev                    | 25                        | 0.1                | 4                  | 1.6    | Middle Bronze Age                    | Arabic           | Edomite?                | NW Semitic/C Semitic      | This study and Cann <i>et al.</i> <sup>34</sup>   |
| Turkey 9                  | 5   | Istanbul                 | 723                       | 0.14               | 5.4                | 1.8    | Early Bronze Age                     | Turkish          | Greek                   | Indo-European             | Cinnioglu <i>et al.</i> <sup>1</sup>              |
| Sudanese Arab             | 26  | Khartoum                 | 150                       | 0.14               | 5.4                | 2      | Neolithic                            | Arabic           | Old Nubian              | Nilo-Saharan              | Tofanelli <i>et al.</i> <sup>18</sup>             |
| Qatar, UAE and Saudi Arab | 102 | Arabian Peninsula        | 85                        | 0.16               | 6.3                | 1.7    | Chalcolithic                         | Arabic           | Arabic                  | C Semitic                 | This study and Cadenas <i>et al.</i> <sup>6</sup> |
| Iraqi                     | 14  | Nassiriya                | 130                       | 0.16               | 6.2                | 2.6    | Ubaid                                | Arabic           | Babylonian              | E Semitic                 | Tofanelli <i>et al.</i> <sup>18</sup>             |
| Yemen                     | 42  | Yemen                    | 170                       | 0.17               | 9                  | 3.3    | Neolithic/Arabian Bifacial Tradition | Arabic           | Old South Arabian       | C Semitic                 | Cadenas <i>et al.</i> <sup>6</sup>                |
| Amhara                    | 10  | Addis Ababa, Ethiopia    | 1100                      | 0.19               | 8.2                | 2      | Epipaleolithic/Neolithic             | Ethiopic         | Ethiopic                | W Semitic                 | Tofanelli <i>et al.</i> <sup>18</sup>             |
| Palestinian               | 16  | Central Highlands        | 600                       | 0.2                | 7.4                | 2.9    | Pottery Neolithic/Chalcolithic       | Arabic           | Hebrew/ Canaanite       | NW Semitic                | This study  |
| Jordan                    | 37  | Jordan                   | 390                       | 0.21               | 8.4                | 2.3    | Pottery Neolithic                    | Arabic           | Ammonite, Moabite       | NW Semitic                | This study  |
| Ismaili                   | 30  | South Syria              | 175                       | 0.23               | 11.3               | 6.1    | Pre-Pottery Neolithic A              | Arabic           | Aramaic                 | NW Semitic                | This study  |
| Oman                      | 45  | Oman                     | 102                       | 0.24               | 12.3               | 3.7    | Epipaleolithic                       | Arabic           | Modern South Arabian    | W Semitic                 | Luis <i>et al.</i> <sup>3</sup>                   |
| Egypt                     | 29  | N Egypt                  | 110                       | 0.25               | 9.9                | 2.4    | Epipaleolithic                       | Arabic           | Egyptian                | Egyptian (Afro-Asiatic)   | Luis <i>et al.</i> <sup>3</sup>                   |
| Iraqi Kurds1              | 7   | Northeast Iraq           | 400                       | 0.27               | 11.6               | 5.1    | Nemrikian, M'lefatian                | Kurdish/Turkish  | Kurdish/ Assyrian       | Indo-European, E Semitic  | This study  |
| Turkey 6                  | 5   | Mediterranean Anatolia   | 1000                      | 0.29               | 9.1                | 6.1    | Pre-Pottery Neolithic B              | Turkish          | Lycian, Luvian, Aramaic | Indo-European, NW Semitic | Cinnioglu <i>et al.</i> <sup>1</sup>              |
| Turkey 1                  | 7   | Marmara/Thrace           | 723                       | 0.29               | 11                 | 3.3    | Epipaleolithic                       | Turkish          | Greek                   | Indo-European             | Cinnioglu <i>et al.</i> <sup>1</sup>              |
| Turkey 4                  | 5   | East Anatolia            | 600                       | 0.34               | 11.8               | 3      | Epipaleolithic                       | Kurdish/Turkish  | Urartian                | Hurrian-Urartian          | Cinnioglu <i>et al.</i> <sup>1</sup>              |
| Alawites                  | 11  | Coastal Syria            | 395                       | 0.37               | 16.1               | 4.5    | Kebaran                              | Arabic           | Phoenician              | NW Semitic                | This study  |
| Assyrian                  | 7   | N Syria/Iraq, Van, Urmia | 400                       | 0.43               | 16.2               | 6.4    | Zarzan                               | Aramaic          | Assyrian/ Aramaic       | E Semitic, NW Semitic     | This study  |
| J1e                       | 453 |                          |                           | 0.24               | 10.1               | 2.8    | Pre-Pottery Neolithic B              |                  |                         |                           |   |
| J1e-YCA 22/22             | 203 |                          |                           | 0.22               | 9.2                | 2.9    | Pre-Pottery Neolithic B              |                  |                         |                           |   |
| J1*(DYS388=13)            | 19  |                          |                           |                    | 9.5                | 3.8    |                                      |                  |                         |                           |   |
| J1* Turks                 | 16  |                          |                           |                    | 20                 | 7.5    |                                      |                  |                         |                           |   |

mean diversity typically just reflect the 22, 22 class (Supplementary Table 2). A network analysis of J1e chromosomes (Figure 2b) also reflects situations of multiple founders.

Although the haplogroup diversification within J1e remains incomplete, the somewhat rare J1e1-M368 provides an insight into the geographical origin of J1e. It has been reported both in the Black Sea region of Turkey<sup>1</sup> and Dagestan in the northeast Caucasus.<sup>18</sup> Furthermore, J1e1-M368 displays the YCAII 19-22 pattern. Although the haplogroup relationships of YCAII alleles are unstable, nevertheless in the context of haplogroup J1, they are suggestive that the prevalent YCAII 22-22 variety may have evolved from a YCAII 19-22 ancestor.

Table 1 lists the current languages and the first millennium BCE Iron Age languages spoken in the geographical regions from which the samples were collected. Tracking back to the Iron Age, all the branches of the Central Semitic languages are represented – NW Semitic, Arabic and Old South Arabian in the Levantine and Yemeni sampling regions. The Assyrian samples and Iraqi Kurdish samples have been drawn from areas in Northern Mesopotamia speaking East Semitic languages at the time. The current data suggest an origin of J1e in the general area of eastern Turkey/northern Iraq associated with the Zarzian horizon,<sup>23–25</sup> as they have similar early pre-agricultural expansions (16 kya, Table 1).

The timing and geographical distribution of J1e is representative of a demic expansion of agriculturalists and herder–hunters from the Pre-Pottery Neolithic B to the late Neolithic era.<sup>24,26</sup> The higher variances observed in Oman, Yemen and Ethiopia suggest either sampling variability and/or demographic complexity associated with multiple founders and multiple migrations. The expansion time associated with Yemen is somewhat older (7000 BCE) and may reflect a migration of herders into southern Arabia.<sup>27</sup> Finally, the more recent expansion times (Table 1) observed in Arabs from the Arabian Peninsula, Negev Bedouins and Sunni Arabs from Hama, Syria, are consistent with a subsequent Chalcolithic/Early Bronze Age (3000–5000 BCE) advance of J1e to the Arab populations of Arabia from near the early attested Arabian-speaking area of Tayma in north central Arabia<sup>28,29</sup> (Figure 1f).

A comparison of the mean annual rainfall and spatial frequency distribution of J1e (Figures 1b and d respectively) indicates J1e peaks in the arid regions of the Arabian Peninsula. We performed a nonparametric Mann–Whitney test to address the hypothesis: is the frequency of J1e higher in arid regions ( $\leq 300$  mm) compared with regions with more rainfall in our sample set of African and Near Eastern populations? We found that the frequency of J1e was significantly greater in the arid than in the non-arid populations ( $P=0.0035$ ). By combining all the arid populations (Supplementary Table 1) into one sample ( $n=16$ ), we circumvented the details of the geographic frequency distribution, such that the J1e frequency pattern was examined primarily with regard to precipitation rather than geography, although the two are correlated.

Although most post-Last Glacial Maximum recolonization events have a typically northward signature,<sup>30,31</sup> our J1e results provide an example of a southward spread during the early Holocene. Although J1e is one of the most frequent haplogroups in the region, haplogroup E-M123 also shows its highest frequency and haplotype diversity in regions of the Fertile Crescent, decreasing toward the Arabian Peninsula.<sup>1,2,6</sup> This co-distribution pattern of Y-chromosome haplogroups J1e and E-M123 resembles mtDNA haplogroups J1b and (PreHV)1 distributions that also display low levels of diversity despite their high frequency in Saudi Arabia.<sup>32,33</sup>

Although on a broad scale the haplogroup J1e frequency distribution and expansion times are consistent with the model that it tracks a

possible expansion of Neolithic agro-pastoralists from the Fertile Crescent into the arid Arabian Peninsula, several caveats must be considered. First, the patchy distribution of J1e frequency in the Levant (Syria, Jordan, Israel and Palestine) may reflect the complex demographic dynamics of religion and ethnicity in the region. Second, even though the highest YSTR variance of J1e lineages is in eastern Anatolia, northern Iraq and northwest Iran, one cannot entirely rule out recent admixture as a contribution to the high variance among ethnic Assyrians.

A recent Bayesian analysis of Semitic languages supports an origin in the Levant 5750 years ago and subsequent arrival in the Horn of Africa from Arabia 2800 years ago,<sup>11</sup> thus providing an indirect support of our phylogenetic clock estimates. It is important to note that the glottochronological dates yield estimates for the break-up and expansion of the Proto-Semitic language. Proto-Semitic, itself, may have been spoken in a localized linguistic community for millennia before its bifurcation into the East and West Semitic branches. In summary, haplogroup J1e data suggest an advance of the Neolithic period agriculturalists/pastoralists into the arid regions of Arabia from the Fertile Crescent and support an association with a Semitic linguistic common denominator.<sup>14</sup>

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank A Silvana Santachiara-Benerecetti for providing the Ethiopian samples to OS. We thank Golden Software, Inc, for assistance with the spatial frequency maps. This work was supported by a grant from the France-Stanford Center for Interdisciplinary Studies to RK and OHLL, Origine de l'Homme des Langues et du Language to JC.

- Cinnoğlu C, King R, Kivisild T *et al*: Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 2004; **114**: 127–148.
- Semino O, Mari C, Benuzzi G *et al*: Origin, diffusion and differentiation of Y-chromosome haplogroups E and J: inferences on the Neolithization of Europe and later migratory events in the Mediterranean area. *Am J of Hum Genet* 2004; **74**: 1023–1034.
- Luis JR, Rowold DW, Regueiro M *et al*: The Levantine versus the Horn of Africa: evidence for bi-directional corridors of human migrations. *Am J of Hum Genet* 2004; **74**: 532–544.
- Di Giacomo F, Luca F, Popa LO *et al*: Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet* 2004; **115**: 357–371.
- Arredi B, Poloni ES, Paracchini S *et al*: A predominantly Neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J of Hum Genet* 2004; **75**: 338–345.
- Cadenas AM, Zhivotovskiy LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ: Y-chromosome diversity characterizes the Gulf of Oman. *Eur J Hum Genet* 2008; **16**: 374–386.
- Zalloua PA, Xue Y, Khalife J *et al*: Y-chromosomal diversity in Lebanon is structured by recent historical events. *Am J Hum Genet* 2008; **82**: 873–882.
- Nebel A, Landau-Tasserion E, Filon D, Oppenheim A, Faerman M: Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa. *Am J Hum Genet* 2002; **70**: 1594–1596.
- Huehnergard J: Comparative semitic linguistics; in Shlomo Izre'el (ed): *Semitic Linguistics: The State of the Art at the Turn of the Twenty-First Century*. Winona Lake, IN: Eisenbrauns Inc, 2002, pp 119–150.
- Fox J: Semitic Noun Patterns. *Harvard Semitic Studies* vol 52. Winona Lake, IN: Eisenbrauns Inc, 2003.
- Kitchen A, Ehret C, Assefa S, Mulligan CJ: Bayesian phylogenetic analysis of Semitic languages identifies an early Bronze Age origin of Semitic in the Near East. *Proc Biol Sci* 2009; **276**: 2703–2710.
- Rubin AD: The subgrouping of the Semitic language. *Lang Linguist Compass* 2008; **2**: 79–102.
- Diakonoff I: The Earliest Semitic Society: linguistic data. *J Semitic Studies* 1998; **43**: 209–219.
- Zarins J: Early pastoral nomadism and the settlement of lower Mesopotamia. *Bull Am Schools Orient Res* 1990; **280**: 31–65.
- Chiaroni J, King R, Underhill P: Correlation of annual precipitation with human Y-chromosome diversity and the emergence of Neolithic agricultural and pastoral economies in the Fertile Crescent. *Antiquity* 2008; **82**: 281–289.

- 16 Parker AG, Eckersley L, Smith MM *et al*: Holocene vegetation dynamics in the northeastern Rub al-Khali, Arabian Peninsula, a phytolith, pollen, and carbon isotope study. *J Quaternary Sci* 2004; **19**: 665–676.
- 17 Sengupta S, Zhivotovsky LA, King R *et al*: Polarity and temporality of high resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J of Hum Genet* 2006; **78**: 202–221.
- 18 Tofanelli S, Ferri G, Bulayeva K *et al*: J1-M267 Y lineage marks climate-driven pre-historical human displacements. *Eur J Hum Genet* 2009. doi:10.1038/ejhg.2009.58.
- 19 Repping S, Van Daalen SK, Brown LG *et al*: High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet* 2006; **38**: 463–467.
- 20 Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF: New binary polymorphisms reshape and increase resolution of the human Y-chromosomal haplogroup tree. *Genome Res* 2008; **18**: 830–838.
- 21 Zhivotovsky LA, Underhill PA, Cinnioglu C *et al*: On the effective mutation rate at Y-chromosome STRs with application to human population divergence time. *Am J Hum Genet* 2004; **74**: 54–61.
- 22 Qamar R, Ayub Q, Mohyuddin A *et al*: Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 2002; **70**: 1107–1124.
- 23 Kozłowski S: *The Eastern Wing of the Fertile Crescent: Late Prehistory of Greater Mesopotamian Lithic Industries*. Oxford: Oxford Archaeopress, 1999.
- 24 Cauvin J: *The Birth of the Gods and the Origins of Agriculture*. Cambridge: Cambridge University Press, 2000.
- 25 Bar-Yosef O: The Natufian Culture in the Levant, threshold to the origins of agriculture. *Evol Anthropol* 1998; **6**: 159–177.
- 26 Bellwood P: *First Farmers: Origins of Agricultural Societies*. Oxford: Oxford Blackwell Publishing Ltd, 2005.
- 27 McCorriston J: Breaking the Rain Barrier and the Tropical Spread of Near Eastern Agriculture into Southern Arabia; in Kennett D, Winterhalder B (eds): *Behavioral Ecology and the Transition to Agriculture*. Berkeley, CA: University of California Press, 2006, pp 217–236.
- 28 Parr PJ: s.v. 'Tayma'; in E. M. Meyers (ed): *The Oxford Encyclopedia of Archaeology in the Near East*. New York/Oxford: Oxford University Press, 1997, pp 160–161.
- 29 Byrne R: Early Assyrian contacts with Arabs and the impact on Levantine Vassal tribute. *BASOR* 2003; **331**: 11–25.
- 30 Hewitt G: The genetic legacy of the Quaternary ice ages. *Nature* 2000; **405**: 907–913.
- 31 Torroni A, Bandelt HJ, Macaulay V *et al*: A signal, from human mtDNA, of Postglacial Recolonization in Europe. *Am J Hum Genet* 2001; **69**: 844–852.
- 32 Abu-Amero KK, González AM, Larruga JM, Bosley TM, Cabrera VM: Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. *BMC Evol Biol* 2007; **7**: 32–47.
- 33 Abu-Amero KK, Larruga JM, Cabrera VM, González AM: Mitochondrial DNA structure in the Arabian Peninsula. *BMC Evol Biol* 2008; **8**: 45–60.
- 34 Cann HM, de Toma C, Cazes L *et al*: A human genome diversity cell panel. *Science* 2002; **296**: 261–262.
- 35 Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ: Iran: tricontinental nexus for Y-chromosome driven migration. *Hum Hered* 2006; **61**: 132–143.
- 36 Shlush LI, Behar DM, Yudkovsky G *et al*: The Druze: a population genetic refugium of the Near East. *PLoS One* 2008; **3**: e2105.
- 37 Hassan HY, Underhill PA, Cavalli-Sforza LL, Ibrahim ME: Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography and history. *Am J Phys Anthro* 2008; **137**: 316–323.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)