

Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs

Yahya M. Khubrani^{1,2}, Jon H. Wetton^{†1}, Mark A. Jobling^{†1}

¹ Department of Genetics & Genome Biology, University of Leicester, University Road, Leicester

² Forensic Genetics Laboratory, General Administration of Criminal Evidence, Public Security, Ministry of Interior, Kingdom of Saudi Arabia

[†] Correspondence can be addressed to either author:

Dr Jon H. Wetton, Department of Genetics & Genome Biology, University of Leicester, University Road, Leicester LE1 7RH, UK

Tel.: +44 (0)116 252 3377 Fax: +44 (0)116 252 3378

Email: jw418@le.ac.uk

Prof Mark A. Jobling, Department of Genetics & Genome Biology, University of Leicester, University Road, Leicester LE1 7RH, UK

Tel.: +44 (0)116 252 3427 Fax: +44 (0)116 252 3378

Email: maj4@le.ac.uk

Running head: Saudi Arabian Y-chromosome diversity

Abstract

Saudi Arabia's indigenous population is organized into patrilineal descent groups, but to date, little has been done to characterize its population structure, in particular with respect to the male-specific region of the Y chromosome. We have used the 27-STR Yfiler® Plus kit to generate haplotypes in 597 unrelated Saudi males, classified into five geographical regions (North, South, Central, East and West). Overall, Yfiler® Plus provides a good discrimination capacity of 95.3%, but this is greatly reduced (74.7%) when considering the reduced Yfiler® set of 17 Y-STRs, justifying the use of the expanded set of markers in this population. Comparison of the five geographical divisions reveals striking differences, with low diversity and similar haplotype spectra in the Central and Northern regions, and high diversity and similar haplotype spectra in the East and West. These patterns likely reflect the geographical isolation of the desert heartland of the peninsula, and the proximity to the sea of the Eastern and Western areas, and consequent historical immigration. We predicted haplogroups from Y-STR haplotypes, testing the performance of prediction by using a large independent set of Saudi Arabian Y-STR+Y-SNP data. Prediction indicated predominance (71%) of haplogroup J1, which was significantly more common in Central, Northern and Southern groups than in East and West, and formed a star-like expansion cluster in a median-joining network with an estimated age of ~2800 years. Most of our 597 participants were sampled within Saudi Arabia itself, but ~16% were sampled in the UK. Despite matching these two groups by home sub-region, we observed significant differences in haplotype and predicted haplogroup constitutions overall, and for most sub-regions individually. This suggests social structure influencing the probability of leaving Saudi Arabia, correlated with different Y-chromosome compositions. The UK-recruited sample is an inappropriate proxy for Saudi Arabia generally, and caution is needed when considering expatriate groups as representative of country of origin. Our study shows the importance of geographical and social structuring that may affect the utility of forensic databases and the interpretation of Y-STR profiles.

Keywords:

Saudi Arabia; Y-STRs; haplogroup; haplotype; population structure

Introduction

Saudi Arabia is the largest country in the Arabian Peninsula. Its population of ~32 million people is distributed highly non-uniformly (Figure 1), with very low densities in its large desert areas, but high densities concentrated around a small number of cities. Its indigenous Arab people (~63% of the population; www.stats.gov.sa, accessed 12/07/17) are historically organized into geographically-differentiated patrilineal descent groups, or tribes [1], with a tradition of consanguinity [2]. This geographical and social organization might be expected to have an effect on patterns of genetic diversity, particularly regarding the male-specific region of the Y chromosome (MSY), which in turn could have implications in interpretation of DNA profiles.

Genetic studies on Saudi Arabia to date are limited. Exome sequencing of a set of samples from the Arabian Peninsula including Saudi individuals demonstrated relatively high inbreeding coefficients [3], consistent with a history of consanguineous marriage. A general analysis of Saudi Arabian mitochondrial DNA (mtDNA) diversity [4] showed a pattern of haplogroups similar to that of other Arabian Peninsula samples. In another mtDNA-based study [5] - the only example to divide Saudi Arabia sub-regionally - central, northern, western and southeastern sub-groups formed a single cluster in a multi-dimensional scaling (MDS) analysis when compared to other Arabian Peninsula samples, but also presented significant inter-group differences. Y-chromosome studies have analysed the seven Y-STRs defining the minimal haplotype [6], or haplogroup-defining SNPs together with 17 Y-STRs (Yfiler®) for one specific haplogroup [7]. The first of these [6] revealed lower diversity in Saudi Arabia than in populations from outside the Arabian Peninsula, and affinity between Saudi Arabia and Yemen, which together were strongly differentiated from Oman and Dubai. It was speculated that this might be due to the influence of patrilineal descent and polygyny. The second study [7] showed that haplogroup J1 was the most prominent lineage (42%) in the Saudi Arabian sample studied, and that genetic distances based on haplogroup frequencies were relatively small among Arabian Peninsula samples. The focus of Y-STR typing on one lineage precludes any population-based conclusions on haplotype

diversity from this study.

To date, therefore, while some general studies have been carried out, little has been done to characterize population structure within Saudi Arabia. Knowledge of any such structure is important in the interpretation of the significance of DNA-based forensic evidence, and in the construction of appropriate databases. Here, we use the 27 Y-chromosomal short-tandem repeats (Y-STRs) in the Yfiler® Plus kit to characterize haplotypes in 597 Saudi males sub-divided by geographical region. We consider the relationships of Y-chromosome diversity between regions within the country and also between Saudi Arabia and other surrounding populations. Finally, we compare the spectrum of Y-chromosome types in males recruited within Saudi Arabia with that of regionally-matched males recruited in the United Kingdom, to ask if social structuring also influences patterns of Y-haplotype diversity.

Materials and Methods

DNA sampling

Five hundred and ninety-seven DNA samples were collected from indigenous Saudi Arabian males who were ethnically and linguistically Arabic. Of these, 503 were extracted from blood spots on FTA cards (Whatman, UK), sampled from individuals recruited within Saudi Arabia itself. The remaining 94 were extracted from buccal swabs [8], or from saliva samples via the Oragene kit (DNA Genotek), from Saudi males resident within the UK. In each case, males with ancestry (to the level of paternal great-grandfather) from five geographical subdivisions of the country shown in Figure 1 (Central, Northern, Southern, Eastern, and Western) were sampled, and consideration of relatedness ensured that all sampled males were separated by at least three generations. Ethical review for recruitment and analysis was provided by the Saudi General Administration for Forensic Evidence and the University of Leicester Research Ethics Committee. Informed consent was provided by all participants.

DNA Extraction and Quantification

DNAs were extracted and purified from FTA blood-spot samples using a fully automated STARlet workstation (Hamilton) and the PrepFiler® Forensic DNA Extraction Kit (Thermo Fisher Scientific), starting from 1.2-mm diameter punches produced using the BSD100 Punching System (Microelectronic Systems). Buccal samples were extracted via QIAamp DNA Mini Kits on a QIAcube robotic workstation (Qiagen). All DNA samples were quantified using the Quantifiler® Human DNA Quantification Kit (Thermo Fisher Scientific) on an Applied Biosystems® 7500 Real-Time PCR System.

DNA amplification and fragment detection

The Yfiler® Plus PCR Amplification Kit was used to generate Y-chromosome haplotypes for the 27 STRs DYS19, DYS385a, DYS385b, DYF387S1a, DYF387S1b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS449, DYS456, DYS458, DYS460, DYS481, DYS518, DYS533, DYS570, DYS576, DYS627, DYS635, and Y-GATA H4. PCRs were conducted as recommended by the manufacturer on a Veriti (Thermo Fisher Scientific). Fragments were detected using an ABI3500 or ABI3130xl Genetic Analyzer (Thermo Fisher Scientific) using the manufacturer's recommended protocols. GeneMapper IDX software V1.4 was used for allele calling and interpretation.

Haplogroup prediction and assessment of accuracy

Y-STR haplotypes were predicted from Y-STR haplotypes using the online Y-DNA Haplogroup Predictor NevGen (<http://www.nevgen.org>), which is based on a previously-implemented Bayesian approach [9], with the additional consideration of pairwise correlation of alleles between different Y-STRs in the calculation of haplogroup probabilities.

We assessed the accuracy of this method by using a dataset of Y-STR and Y-SNP haplotypes from 743 self-declared Saudi Arabian males downloaded from FamilyTreeDNA (www.familytreedna.com). The NevGen method requires input of a standard set of Y-STRs, but as yet this does not include the Yfiler® Plus set. We therefore input the PowerPlex Y23 (PPY23) set of 23 Y-STRs with the two

markers DYS549 and DYS643 marked as missing data, since these are present in PPY23 but not in Yfiler® Plus. Y-STR resolution varied in the FamilyTreeDNA dataset, depending on whether custom Y-STR typing or Y-chromosome resequencing had been carried out, and we standardized the level of haplogroup definition to a broad resolution of the thirteen haplogroups, A, B, E1b1b, E1b1a, G, H, J1, J2, L, Q, R1a, R1b and T (nomenclature according to [10]). For the set of 743 FamilyTreeDNA Y chromosomes, NevGen predicted a compatible haplogroup in 738 cases (99.3%). The five incompatible predictions all involved mis-prediction of haplogroup E sublineages as either hg I2a2a, or D1a. In applying NevGen to our own dataset, we therefore accepted all predictions except those for hgs I (n=5) and D (n=1), and in addition rejected predictions of haplogroups (hg N; n=1, and hg O; n=2) for which examples were not found in the FamilyTreeDNA dataset. We defined this total set of nine haplotypes as ‘unpredicted’. To further understand the relationship between Y-STR haplotypes and predicted haplogroups, we also constructed a median-joining network combining haplotypes both from our dataset and from the FamilyTreeDNA dataset (Figure S1); this demonstrates coherence of haplogroup prediction and haplotype clustering.

Median-joining networks

Median-joining networks [11] were constructed using the software Network 5.0 and Network Publisher (<http://www.fluxus-engineering.com/sharenet.htm>). In the case of intermediate alleles, repeat numbers were rounded to the nearest integer; constitutively duplicated loci (DYS385a,b; DYF387S1 a/b) were removed for network construction. Deletion alleles were coded ‘99’ in input files, and thereby considered as missing data.

Forensic and population genetic parameters

For each sample or sub-sample, haplotype diversity was calculated using the formula $n(1-\sum p_i^2)/(n-1)$, where n is the sample size and p_i the frequency of the i th haplotype. Haplotype match probability (HMP) was estimated as the sum of squares of the haplotype frequencies. Discrimination capacity (DC) was calculated as the ratio between the number of distinct haplotypes and the total

number of haplotypes in the sample.

Rst calculations based on Y-STR data and multi-dimensional scaling (MDS) plots [12] were carried out using comparative population data and the calculation tool within the online Y Haplotype Reference Database ([13]; <https://yhrd.org/amova>). The square plots produced by this approach were graphically adjusted to Euclidean space for display purposes.

Population differentiation tests based on predicted haplogroup frequencies were carried out within Arlequin [14] based on predicted haplogroups, using a method analogous to Fisher's exact test. Haplogroup-based gene diversity was also calculated in Arlequin.

Based on published comparisons of the performance of sequence- and STR-based dating [15], time-to-most-recent-common-ancestor (TMRCA) was estimated by the average-squared distance (ASD) method [16, 17], using 23 STRs omitting the duplicated STRs (DYS385a,b; DYF387S1 a/b), and also using a reduced set of 18 omitting the Rapidly Mutating STRs (DYS449, DYS518, DYS570, DYS576 & DYS627). We used the modal haplotype as a root, and the mean pedigree mutation rate across STRs as measured in father-son pairs and available from yhrd.org/pages/resources/mutation_rates.

Results

The 27 Y-STRs contained in the Yfiler® Plus kit were amplified in DNAs from 597 Saudi Arabian males. Table S1 contains a full list of haplotypes, as well as other sample information; data are also available from YHRD, release 55 (YA004270 - Central; YA004271 – East; YA004272 – North; YA004273 – South; YA004274 – West). We also predicted haplogroups from each STR haplotype, using the prediction tool NevGen, and tested prediction accuracy based on a large independent set of Y-STR data and known Y-SNP haplogroups, as described in Materials and Methods.

Y-STR allele and haplotype diversity within Saudi Arabia

Considering allelic diversity, our dataset is characterized by a very high proportion of individuals (424/597; 71%) carrying intermediate alleles, in particular .2 alleles at DYS458; this is a known characteristic of haplogroup J1 [18], and immediately suggests that this haplogroup strongly predominates in the sample. In addition, twenty-four haplotypes carry .2 alleles at DYS627, and 22 of these are also predicted to belong to hg J1. The allele 23.2, present in a single individual of ‘unpredicted’ haplogroup, is not yet catalogued in YHRD (Release 54, June 2017). One copy of the duplicated STR DYF387S1 a/b carries a .2 allele in 14 haplotypes, and of these haplotypes, ten are predicted to belong to the generally rare hg B, and are therefore likely identical by descent - these have been observed previously in the same haplogroup [19]. Sporadic examples of other intermediate alleles are observed at DYS390, DYS392, DYS448, DYS449 (including allele 31.2, not listed in YHRD) and DYS570.

We observe one allele duplication among our data - a tri-allelic pattern at DYF387S1 a/b. We also observe one example of a deleted allele at DYS448 in a haplotype predicted to belong to haplogroup J1. DYS448 deletions have been described previously [20], though not (to our knowledge) in this haplogroup; deletions are recurrent, and driven by unequal recombination between flanking large repeats. Deletion is also observed in a single case for both copies of DYF387S1 a/b; five such cases are included in the 18,921 haplotypes in the YHRD.

Turning to haplotypes, Table 1 lists diversity summary statistics for the whole dataset, and for the five geographical subdivisions. The 597 males carry 543 distinct haplotypes, including 25 identical pairs, and two trios, providing a discrimination capacity of 95.3%. However, when we consider the sub-set of 17 Y-STRs making up the Yfiler® haplotype, we see a much higher level of haplotype sharing in the dataset: one haplotype, for example, is represented 22 times, and the discrimination capacity is only 74.7%; this compares, for example, with discrimination capacities of 98.5% and 95.7% for Yfiler® Plus and Yfiler® respectively in a US Caucasian sample [21].

Considering the five geographical subdivisions, comparisons of discrimination capacity at the levels of Yfiler® Plus and Yfiler® reveal striking differences in diversity between regions (Table 1). Values for Yfiler® Plus range from ~94% to 100%, but the range of values for Yfiler® haplotypes is much broader; in particular, the Central sample shows a discrimination capacity of only 72%, while the corresponding value for the Eastern sample is over 95%. In general, this points to relatively low diversity in the Central and Northern samples, with relatively high diversity in the East and West.

We compared our total Saudi Arabian population sample with other samples from the Arabian Peninsula, using multi-dimensional scaling based on Rst distances calculated from Y-STR haplotypes (Figure 2a). In the first dimension of the plot, Iraq and Qatar lie at the extremes, with a cluster of other populations between them. Our Saudi Arabian sample (KSA) lies midway between this cluster and Qatar, and close to a previously published Saudi Arabian sample [6]. However, when we subdivide our sample into its five geographically-defined subsamples, two of these (Northern and Central) overlap with each other and cluster with Qatar, whereas the Eastern and Western subsamples overlap with each other and show affiliation with the major cluster of populations in the middle of the plot (Jordan, Kuwait, Oman, UAE and Yemen). The Southern sample lies at a similar first-dimension position to these, but is shifted in the second dimension of the plot, suggesting a distinct haplotype distribution to that of the Eastern and Western samples. Inclusion of population samples [19, 22-26] from a wider surrounding geographical region (Figure S2) does not change these relationships substantively.

Analysis of diversity via network analysis and haplogroup prediction

In order to understand the relationships between Y-STR haplotypes in the dataset, we constructed a median-joining network (Figure 3a). Based on NevGen predictions (Materials & Methods), haplogroups were assigned to haplotypes within the network (Table S1). Most predicted haplogroups form coherent clusters, with the exception of haplogroup E1b1b, which forms two well-separated clusters, possibly indicating distinct sub-lineages that cannot be reliably distinguished by the prediction method used; the same split of haplogroup E1b1b is seen in a network containing the FamilyTreeDNA haplotypes of known haplogroup (Figure S1). Network sub-structures for most haplogroups are generally extended, although a cluster of related haplotypes exists among the predicted haplogroup E1b1b chromosomes. However, the network's major feature is a central star-like cluster of closely related haplotypes assigned to haplogroup J1 (71% of the total sample), suggesting a recent expansion for this set of lineages. We estimated the TMRCA of this cluster using the average-squared distance method and a mean pedigree-based STR mutation rate. Considering the total set of 23 non-duplicated Y-STRs, this yields a TMRCA of 2494 ± 487 years; removal of rapidly-mutating markers, which might be expected to bias the estimate, reduces the number of STRs to 18 and increases the age slightly to 2754 ± 389 years. Application of the same methods to the FamilyTreeDNA dataset of confirmed haplogroup J1 haplotypes yields very similar estimates, for example 2783 ± 394 years for the reduced set of 18 Y-STRs. It is worth noting that use of the so-called 'evolutionary' average mutation rate of 6.9×10^{-4} per STR per generation [27] yields greatly elevated and very divergent TMRCA estimates for 23 and 18 Y-STRs in our own dataset of $19,835 \pm 3874$ years and 9809 ± 1916 years respectively.

Figure S3 shows the same network, but with haplotypes coloured by region of origin. There is little evidence from inspecting this network of geographical substructuring, although the haplogroup E1b1b sub-cluster mentioned above is mostly formed by Western samples. Table 2 and Figure 3b present predicted haplogroup distributions for the geographically defined sub-

samples. One striking feature is the difference in the frequency of predicted haplogroup J1 in the Northern+Central+Southern samples (93% collectively) than that in the Eastern+Western pair (50%; exact test $p<0.001$); on the other hand, the latter pair has a significantly higher frequency of predicted haplogroup E1b1b (19% vs 6%; exact test $p<0.001$). Considering gene diversity values from haplogroup frequencies (Table 2), the Northern+Central pair shows significantly lower diversity than the Southern sample, which in turn is significantly lower than the Eastern+Western pair ($p<0.05$).

Comparison of cohorts recruited in Saudi Arabia and the UK

Of our 597 samples, 94 (~16%) were recruited not in Saudi Arabia itself, but in the UK. To ask whether place of recruitment influenced the spectrum of haplotypes observed, we compared Y-STR haplotype diversities in the two differently recruited samples (Table S2), considering the same parameters as in Table 1. In the Saudi-recruited sample, both the proportion of unique haplotypes (~90%) and the discrimination capacity (~95%) are similar to the corresponding values in the cohort as a whole (~91% and ~95%). However, the UK-recruited sample shows much higher values - ~98% and 99% respectively, indicating that this mode of recruitment is sampling a more diverse subset of the Saudi Arabian population, despite the two groups being geographically matched. Predicted haplogroup frequencies in the Saudi- and UK-recruited samples (Figure 4a) are significantly different (exact test $p<0.001$), including a much higher frequency of predicted haplogroup J1 in the former. Similar comparisons at the sub-regional level (Figure 4a) show significant differences between predicted haplogroup frequencies (Table S3) for the Saudi- and UK-recruited samples from Central, Western, and Southern regions ($p\leq0.02$). For the North, the difference is not significant, probably due to the small sample size ($n=10$) of the UK-recruited sample. For the East, however, the corresponding sample size is larger ($n=17$), and the lack of significant difference ($p=0.795$) probably indicates true homogeneity of the Saudi- and UK-recruited samples for this region.

Discussion

In this study, we have determined the Yfiler® Plus haplotypes of a set of 597 Saudi Arabian males, and also considered how haplotype composition is affected by division into five geographically-defined sub-groups, and by two different countries of recruitment (Saudi Arabia itself, and the United Kingdom).

The Yfiler® Plus system provides a discrimination capacity of 95.3% in the overall sample, which, while lower than that for US Caucasian, US Hispanic and African-American samples [21], exceeds that for an US Asian sample (94.4%). However, the added value of using the extended set of 27 Y-STRs contained in Yfiler® Plus is clearly demonstrated by a comparison with the 17 STRs defining the Yfiler® system – discrimination capacity in the Saudi Arabian sample falls much more markedly than in the US samples, to only 74.7%. This suggests that, despite our care in avoiding related males, there are many individuals in the sample whose haplotypes are similar because of deeper patrilineal descent from shared ancestors. This probably reflects a general property of many Middle Eastern populations: of all global regions, the Middle East was previously shown to exhibit the greatest difference between diversity assessed by Yfiler® STRs and RM-YSTRs [28], and the lowest Yfiler® discrimination capacity (~84%). We note that the benefits of using Yfiler Plus compared to Yfiler have also been reported in low-diversity patrilineal groups in East Africa [19].

As well as determining Y-STR haplotypes, we have predicted Y-SNP haplogroups from these haplotypes. A number of prediction methods exist, taking different approaches including phylogenetic trees with STR mutation rates (YPredictor; <http://predictor.ydna.ru/>), machine learning [29], partitional clustering [30], simple allele frequencies [31], and Bayesian allele frequencies ([9]; NevGen: <http://www.nevgen.org>). Evaluating prediction methods is not straightforward because some have been produced by the genetic genealogy community and are therefore not published in mainstream journals [9, 31] or peer reviewed (YPredictor; NevGen); exact methodology is sometimes unclear. There has been debate about the accuracy of prediction; for example, criticism [32] of one widely used method [9], was counter-criticized [33] for using only 7

Y-STRs in evaluation. It seems clear that the larger the number of STRs, the better, and here we have used 21 STRs from the Yfiler® Plus set. In addition, any prediction method is only as good as the Y-SNP+Y-STR comparative datasets it uses for training or classification, and sometimes these are not well described – comparative datasets that are too small, or that do not include populations appropriate to the sample being predicted, may give unreliable results. In order to address this problem, we used a large set of independent Y-SNP+Y-STR data from the same population (Saudi Arabia) as that under study to test prediction performance. The chosen method, NevGen, performs well, and provides a >99% accuracy in our sample, but we recognize the need to undertake SNP typing for definitive haplogroup determination [34].

The median-joining network of haplotypes (Figure 3a) exhibits a large central star-like cluster that corresponds to predicted haplogroup J1, and contains many of the identical or highly similar haplotypes. Such features are commonly interpreted as past male-lineage expansions [35]. Star-like features of haplotypes comprising haplogroup J1 have been reported before in specific Arabian populations [36] and in broader Middle Eastern samples [37]. Interpretations of its origins initially focused on the 7th-century Muslim expansion [38], and were supported by some later studies [39], but some other authors have interpreted it in terms of much earlier spread in the Neolithic [37, 40] followed by Bronze Age expansion in the Arabian Peninsula [37]. The age of the expansion is clearly crucial, and this in turn is affected by the method chosen, but most strongly by the choice of mutation rate. The mean ‘evolutionary’ rate of 6.9×10^{-4} mutations per generation [27] has been widely used [7, 37, 40], and has been reported as performing better the directly-determined ‘pedigree’ rate for the dating of ancient events such as the coalescence of the whole Y phylogeny [15, 41]. However, this rate was estimated from a relatively small set of 7-10 STRs [27], not including RM-YSTRs, so certainly cannot be universally applied. Furthermore, for haplogroups showing star-like patterns in networks, and for which Y-chromosome resequencing data indicate recent TMRCAAs (<10 thousand years), the pedigree mutation rate has been shown to perform best [15]. We therefore used this rate, and obtain a TMRCA for predicted haplogroup J1 of around 2800 years. This is several-fold younger than published estimates [7, 37],

40], and if correct, suggests late Bronze Age dispersion, possibly followed by later spread during the Islamic expansion.

Aside from the dominant predicted haplogroup J1, the range of other predicted lineages (Table 2, Table S3) is similar to that seen in a previous Y-SNP-based study of Saudi Arabia [7]. However, the frequency of these haplogroups differs very significantly between our sample and the published sample [7] (Figure 4b, Table S3). This suggests that there is considerable Saudi Arabian heterogeneity, and that the sub-populations sampled in these two studies are very different. Such heterogeneity is confirmed when we subdivide our sample into five geographical regions. In both Y-STR- (Figure 2b, Table 1) and predicted-haplogroup-based (Figure 3b, Table 2) comparisons, the Central and Northern regions are highly similar in composition, and, surprisingly, the Eastern and Western samples are also highly similar. The Central + Northern pair is highly diverged from the Eastern + Western pair. The Southern region is somewhat distinct from all other regions with respect to both its spectrum of predicted haplogroups (Figure 3b, Table 2) and Y-STR haplotypes (Figure 2b). Considering discrimination capacity at the level of Yfiler® (Table 1), the Central and Northern samples show similarly low values, with higher and increasing values in the Southern, Western and Eastern samples respectively. Similar results are shown by comparing predicted haplogroup distributions (Table 2). The low diversity and similarity of Central and Northern areas reflect their relative geographical isolation within the desert heartland of the country, and possible bottleneck associated with the onset of desertification around 3000 years ago [42]. By contrast, the relatively high diversity of the Eastern and Western areas reflects their closeness to the sea and outside influences from other populations that may historically have brought in migrants.

While most of our 597 participants were contacted and recruited within Saudi Arabia itself, about 16% were recruited in the UK. Despite attempting to match the two groups by geographical sub-region, there are significant differences in the haplotype and predicted haplogroup constitutions of these two groups overall (Figure 4a, Table S2, Table S3), and for the Central, Southern and Western sub-groups in particular. The UK-recruited sample size for the Northern region is very small so it is hard to draw any conclusion, while for the Eastern

region the two differently-recruited sub-samples seem genuinely similar in composition. Taken together, this suggests that there is social structure in the country, which influences the probability of males leaving Saudi Arabia and undertaking study in the UK, and that this structure correlates with different sub-groups having different haplotype compositions. This social structuring appears to be less marked in the East of the country than elsewhere. It means that the UK-recruited sample is an inappropriate proxy for Saudi Arabia generally, and indicates that caution is needed when considering expatriate groups as representative of their country of origin.

The strong geographical and social structure we have observed in Saudi Arabia has important implications for the interpretation of Y-STR profiles in casework. Geographically appropriate databases must be used for assessment of evidential weight, and more work should be done to understand the social structuring reflected in the two differently recruited cohorts. Given the patrilineal descent structures of the tribal system, analysis of tribal names and surnames together with Y haplotypes should be illuminating. Other tribally-based Middle Eastern countries may also show marked population structure.

It will also be important to ask whether autosomal STR diversity is affected by the same factors that give rise to low diversity and a high degree of population structure among Y-haplotypes in Saudi Arabia. Sex-bias in population structure could also be addressed using maternally-inherited mtDNA. Indeed, inspection of a published study in which the country is subdivided [5] shows that the mtDNA haplogroup spectrum of the Central region differs significantly from those of other regions except the North ($p \leq 0.05$, Bonferroni-corrected; our analysis). This indication of structuring among maternal lineages, as well as paternal lineages, suggests that further analysis in our samples would be worthwhile.

Acknowledgments

YMK was supported by the Saudi Arabian Ministry of Interior, and by a PhD studentship grant from the Saudi Arabian Cultural Bureau, London. We thank members of the Forensic Genetics Laboratory, General Administration of

Criminal Evidence, Riyadh, for technical assistance. Milos Cetkovic Gentula and Aco Nevski of NevGen for running batch haplogroup predictions for us, and Jon Kyte and Thermo Fisher for their support.

Conflicts of interest

None

References

- [1] M. Al-Rasheed, A History of Saudi Arabia, Cambridge University Press, Cambridge, 2010.
- [2] L. Al-Gazali, H. Hamamy, S. Al-Arrayad, Genetic disorders in the Arab world, *BMJ* 333 (2006) 831-4.
- [3] E.M. Scott, A. Halees, Y. Itan, E.G. Spencer, Y. He, M.A. Azab, S.B. Gabriel, A. Belkadi, B. Boisson, L. Abel, A.G. Clark, C. Greater Middle East Variome, F.S. Alkuraya, J.L. Casanova, J.G. Gleeson, Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery, *Nat. Genet.* 48 (2016) 1071-6.
- [4] K.K. Abu-Amero, A.M. Gonzalez, J.M. Larruga, T.M. Bosley, V.M. Cabrera, Eurasian and African mitochondrial DNA influences in the Saudi Arabian population, *BMC Evol. Biol.* 7 (2007) 32.
- [5] K.K. Abu-Amero, J.M. Larruga, V.M. Cabrera, A.M. Gonzalez, Mitochondrial DNA structure in the Arabian Peninsula, *BMC Evol. Biol.* 8 (2008) 45.
- [6] F. Alshamali, L. Pereira, B. Budowle, E.S. Poloni, M. Currat, Local population structure in Arabian Peninsula revealed by Y-STR diversity, *Hum. Hered.* 68 (2009) 45-54.
- [7] K.K. Abu-Amero, A. Hellani, A.M. Gonzalez, J.M. Larruga, V.M. Cabrera, P.A. Underhill, Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions, *BMC Genet.* 10 (2009) 59.
- [8] T.E. King, S.J. Ballereau, K. Schürer, M.A. Jobling, Genetic signatures of coancestry within surnames, *Curr. Biol.* 16 (2006) 384-388.
- [9] T.W. Athey, Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach, *J. Genet. Geneal.* 2 (2006) 34-39.
- [10] T.M. Karafet, F.L. Mendez, M. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y-chromosomal haplogroup tree, *Genome Res.* 18 (2008) 830-838.
- [11] H.-J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, *Mol. Biol. Evol.* 16 (1999) 37-48.
- [12] J.B. Kruskal, Multidimensional scaling by optimizing a goodness of fit test to a nonmetric hypothesis, *Psychometrika* 19 (1964) 1-27.
- [13] S. Willuweit, L. Roewer, The new Y Chromosome Haplotype Reference Database, *Forensic Sci Int Genet* 15 (2015) 43-8.
- [14] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564-7.
- [15] P. Hallast, C. Batini, D. Zadik, P. Maisano Delser, J.H. Wetton, E. Arroyo-Pardo, G.L. Cavalleri, P. de Knijff, G. Destro Bisol, B.M. Dupuy, H.A. Eriksen, L.B. Jorde, T.E. King, M.H. Larmuseau, A. Lopez de Munain, A.M. Lopez-Parra, A. Loutradis, J. Milasin, A. Novelletto, H. Pamjav, A. Sajantila, W. Schempp, M. Sears, A. Tolun, C. Tyler-Smith, A. Van Geystelen, S. Watkins, B. Winney, M.A. Jobling, The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades, *Mol. Biol. Evol.* 32 (2015) 661-73.
- [16] D.B. Goldstein, A.R. Linares, L.L. Cavalli-Sforza, M.W. Feldman, An evaluation of genetic distances for use with microsatellite loci, *Genetics* 139 (1995) 463-471.
- [17] D.B. Goldstein, A.R. Linares, L.L. Cavalli-Sforza, M.W. Feldman, Genetic

- absolute dating based on microsatellites and the origin of modern humans, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 6723-6727.
- [18] G. Ferri, C. Robino, M. Alu, D. Luiselli, S. Tofanelli, L. Caciagli, C. Onofri, S. Pelotti, C. Di Gaetano, F. Crobu, G. Beduschi, C. Capelli, Molecular characterisation and population genetics of the DYS458 .2 allelic variant, *Forensic Sci. Int. Genet. Suppl. Ser.* 1 (2008) 203–205.
- [19] G. Iacovacci, E. D'Atanasio, O. Marini, A. Coppa, D. Sellitto, B. Trombetta, A. Berti, F. Cruciani, Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries, *Forensic Sci. Int. Genet.* 27 (2017) 123-131.
- [20] P. Balaresque, G.R. Bowden, E.J. Parkin, G.A. Omran, E. Heyer, L. Quintana-Murci, L. Roewer, M. Stoneking, I. Nasidze, D.R. Carvalho-Silva, C. Tyler-Smith, P. de Knijff, M.A. Jobling, Dynamic nature of the proximal AZFc region of the human Y chromosome: multiple independent deletion and duplication events revealed by microsatellite analysis, *Hum. Mutat.* 29 (2008) 1171-1180.
- [21] Applied Biosystems, Yfiler™ Plus PCR Amplification Kit User Guide, Life Technologies Corporation, Carlsbad, CA, 2016.
- [22] F. Manni, P. Leonardi, A. Barakat, H. Rouba, E. Heyer, M. Klintschar, K. McElreavey, L. Quintana-Murci, Y-chromosome analysis in Egypt suggests a genetic regional continuity in Northeastern Africa, *Hum. Biol.* 74 (2002) 645-58.
- [23] I. Nasidze, H. Schadlich, M. Stoneking, Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci, *Forensic Sci. Int.* 137 (2003) 85-93.
- [24] L. Roewer, S. Willuweit, M. Stoneking, I. Nasidze, A Y-STR database of Iranian and Azerbaijani minority populations, *Forensic Sci. Int. Genet.* 4 (2009) e53-5.
- [25] C. Hallenberg, B. Simonsen, J. Sanchez, N. Morling, Y-chromosome STR haplotypes in Somalis, *Forensic Sci. Int.* 151 (2005) 317-21.
- [26] J. Piatek, A. Ossowski, M. Parafiniuk, A. Pudlo, K. Jasionowicz, K. Jalowinska, A. Niemcunowicz-Janica, M. Konarzewska, W. Pepinski, Y-chromosomal haplotypes for the AmpFLSTR Yfiler PCR amplification kit in a population sample of Bedouins residing in the area of the Fourth Nile Cataract, *Forensic Sci. Int. Genet.* 6 (2012) e176-7.
- [27] L.A. Zhivotovsky, P.A. Underhill, C. Cinnioglu, M. Kayser, B. Morar, T. Kivisild, R. Scozzari, F. Cruciani, G. Destro-Bisol, G. Spedini, G.K. Chambers, R.J. Herrera, K.K. Yong, D. Gresham, I. Tournev, M.W. Feldman, L. Kalaydjieva, The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time, *Am. J. Hum. Genet.* 74 (2004) 50-61.
- [28] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. de Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, *Forensic Sci. Int. Genet.* 6 (2012) 208-18.
- [29] J. Schlecht, M.E. Kaplan, K. Barnard, T. Karafet, M.F. Hammer, N.C. Merchant, Machine-learning approaches for classifying haplogroup from Y chromosome STR data, *PLoS Comput. Biol.* 4 (2008) e1000093.
- [30] A. Seman, Z.A. Bakar, M.N. Isa, An efficient clustering algorithm for partitioning Y-short tandem repeats data, *BMC Res. Notes* 5 (2012) 557.
- [31] T.W. Athey, Haplogroup prediction from Y-STR values using an allele frequency approach, *J. Genet. Geneal.* 1 (2005) 1-7.
- [32] M. Muzzio, V. Ramallo, J.M. Motti, M.R. Santos, J.S. Lopez Camelo, G. Bailliet,

- Software for Y-haplogroup predictions: a word of caution, *Int. J. Legal Med.* 125 (2011) 143-7.
- [33] W. Athey, Comments on the article, "Software for Y haplogroup predictions, a word of caution", *Int. J. Legal Med.* 125 (2011) 901-3; author reply 905-6.
- [34] L. Gusmao, J.M. Butler, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, A. Carracedo, Revised guidelines for the publication of genetic population data, *Forensic Sci. Int. Genet.* doi: 10.1016/j.fsigen.2017.06.007 (2017).
- [35] C. Batini, M.A. Jobling, Detecting past male-mediated expansions using the Y chromosome, *Hum Genet* 136 (2017) 547-557.
- [36] T. Mohammad, Y. Xue, M. Evison, C. Tyler-Smith, Genetic structure of nomadic Bedouin from Kuwait, *Heredity (Edinb)* 103 (2009) 425-33.
- [37] J. Chiaroni, R.J. King, N.M. Myres, B.M. Henn, A. Ducourneau, M.J. Mitchell, G. Boetsch, I. Sheikha, A.A. Lin, M. Nik-Ahd, J. Ahmad, F. Lattanzi, R.J. Herrera, M.E. Ibrahim, A. Brody, O. Semino, T. Kivisild, P.A. Underhill, The emergence of Y-chromosome haplogroup J1e among Arabic-speaking populations, *Eur. J. Hum. Genet.* 18 (2010) 348-53.
- [38] A. Nebel, E. Landau-Tasseron, D. Filon, A. Oppenheim, M. Faerman, Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa, *Am. J. Hum. Genet.* 70 (2002) 1594-6.
- [39] P.A. Zalloua, Y. Xue, J. Khalife, N. Makhoul, L. Debiane, D.E. Platt, A.K. Royyuru, R.J. Herrera, D.F. Hernanz, J. Blue-Smith, R.S. Wells, D. Comas, J. Bertranpetti, C. Tyler-Smith, Y-chromosomal diversity in Lebanon is structured by recent historical events, *Am. J. Hum. Genet.* 82 (2008) 873-882.
- [40] D.E. Platt, M. Haber, M.B. Dagher-Kharrat, B. Douaihy, G. Khazen, M. Ashrafian Bonab, A. Salloum, F. Mouzaya, D. Luiselli, C. Tyler-Smith, C. Renfrew, E. Matisoo-Smith, P.A. Zalloua, Mapping post-glacial expansions: the peopling of Southwest Asia, *Sci. Rep.* 7 (2017) 40338.
- [41] W. Wei, Q. Ayub, Y. Xue, C. Tyler-Smith, A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping, *Forensic Sci. Int. Genet.* 7 (2013) 568-72.
- [42] H.S. Groucutt, M.D. Petraglia, The prehistory of the Arabian peninsula: deserts, dispersals, and demography, *Evol. Anthropol.* 21 (2012) 113-25.
- [43] M. El-Sibai, D.E. Platt, M. Haber, Y. Xue, S.C. Youhanna, R.S. Wells, H. Izaabel, M.F. Sanyoura, H. Harmanani, M.A. Bonab, J. Behbehani, F. Hashwa, C. Tyler-Smith, P.A. Zalloua, Geographical structure of the Y-chromosomal genetic landscape of the Levant: a coastal-inland contrast, *Ann. Hum. Genet.* 73 (2009) 568-81.
- [44] Z. Taqi, M. Alenizi, H. Alenizi, S. Ismael, A.A. Dukhyil, M. Nazir, S. Sanqoor, E. Al Harbi, J. Al-Jaber, J. Theyab, R.S. Moura-Neto, B. Budowle, Population genetics of 23 Y-STR markers in Kuwaiti population, *Forensic Sci. Int. Genet.* 16 (2015) 203-4.
- [45] S. Triki-Fendri, S. Alfadhli, I. Ayadi, N. Kharrat, H. Ayadi, A. Rebai, Genetic structure of Kuwaiti population revealed by Y-STR diversity, *Ann. Hum. Biol.* 37 (2010) 827-35.
- [46] A.M. Cadenas, L.A. Zhivotovsky, L.L. Cavalli-Sforza, P.A. Underhill, R.J. Herrera, Y-chromosome diversity characterizes the Gulf of Oman, *Eur. J. Hum. Genet.* 16 (2008) 374-86.
- [47] M. Nazir, H. Alhaddad, M. Alenizi, H. Alenizi, Z. Taqi, S. Sanqoor, A. Alrazouqi, A. Hassan, R. Alfalasi, S. Gaur, J. Al Jaber, J. Ziab, E. Al-Harbi, R.S. Moura-Neto, B.

Budowle, A genetic overview of 23Y-STR markers in UAE population, *Forensic Sci. Int. Genet.* 23 (2016) 150-2.

Figure legends

Figure 1: Map of Saudi Arabia, showing population density and sub-regional divisions used in this study.

Population density is indicated by shading as shown in the key, top right, and locations of some major cities are shown. Adapted from Global Rural-Urban Mapping Project (sedac.ciesin.columbia.edu/gpw/), under a Creative Commons 3.0 Attribution License. Administratively, Saudi Arabia is divided into 13 regions which we here consider as five larger geographical areas, namely: Central (Riyadh, Al-Qassim), Northern (Northern borders region, Tabuk, Al-Jawf and Hail), Southern (Asir, Jazan, Bahah and Najran), Eastern (Eastern province) and Western (Mecca and Medina).

Figure 2: Multidimensional scaling (MDS) plots of Arabian Peninsula populations based on Y-STR haplotypes.

Comparison with other datasets required reduction of the number of STRs to a shared set of nine.

- a) Comparison of our total dataset (KSA) with an independent dataset of 106 Y-STR haplotypes from the same country ('Saudi Arabia' [6]), and other datasets from the Arabian peninsula: Iraq (n=249; YHRD data), Jordan (n=254 [43]), Kuwait (n=645 [44, 45]), Oman (n=262 [6]), Qatar (n=46 [46]), UAE (n=684 [6, 47]), Yemen (n=375 [6], plus YHRD data).
- b) Comparison of our dataset divided into five geographical sub-groups (KSA-N, -S, -E, -W, -C) with other Arabian Peninsula datasets as in part (a).

Figure 3: Median-joining network of Y-STR haplotypes, and geographical distribution of predicted haplogroups.

- a) Median-joining network for 597 Saudi Arabian haplotypes, constructed from data on 23 Y-STRs. Circles represent haplotypes, with area proportional to sample size, and lines between them proportional to the number of mutational steps. Colours represent haplogroups given in the key, top left. b) Map showing distributions of predicted haplogroups in five regional samples as pie-charts, not

to scale; haplogroup distribution in the overall sample is represented in the pie-chart inset top right. Colours of sectors indicate haplogroups as shown in the key.

Figure 4: Comparison of Saudi- and UK-recruited cohorts by frequency of predicted haplogroups.

a) Map showing distributions of predicted haplogroups in the five geographical regions, each divided into Saudi-recruited (outer pie-chart) and UK-recruited (inner pie-chart) samples. Pie-charts are not to scale. Haplogroup distributions in Saudi-recruited and UK-recruited samples for the total dataset are shown in the pie-chart inset top right. In each case, the *p*-value of a population differentiation test between Saudi- and UK-recruited samples is given. Colours of sectors indicate haplogroups as show in the key bottom right. b) Comparison of the haplogroup distribution in the total dataset (top) with that in published data [7] (bottom), with the *p*-value of the population differentiation test given.

Tables

Table 1: Diversity summary statistics for Y-STR haplotypes in the entire sample set, and by geographical subdivisions.

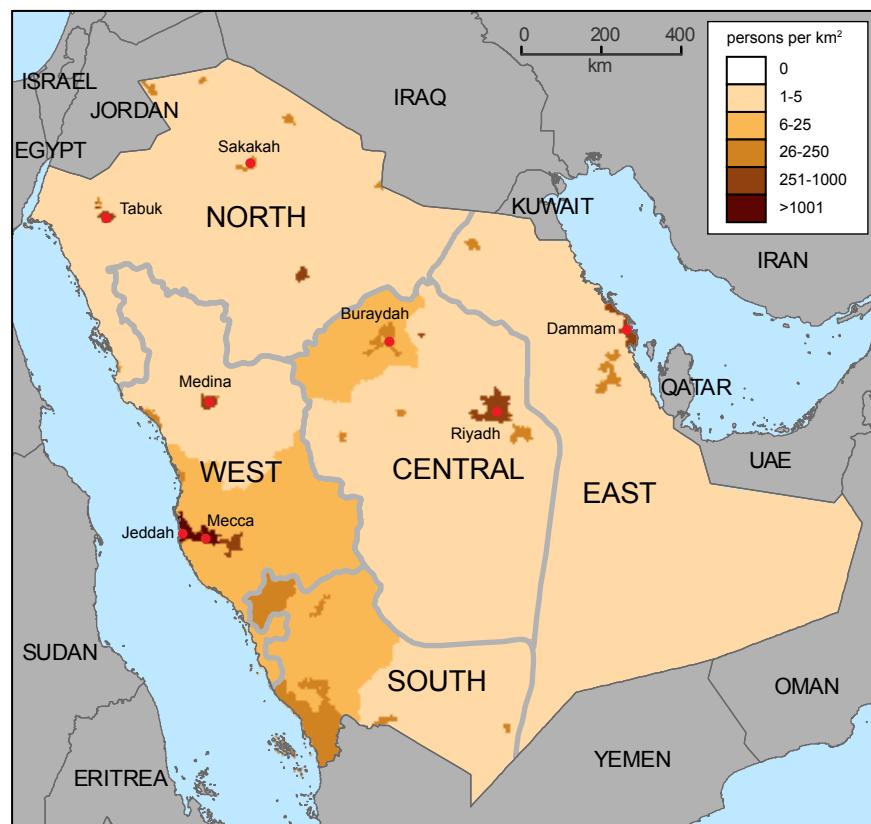
		Yfiler® plus							Yfiler® comparison			
pop	n	No. unique hts	No. pair hts	No. trio hts	HMP	Haplotype Diversity	% unique hts	DC	HMP	Haplotype Diversity	% unique hts	DC
All	597	541	25	2	0.0018	0.9998	91.0	95.3	0.0048	0.9969	63.0	74.7
C	125	117	4		0.0085	0.9995	93.6	96.8	0.0245	0.9834	62.4	72.0
E	110	110			0.0091	1.0000	100.0	100.0	0.0099	0.9992	90.9	95.5
N	106	92	7		0.0107	0.9987	86.8	93.4	0.0176	0.9917	57.5	74.5
S	140	125	6	1	0.0081	0.9991	89.3	94.3	0.0106	0.9965	72.1	83.6
W	116	106	5		0.0094	0.9993	91.4	95.7	0.0113	0.9973	79.3	87.9

pop: population; n: number of individuals; ht: haplotype; HMP: haplotype match probability; DC: discrimination capacity; C: central; E: eastern; N: northern; S: southern; W: western. Yfiler® comparison: lists statistics considering only those STRs included in the 17-STR Yfiler® kit.

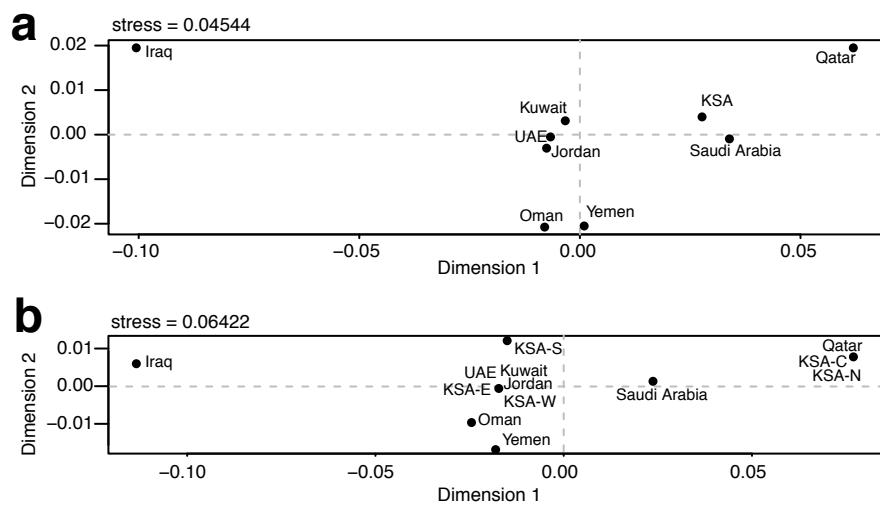
Table 2: Predicted haplogroup distributions and diversities in the entire sample set, and by geographical subdivisions.

		Predicted haplogroup															
pop	n	A	B	E1b1a	E1b1b	G	H	J1	J2	L	Q	R1a	R1b	T	UP	h ± s.d.	
All	597	5	10	9	66	8	2	424	16	6	8	14	5	15	9	0.481 ± 0.024	
C	125	1	0	0	7	1	0	107	2	1	1	1	1	1	2	0.265 ± 0.052	
E	110	2	3	3	20	3	2	51	7	4	3	7	2	3	0	0.745 ± 0.037	
N	106	0	0	1	7	2	0	91	2	0	0	0	1	0	2	0.260 ± 0.056	
S	140	0	2	4	9	0	0	113	3	0	1	3	1	4	0	0.344 ± 0.052	
W	116	2	5	1	23	2	0	62	2	1	3	3	0	7	5	0.671 ± 0.041	

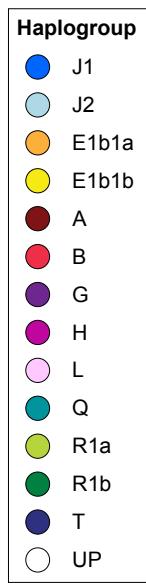
pop: population; n: number of individuals; UP: unpredicted haplogroup; h: gene diversity; s.d.: standard deviation; C: central; E: eastern; N: northern; S: southern; W: western.



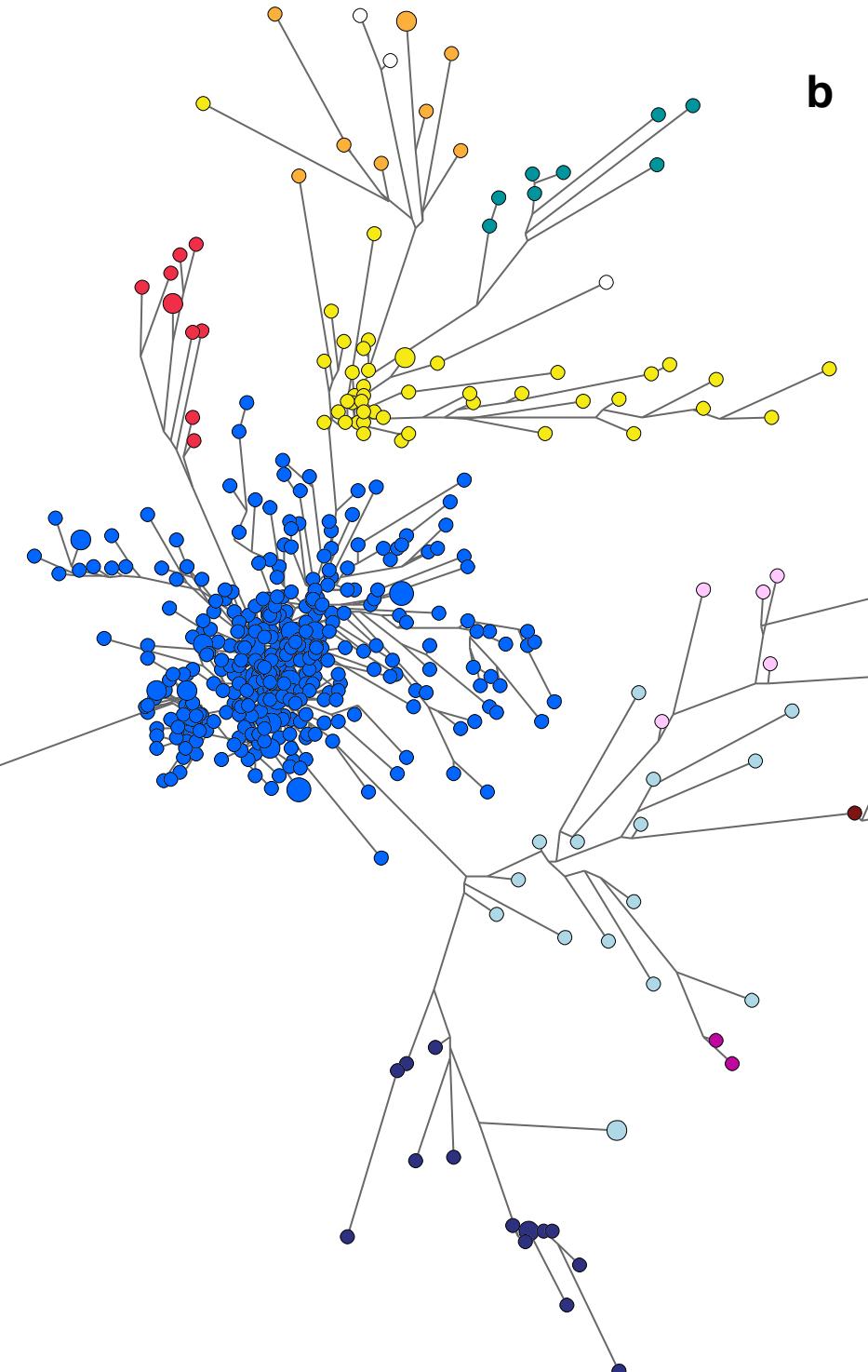
Khubrani et al., Figure 1.



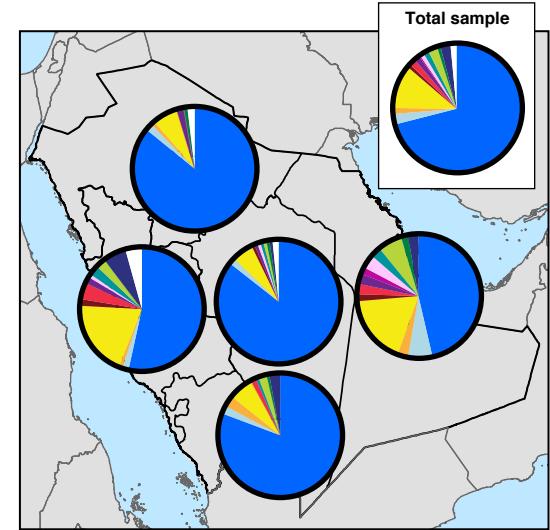
Khubrani et al., Figure 2.



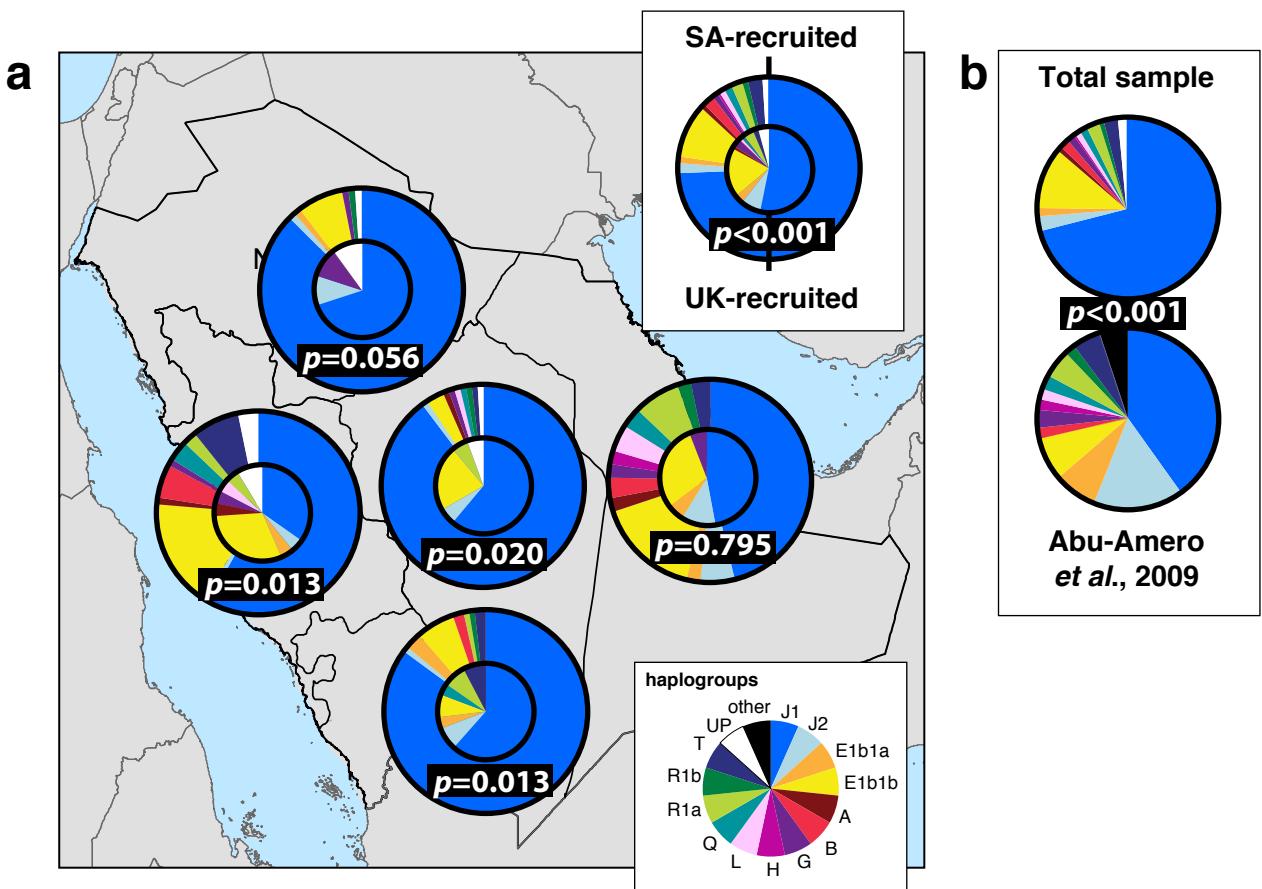
a



b



Khubrani et al., Figure 3.



Khubrani et al., Figure 4.

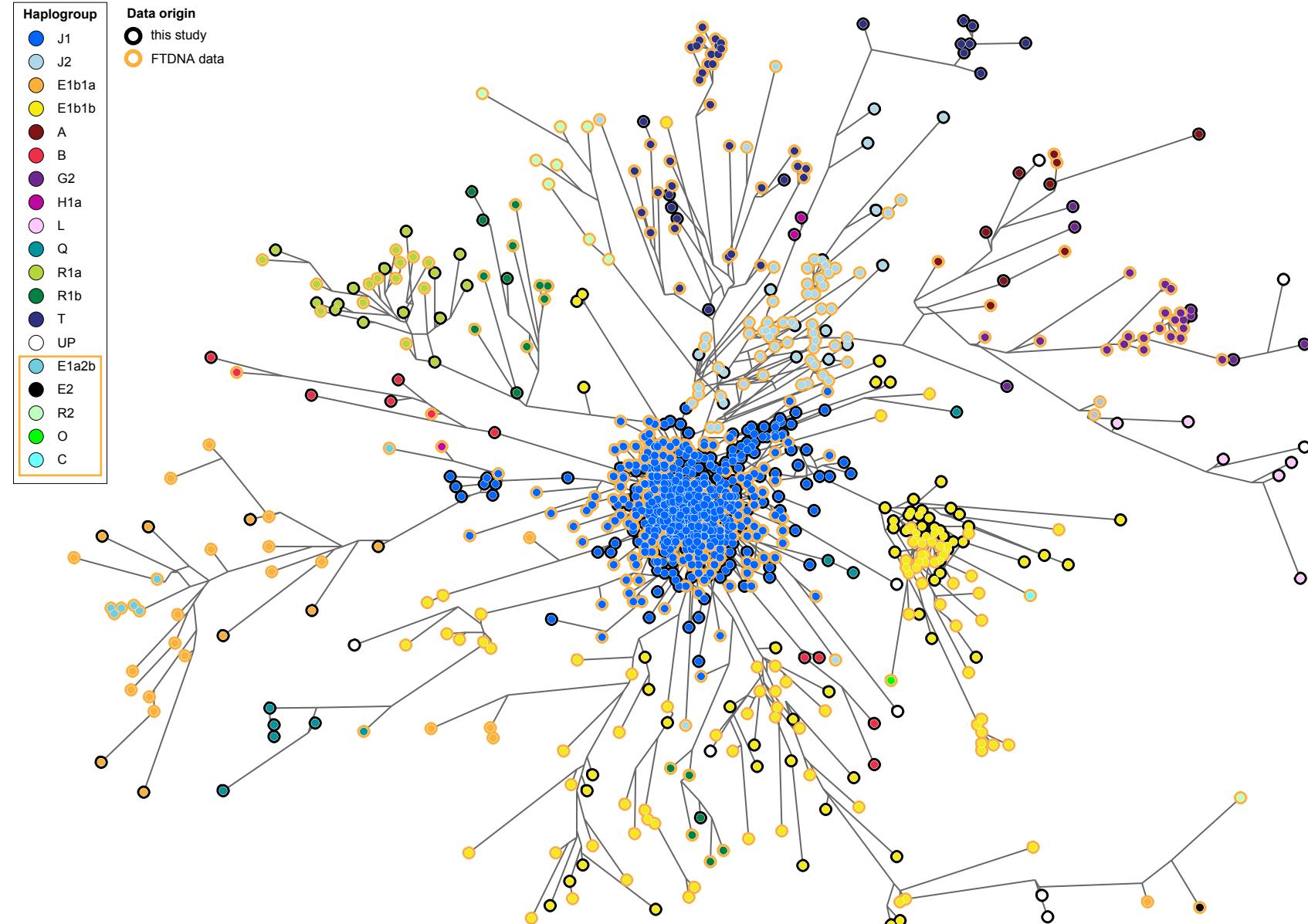


Figure S1: Median-joining network of Y-STR haplotypes, including FT-DNA samples and samples described in this study.

Median-joining network constructed from 23-Y-STR haplotypes in the 597 males of the current study, combined with 743 haplotypes from the Saudi Arabian FamilyTreeDNA (FTDNA) dataset. Circles represent haplotypes, with area proportional to sample size, and lines between them proportional to the number of mutational steps. Circle fill colours represent haplogroups, and circle edge colours indicate the two different sample origins, as shown in the key, top left. Within the key panel, the orange box highlights the five haplogroups found only in the FamilyTreeDNA dataset.

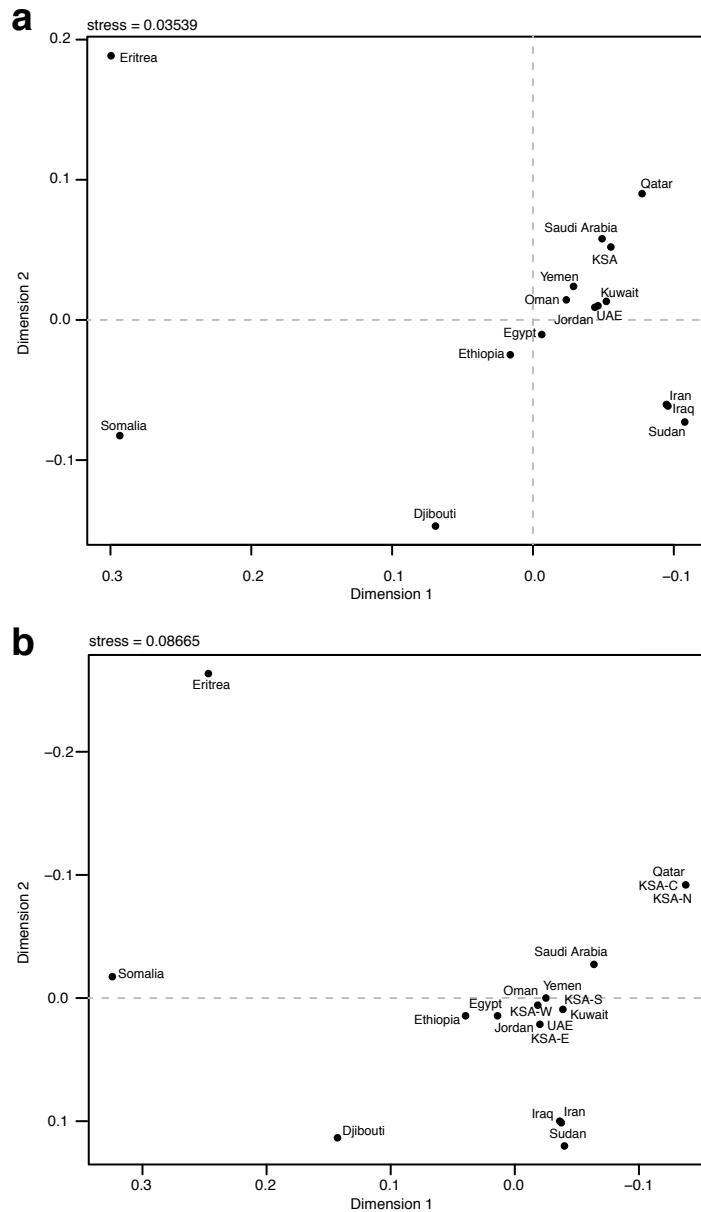


Figure S2: Multidimensional scaling (MDS) plots based on Y-STR haplotypes, including regional populations.

Comparison with other datasets required reduction of the number of STRs to a shared set of nine.

a) Comparison of our total dataset (KSA) with an independent dataset of 106 haplotypes from the same country ('Saudi Arabia' - [1]), and other datasets from the Arabian Peninsula: Iraq (n=249; YHRD data), Jordan (n=254 [2]), Kuwait (n=645 [3, 4]), Oman (n=262 [1]), Qatar (n=46 [5]), UAE (n=684 [1, 6]), Yemen (n=375 [1], plus YHRD data), plus a geographically broad dataset from the Arabian Peninsula and additional surrounding nations (Djibouti (n=54; YHRD data), Egypt (n=289 [7]), Eritrea (n=161 [8]), Ethiopia (n=275 [8]), Iran (n=1887 [1, 9, 10]), Somalia (n=201 [11]), Sudan (n=64 [12])).

b) Comparison of our dataset divided into five geographical sub-groups (KSA-N, -S, -E, -W, -C) with other datasets as in part (a).

References: [1] Alshamali et al., Local population structure in Arabian Peninsula revealed by Y-STR diversity, *Hum. Hered.* 68 (2009) 45-54. [2] El-Sibai et al., Geographical structure of the Y-chromosomal genetic landscape of the Levant: a coastal-inland contrast, *Ann. Hum. Genet.* 73 (2009) 568-81. [3] Taqi et al., Population genetics of 23 Y-STR markers in Kuwaiti population, *Forensic Sci. Int. Genet.* 16 (2015) 203-4. [4] Triki-Fendri et al., Genetic structure of Kuwaiti population revealed by Y-STR diversity, *Ann. Hum. Biol.* 37 (2010) 827-35. [5] Cadena et al., Y-chromosome diversity characterizes the Gulf of Oman, *Eur. J. Hum. Genet.* 16 (2008) 374-86. [6] Nazir et al., A genetic overview of 23Y-STR markers in UAE population, *Forensic Sci. Int. Genet.* 23 (2016) 150-2. [7] Manni et al., Y-chromosome analysis in Egypt suggests a genetic regional continuity in Northeastern Africa, *Hum. Biol.* 74 (2002) 645-58. [8] Iacovacci et al., Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries, *Forensic Sci. Int. Genet.* 27 (2017) 123-131. [9] Nasidze et al., Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci, *Forensic Sci. Int.* 137 (2003) 85-93. [10] Roewer et al., A Y-STR database of Iranian and Azerbaijani minority populations, *Forensic Sci. Int. Genet.* 4 (2009) e53-5. [11] Hallenberg et al., Y-chromosome STR haplotypes in Somalis, *Forensic Sci. Int.* 151 (2005) 317-21. [12] Piatek et al., Y-chromosomal haplotypes for the AmpFISTR Yfiler PCR amplification kit in a population sample of Bedouins residing in the area of the Fourth Nile Cataract, *Forensic Sci. Int. Genet.* 6 (2012) e176-7.

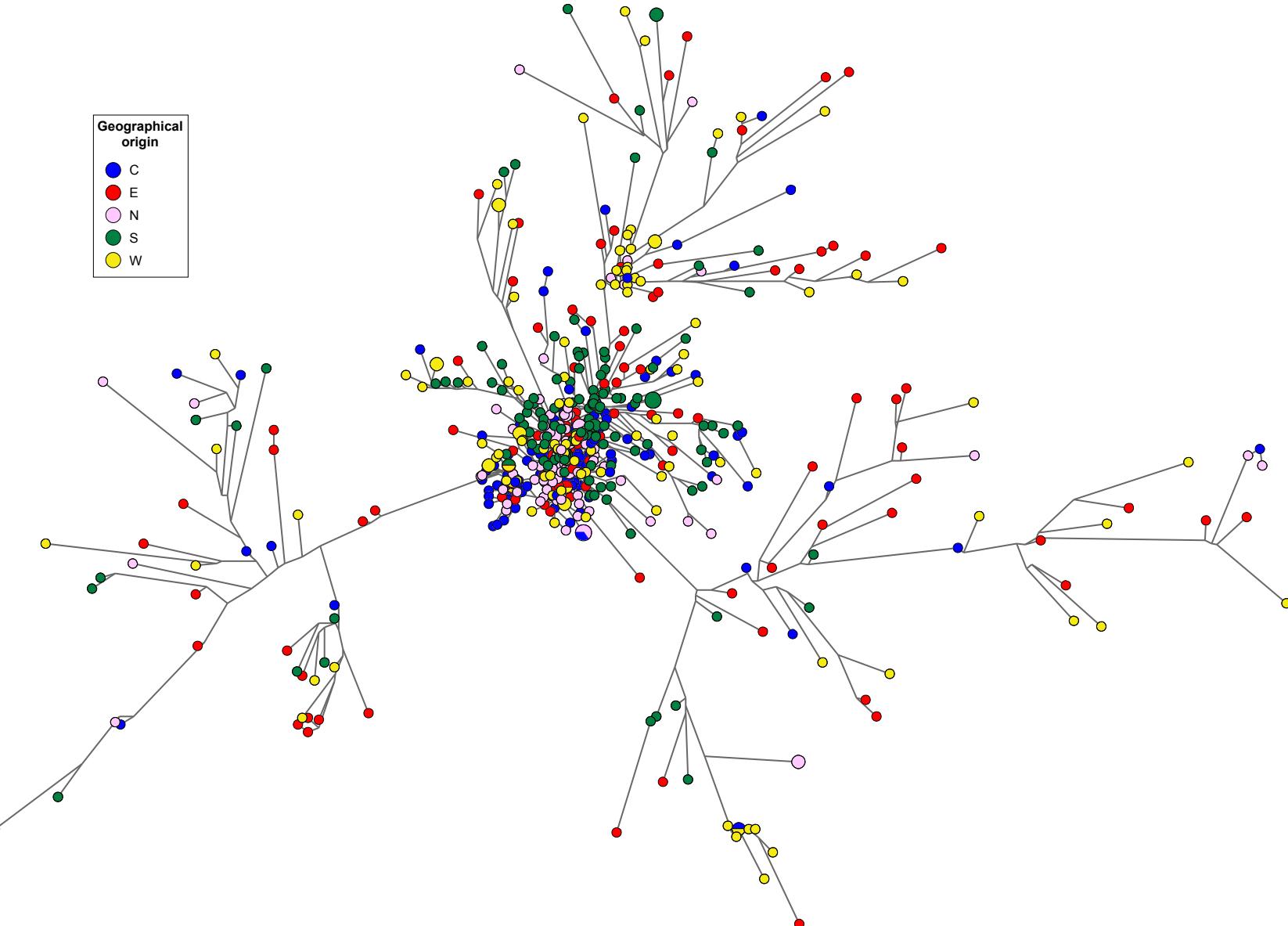


Figure S3: Median-joining network of Y-STR haplotypes, showing sub-regions of origin.

Median-joining network for 597 Saudi Arabian haplotypes, constructed from data on 23 Y-STRs. Circles represent haplotypes, with area proportional to sample size, and lines between them proportional to the number of mutational steps. Colours represent the five different geographical sub-regions of origin, as given in the key, top left.

Table S1: Sub-regional affiliations, Y-STR haplotypes, and predicted Y-haplogroups for 597 Saudi Arabian males.

Sample ^a	Recruitment	DYS387S1a	DYS19	DYS385a,b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4	Neogen predicted haplogroup ^b	Haplogroup probability	Haplogroup fitness
C1	SA	33,39,2	14	10,10	13	27	25	11	11	13	14	10	11	18	30	17	17	11	27	40	11	16	17	20	21	12	UP	90.06	8.03
C10	SA	37,37	14	13,18	13	30	23	10	11	12	14	10	11	20	25	14	19,2	10	26	40	12	17	17	20	22	11	J1	99.42	49
C100	SA	36,39	14	13,18	13	30	23	10	11	12	14	10	11	20	25	15	17,2	11	25	38	12	19	17	22	21	11	J1	100	55.64
C101	SA	36,36	15	12,12	13	29	21	9	11	13	15	10	11	21	35	15	19	10	25	38	12	18	17	22	19	12	A	98.83	25.64
C104	SA	36,37	14	13,19	13	29	23	12	11	12	14	10	11	20	25	14	18,2	10	25	41	12	16	17	22	21	11	J1	100	45.24
C105	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	19,2	10	26	40	11	18	17	23	21	11	J1	100	64
C106	SA	36,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	19,2	10	23	40	11	18	18	23	21	11	J1	100	58.17
C107	SA	36,37	14	12,17	14	31	23	10	11	12	14	10	11	20	26	15	17,2	10	25	37	11	17	19	21	20	11	J1	100	44.87
C108	SA	37,37	14	13,21	13	30	23	11	11	12	14	10	11	20	25	14	18,2	10	26	38	11	18	18	20	21	11	J1	100	57.07
C109	SA	36,36	14	13,20	13	30	23	11	11	12	14	10	11	20	25	14	18,2	11	26	39	11	18	18	21	21	11	J1	100	62.07
C11	SA	37,37	14	13,19	13	30	23	10	11	12	14	10	11	20	26	14	18,2	10	24	40	12	19	18	21	21	11	J1	100	53.01
C110	SA	36,37	14	13,21	13	31	23	11	11	12	14	10	11	20	25	14	18,2	10	26	39	11	18	18	22	21	11	J1	100	51.73
C111	SA	36,37	14	13,21	13	31	23	11	11	12	14	10	11	20	25	14	19,2	10	26	39	11	18	17	20	21	11	J1	100	49.81
C112	SA	37,37	14	13,19	13	30	23	12	11	12	14	10	11	20	25	14	18,2	10	26	40	12	19	18	21	21	12	J1	100	46.78
C113	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	18,2	10	25	38	11	18	16	20,2	21	12	J1	100	46.49
C114	SA	36,37	14	13,20	13	30	23	12	11	12	14	10	11	20	25	14	18,2	10	26	39	11	18	19	20	21	11	J1	100	52.97
C115	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	18,2	10	25	38	11	18	18	20,2	21	12	J1	100	51.03
C116	SA	36,39	14	13,18	14	31	23	10	11	12	14	10	11	20	24	14	17,2	11	25	40	11	18	17	19	21	11	J1	100	54.27
C117	SA	36,37	14	13,20	13	30	23	12	11	12	14	9	11	20	26	14	18,2	10	26	39	11	18	18	21	22	11	J1	100	39.32
C118	SA	37,37	14	13,20	14	30	23	11	11	12	14	10	11	20	25	14	18,2	11	26	40	13	17	18	21	21	11	J1	99.86	43.41
C119	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	19,2	10	24	41	12	18	18	21	21	11	J1	100	56.7
C12	SA	37,38	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	18,2	11	25	37	11	18	18	21	21	11	J1	100	62.79
C120	SA	37,37	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	18,2	10	24	39	12	17	18	21	21	11	J1	100	53.3
C13	SA	36,39	13	15,16	14	30	23	11	16	13	15	9	13	19	32	15	16	11	26	38	11	17	18	21	22	11	T	100	37.81
C14	SA	37,37	14	13,19	13	30	23	13	11	12	14	10	11	20	25	14	18,2	10	24	40	12	18	18	21	21	11	J1	100	45.39
C15	SA	37,38	15	13,18	13	30	22	10	11	12	14	10	12	20	27	15	17,2	10	24	39	11	19	19	21	21	12	J1	99.99	37.23
C16	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	26	14	18,2	10	24	39	12	17	18	21	21	11	J1	100	57.24
C17	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	18,2	10	24	40	12	17	18	22	21	11	J1	100	58.87
C18	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	18,2	10	24	40	12	18	18	21	21	11	J1	100	56.04
C2	SA	37,37	14	12,19	13	30	24	11	11	12	14	10	12	20	25	14	18,2	10	26	40	11	18	18	22	21	11	J1	99.94	50.93
C20	SA	36,37	14	12,17	14	31	23	10	11	12	14	10	12	21	26	15	17,2	10	25	37	11	17	18	22	20	11	J1	99.86	40.62
C21	SA	37,37	14	12,19	13	29	23	11	11	12	14	10	11	20	25	14	20,2	10	25	40	12	18	17	21	21	11	J1	99.99	45.76
C22	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	13	18,2	10	27	40	12	18	18	20	21	12	J1	100	37.63
C23	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	19,2	10	24	41	12	18	18	21	21	11	J1	100	56.7
C24	SA	36,39	14	13,18	13	30	23	10	11	12	14	10	11	20	24	14	17,2	11	26	40	11	18	17	19	21	11	J1	100	58.09
C25	SA	36,37	14	13,19	13	30	23	12	11	12	14	10	11	20	25	14	20,2	10	26	40	11	18	17	23	21	11	J1	100	53.88
C26	SA	36,37	14	13,20	13	30	23	11	11	12	14	10	11	20	25	14	19,2	10	26	40	11	18	18	23	21	11	J1	100	61.81
C28	SA	36,36	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	18,2	11	24	40	11	18	17	22	22	11	J1	100	51.46
C29	SA	37,37	14	12,19	13	29	23	11	11	12	14	10	11	20	25	14	20,2	10	25	41	13	17	18	21	21	11	J1	99.45	42.06
C30	SA	37,37	14	13,19	13	29	23	11	13	12	14	10	11	20	25	14	18,2	10	26	39	12	17	18	23	21	11	J1	99.43	31.52
C31	SA	37,37	14	13,18	13	30	23	9	11	12	14	10	11	20	26	13	19,2	11	21	40	11	19	19	22	22	10	J1	99.66	24.06
C32	SA	36,36	14	13,20	13	29	23	12	11	12	14	10	11	20	25	14	18,2	10	28	41	11	18	17	21	21	11	J1	100	41.61
C33	SA	36,41	15	9,15	13	29	22	10	14	12	15	10	12	19	29	14	17	11	24	37	13	16	15	21	22	11	UP	100	35.13
C34	SA	37,37	14	13,17	13	30	23	11	11	12	14	10	11	20	25	14	18,2	11	26	36	11	20	19	22	21	11	J1	100	50.6
C35	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	18,2	10	24	40	12	17	18	21	21	11	J1	100	57.24
C36	SA	39,40	14	16,17	14	31	24	10	11	13	14	10	13	21	29	15	18	10	25	40	10	20	18	21	21	11	E1b1b	100	33.3</

Sample ^a	Recruitment	DYS387S1	aDYS19	DYS385a,b	DYS389I	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4	Nevgen predicted haplogroup ^b	Haplogroup probability	Haplogroup fitness	
C55	SA	37,37	14	13,19	13	28	23	11	11	12	14	10	10	20	26	14	19.2	10	25	40	12	20	18	22	21	11	J1	99.99	43.56
C56	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	10	20	26	14	19.2	10	25	40	12	20	18	22	21	11	J1	100	44.29
C57	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	24	39	12	18	18	21	21	11	J1	100	58.87
C58	SA	37,38	14	13,19	13	30	23	11	11	12	14	10	11	19	25	14	18.2	10	24	38	11	18	16	20.2	21	12	J1	100	55.32
C59	SA	37,37	14	14,18	13	30	23	12	11	12	14	10	11	19	25	14	18.2	10	24	38	11	18	16	20.2	21	12	J1	100	34.16
C6	SA	37,37	14	13,18	13	29	24	11	11	12	14	10	11	20	24	12	18.2	11	25	39	11	18	17	20	21	11	J1	100	41.24
C60	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	25	38	11	18	16	20.2	21	12	J1	100	46.49
C61	SA	36,37	14	13,21	13	30	23	12	11	12	14	10	11	20	25	14	18.2	10	26	38	11	18	19	22	21	11	J1	100	48.71
C62	SA	36,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	19.2	10	26	40	11	18	18	22	21	11	J1	100	66.19
C63	SA	36,36	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	20.2	12	24	39	11	16	19	20	21	11	J1	100	45.41
C64	SA	36,39	16	13,14	14	30	25	10	13	14	14	12	12	19	25	15	17	10	22	39	11	20	17	21	23	12	R1b	99.74	27.1
C65	SA	37,37	14	13,18	13	30	23	12	11	12	14	10	11	19	25	15	19.2	10	25	38	11	18	16	20.2	21	12	J1	99.99	36.87
C66	SA	36,38	14	13,18	13	30	23	10	11	12	14	10	11	20	24	14	17.2	11	25	40	11	18	19	21	11	J1	100	59.95	
C67	SA	37,37	14	12,17	14	31	23	11	11	12	14	10	11	20	25	14	18.2	11	26	36	11	19	18	22	21	11	J1	99.99	45.54
C68	SA	37,39	14	13,18	13	30	23	10	11	12	14	10	11	20	24	14	17.2	11	25	37	11	17	17	19	21	12	J1	100	52.38
C69	SA	36,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	19.2	10	26	40	11	18	18	22	21	11	J1	100	59.91
C7	SA	36,40	14	14,17	13	30	23	12	11	12	14	10	11	19	25	14	18.2	11	25	40	12	17	16	22	21	11	J1	99.99	36.17
C70	SA	37,38	15	13,19	13	30	23	11	13	12	14	10	11	20	25	14	18.2	10	26	39	12	17	18	22	21	11	J1	99.99	29.61
C71	SA	37,37	14	13,18	13	32	23	11	11	12	14	10	11	20	25	14	18.2	11	25	41	12	20	17	22	24	11	J1	99.95	31.06
C72	SA	37,37	14	14,18	13	30	23	12	11	12	14	10	11	19	25	14	18.2	10	24	39	11	18	16	20.2	21	12	J1	100	34.16
C73	SA	36,37	14	13,21	13	30	23	12	11	12	14	10	11	20	25	14	18.2	10	26	39	11	18	18	21	21	11	J1	100	50.47
C74	SA	36,37	14	13,21	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	26	39	11	18	18	22	21	11	J1	100	57.07
C75	SA	35,36	14	17,17	14	31	24	10	11	13	14	10	12	20	32	15	18	11	21	37	12	20	17	21	20	12	E1b1b	100	29.36
C76	SA	36,37	14	13,21	13	30	23	12	11	12	14	10	11	20	25	14	18.2	10	26	39	11	18	18	20	21	11	J1	100	50.47
C77	SA	37,37	14	13,19	13	30	23	14	11	12	14	10	12	20	25	14	18.2	10	25	40	12	18	18	22	22	11	J1	99.98	33.88
C78	SA	36,37	14	13,20	13	30	23	12	11	12	14	10	11	20	26	14	18.2	10	26	39	11	18	19	21	21	11	J1	100	52.97
C8	SA	38,38	14	14,19	13	31	23	10	11	12	14	10	11	20	26	14	18.2	11	24	40	11	18	17	21	21	11	J1	100	47.05
C80	SA	37,38	14	13,18	14	31	23	11	11	13	14	10	12	20	25	14	18.2	12	25	40	11	17	19	22	22	11	J1	99.99	40.87
C81	SA	37,37	14	13,20	13	30	23	11	11	12	14	9	11	20	26	14	18.2	10	27	39	11	18	18	21	21	11	J1	100	45.01
C82	SA	37,38	15	15,17	12	29	22	10	11	13	15	10	11	20	28	17	16	11	21	39	9	17	17	21	22	11	G	100	49.93
C83	SA	38,39	13	13,17	13	31	23	10	11	12	14	10	13	20	26	15	17.2	11	26	39	11	18	18	20	21	11	J1	99.95	36.24
C84	SA	37,38	15	13,19	13	30	22	10	11	12	14	10	11	20	27	15	17.2	10	24	41	11	19	18	21	20	12	J1	100	35.98
C85	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	24	39	12	19	18	22	21	11	J1	100	55.32
C86	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	24	40	12	18	18	21	21	11	J1	100	58.87
C87	SA	35,36	12	15,16	13	29	22	10	15	13	14	11	12	19	26	15	19	11	24	37	11	16	16	19	22	11	Q	100	43.58
C88	SA	36,37	14	13,22	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	26	39	11	18	18	20	21	11	J1	100	53.47
C89	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	18.2	10	25	38	11	18	18	20	21	12	J1	100	46.49
C9	SA	37,37	13	13,18	13	29	23	11	11	12	14	10	14	20	26	13	18.2	11	27	40	11	17	18	23	21	11	J1	99.97	31.65
C90	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	24	38	11	18	16	20.2	21	11	J1	100	49.53
C94	SA	36,37	14	13,20	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	26	39	11	18	18	20	21	11	J1	100	62.07
C95	SA	37,38	15	13,19	13	30	22	10	11	12	14	10	11	20	28	15	18.2	10	24	40	11	19	18	21	20	12	J1	99.99	35.07
C96	SA	36,37	14	13,18	13	29	22	10	11	12	14	10	13	20	25	13	19.2	10	24	40	11	19	16	20	21	12	J1	99.54	28.49
C97	SA	37,37	14	13,18	13	32	23	11	11	12	14	10	11	19	25	14	18.2	10	25	38	11	19	16	20.2	21	12	J1	100	36.64
E100	SA	36,40	14	13,16	13	29	23	10	11	12	15	9	11	20	31	15	17	10	24	39	12	16	16	21	22	11	J2	99.94	56.41
E101	SA	39,39	14	12,20	14	30	24	10	11	12	14	10	11	21	27	15	17.2	11	24	39	11	18	18	21	21	11	J1	99.65	35.73
E102	SA	37,39	15	16,17	13	30	21	11	11	14	14	14	11	21	30	15	16	11	28	38	11	20	16	22	21	12	E1b1a	100	44.36
E103	SA	37,42	14	13,16	13	29	23	10</																					

Sample ^a	Recruitment	DYS387S1	aDYS19	DYS385a,b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4	Neogen predicted haplogroup ^b	Haplogroup probability	Haplogroup fitness
E20	SA	36,39	14	13,18	13	30	23	10	11	12	14	10	11	20	25	14	18.2	11	26	39	11	18	18	20	22	11	J1	100	43.79
E22	SA	36,37	14	13,18	14	31	23	10	11	12	14	10	11	20	25	14	18.2	11	26	39	11	18	18	20	22	11	J1	99.89	51.61
E24	SA	37,41	14	13,17	13	30	23	10	11	12	14	10	12	19	25	14	18.2	11	25	40	12	18	16	21	21	11	J1	99.99	42.74
E25	SA	37,37	14	13,17	13	30	23	10	11	12	14	10	11	20	26	14	19.2	12	25	41	11	17	18	21	21	11	J1	100	57.66
E27	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	12	20	26	14	17.2	11	25	38	11	19	18	21	21	11	J1	100	52.77
E29	SA	38,39	15	18,18	13	29	24	10	11	12	14	10	11	19	27	15	17.2	10	25	42	11	18	19	22	21	12	J1	99.59	28.42
E30	SA	35,39	14	17,19	14	32	24	10	11	13	14	10	13	21	34	15	16	11	23	37	11	18	16	22	20	12	E1b1b	99.98	24.09
E32	SA	40,41,2	14	11,12	13	32	23	10	11	15	15	10	12	20	30,2	13	18	11	25	41	12	20	17	18	17	11	B	100	16.79
E34	SA	39,39	14	16,17	13	31	25	10	11	13	14	10	12	21	29	16	19	10	27	39	10	19	18	18	21	11	E1b1b	99.87	24.46
E35	SA	37,39	15	15,17	14	30	22	10	11	12	15	9	11	19	32	16	15	10	25	40	11	18	16	19	20	10	H	90.08	21.78
E36	SA	36,37	14	13,19	13	30	23	11	11	12	14	10	12	20	25	15	18.2	11	26	40	12	18	18	21	21	11	J1	100	54.78
E37	SA	37,37	14	13,20	14	32	23	10	11	12	14	10	11	20	26	15	18.2	11	25	37	11	18	18	22	21	11	J1	99.95	48.41
E38	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	25	38	11	19	16	19,2	21	12	J1	100	43.86
E39	SA	39,39	14	16,18	14	32	24	10	11	13	14	10	12	20	29	16	17	9	24	40	10	18	18	18	21	11	E1b1b	100	36.61
E40	SA	36,36	15	11,13	13	29	21	9	11	14	16	10	12	21	33	15	15	10	16	39	10	17	17	21	20	12	G	50.57	12.88
E41	SA	39,43,2	15	11,11	13	31	23	11	11	13	14	10	12	20	32	13	18	11	25	37	11	17	19	19	17	11	B	100	44.1
E43	SA	35,38	14	13,17	13	31	23	10	11	12	14	10	11	20	26	15	17.2	10	25	40	11	19	19	21	20	12	J1	100	45.04
E44	SA	37,38	14	13,19	13	30	24	11	11	12	14	10	11	20	25	14	18.2	10	23	39	11	18	18	21	21	11	J1	100	52.17
E45	SA	37,38	11	16,16	13	30	24	10	12	13	14	11	11	21	34	15	14	10	24	38	11	18	17	19	21	11	E1b1b	87.08	22.14
E46	SA	39,39	14	16,16	13	30	24	10	11	13	14	10	13	21	29	16	18	10	25	38	10	19	18	18	22	10	E1b1b	100	28.75
E47	SA	35,38	14	13,17	13	32	23	10	11	13	14	10	11	20	26	15	17.2	10	26	40	11	19	19	20	20	12	J1	99.93	33.42
E49	SA	35,37	15	14,14	12	28	22	10	11	13	16	10	11	21	29	15	18	11	20	35	9	19	15	19	21	11	G	99.42	41.89
E50	SA	36,37	15	13,19	13	29	23	10	11	12	14	10	11	20	25	14	19.2	11	24	40	11	19	18	21	21	11	J1	99.87	45.84
E51	SA	35,37	12	16,20	13	31	25	10	12	13	14	11	11	20	31	15	13	11	23	41	11	20	18	21	22	12	E1b1b	67.83	18.71
E53	SA	36,38	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	18.2	11	24	39	11	19	19	21	22	12	J1	100	42.37
E54	SA	35,39	14	14,17	14	31	23	10	15	14	15	9	11	19	30	15	15	11	27	37	11	18	17	22	21	12	T	99.99	33.44
E55	SA	36,37	14	13,20	13	30	23	11	11	12	14	10	11	21	26	14	18.2	10	25	39	11	18	19	22	21	11	J1	100	50.9
E56	SA	37,37	14	13,20	13	30	23	11	11	12	14	10	12	20	25	13	18.2	10	24	42	12	19	17	21	21	11	J1	100	39.75
E57	SA	36,39	15	13,15	13	30	23	10	12	14	14	11	12	19	29	15	16	11	22	38	12	19	17	23	23	12	R1b	99.98	31.53
E58	SA	35,37	14	17,19	13	30	23	10	11	13	14	10	13	19	27	15	15	11	21	39	11	19	17	22	21	12	E1b1b	99.72	30.38
E59	SA	35,40	14	13,18	12	27	22	10	14	11	15	10	11	19	27	15	15	11	23	39	12	16	16	22	26	12	UP	100	41.89
E60	SA	36,38	14	13,19	13	30	23	11	11	12	14	10	12	20	26	14	18.2	10	27	39	11	18	19	22	21	11	J1	100	55.51
E61	SA	39,39	14	16,16	13	30	24	10	11	13	14	10	13	21	30	16	18	10	25	39	10	19	18	22	21	10	E1b1b	100	28.75
E62	SA	39,42	14	16,17	13	31	24	10	11	13	14	10	12	21	29	15	18	10	25	39	10	19	18	21	21	11	E1b1b	100	39.91
E63	SA	34,37	11	16,17	14	32	25	10	12	14	14	12	11	21	31	16	13	10	24	41	12	18	17	19	21	11	E1b1b	99.99	17.87
E64	SA	37,38	14	13,20	13	29	23	11	11	12	14	10	12	20	25	14	18.2	10	26	40	12	18	18	21	22	11	J1	99.99	46.3
E65	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	19.2	10	26	40	11	18	18	21	21	11	J1	100	65.58
E66	SA	37,43	15	9,16	13	29	22	10	14	12	16	10	11	18	27	15	17	11	23	37	12	15	20	18	21	11	UP	100	41.05
E67	SA	35,40	14	13,19	12	27	22	10	14	11	15	10	11	19	28	16	15	11	22	38	12	17	15	21	23	12	UP	100	30.65
E68	SA	37,39	14	13,18	13	30	23	10	11	12	14	10	11	20	26	14	18.2	11	25	38	11	19	17	22	22	11	J1	100	53.48
E69	SA	36,40	14	16,16	13	30	21	10	11	14	14	11	11	21	29	15	16	10	24	39	10	18	15	20	22	11	E1b1a	99.99	29.83
E70	SA	34,37	14	17,18	13	32	25	11	11	13	14	10	13	19	31	14	14	11	23	38	12	19	19	21	23	11	E1b1b	100	21.35
E71	SA	38,39	15	13,16	13	30	23	10	11	13	15	9	12	21	31	15	15	10	23	38	12	19	17	21	21	12	J2	99.97	53.28
E72	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	19.2	10	27	41	12	18	18	20	21	11	J1	100	47.16
E73	SA	36,38	14	13,14	13	30	23	10	11	13	14	10	11	20	26	15	17.2	10	24	39	10	17	17	21	21	11	J1	99.95	38.6
E74	SA	37,37	13	13,18	13	29	23	11	11	12	14	10	11	20	25	13	19.2	11	24	39	11	18	17	21	21	11	J1	100	40.08
E75	SA	37,37	14	13,18	13																								

Sample ^a	Recruitment	DYS387S1	aDYS19	DYS385a,b	DYS389I	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4	Nevgen predicted haplogroup ^b	Haplogroup probability	Haplogroup fitness	
E96	SA	38,39	14	16,17	14	31	24	10	11	13	14	11	21	32	15	16	12	23	39	12	19	19	17	23	13	E1b1b	100	37.07	
E97	SA	36,39	15	12,14	13	30	25	10	11	13	14	11	21	32	15	16	12	23	39	12	19	19	17	23	13	R1a	100	41.44	
E98	SA	37,37	14	14,18	13	31	23	11	11	12	14	10	11	20	25	14	20.2	10	26	40	12	18	17	21	11	J1	100	44.81	
E99	SA	36,38	14	13,18	13	29	23	11	11	12	14	10	11	18	34	16	16	11	22	38	12	19	15	18	18	11	E1b1b	99.89	13.95
N100	SA	37,39	14	16,19	13	32	23	11	11	13	14	10	11	18	34	16	16	11	22	38	12	19	15	18	18	11	E1b1b	99.89	13.95
N101	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	13	18.2	11	25	38	11	18	18	21	21	11	J1	100	58.03
N102	SA	37,37	14	12,19	13	30	23	11	11	12	14	10	12	20	25	14	18.2	10	26	39	11	18	17	21	21	11	J1	100	57.03
N103	SA	37,38	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	18.2	11	24	39	11	18	18	21	22	11	J1	100	53.22
N104	SA	37,38	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	18.2	11	25	37	11	18	18	21	21	12	J1	100	55.43
N105	SA	36,37	14	13,17	13	29	22	11	11	12	14	10	11	20	25	13	18.2	11	25	39	11	18	18	20	21	11	J1	100	45.66
N107	SA	36,38	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	18.2	11	25	38	11	17	19	22	21	11	J1	100	58.14
N108	SA	37,38	14	13,18	13	29	23	11	11	12	14	10	11	20	24	14	18.2	10	24	38	11	17	17	21	21	11	J1	100	55.54
N11	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	17.2	10	25	39	11	17	18	21	21	11	J1	100	57.99
N110	SA	36,36	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	21.2	12	24	39	11	17	19	21	21	11	J1	100	45.87
N111	SA	38,38	14	13,18	13	29	24	11	11	12	14	10	11	20	25	14	18.2	10	24	38	11	17	17	21	21	11	J1	100	49
N113	SA	37,37	14	13,20	12	29	23	11	11	12	14	10	11	20	25	14	19.2	10	26	39	12	18	17	19	21	11	J1	100	48.04
N114	SA	39,40	14	16,17	13	30	24	10	11	13	14	10	12	20	29	15	17	10	25	39	10	19	20	18	21	11	E1b1b	100	40.98
N115	SA	37,37	14	14,19	13	30	23	11	11	12	14	10	11	20	26	14	18.2	10	26	39	13	18	18	21	21	11	J1	100	49.78
N119	SA	36,37	14	13,20	13	30	23	12	11	12	14	10	11	20	25	14	18.2	10	26	39	11	19	20	22	21	11	J1	100	44.87
N12	SA	37,38	15	15,17	12	29	22	10	11	13	15	10	11	21	28	17	16	11	21	39	9	17	17	21	21	11	G	100	55.05
N13	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	17.2	10	25	39	11	17	18	21	21	11	J1	100	57.99
N15	SA	37,37	14	13,20	13	29	23	11	11	12	14	10	11	20	25	14	20.2	10	26	40	11	18	18	23	21	11	J1	100	53.74
N19	SA	36,37	14	12,19	14	31	23	11	11	12	14	10	12	20	25	14	18.2	10	26	40	11	18	18	22	21	11	J1	100	53.29
N20	SA	36,39	15	13,15	12	28	23	9	14	11	14	10	12	19	26	15	17	11	25	38	12	16	17	23	24	11	UP	71.45	23.23
N21	SA	37,38	14	13,17	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	23	38	11	18	18	23	21	11	J1	100	54.18
N22	SA	37,37	14	13,19	13	29	23	11	11	12	14	10	11	20	25	14	20.2	10	27	40	11	18	17	22	21	11	J1	100	51.71
N23	SA	36,38	14	13,18	13	30	23	10	11	13	14	10	11	20	27	15	17.2	12	25	38	11	17	19	22	22	11	J1	100	59.99
N24	SA	36,37	14	13,20	13	29	24	11	11	12	14	10	11	20	25	15	17.2	10	26	40	11	18	18	23	21	11	J1	100	41.22
N25	SA	37,37	14	13,19	14	31	23	11	11	12	14	10	11	20	25	14	19.2	10	27	40	12	18	18	20	22	12	J1	100	59.14
N26	SA	36,37	15	13,18	13	29	23	10	11	13	14	10	11	19	25	14	19.2	11	25	40	11	18	18	21	21	11	J1	100	59.87
N28	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	23	38	11	17	18	21	21	11	J1	100	57.96
N29	SA	36,36	14	13,18	13	29	23	10	11.1	12	14	10	11	20	25	14	20.2	11	25	39	11	17	19	21	21	11	J1	100	53.39
N30	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	26	14	18.2	11	24	38	11	19	19	22	22	11	J1	100	50.86
N31	SA	36,37	15	13,18	13	29	23	10	11	12	14	10	11	20	25	14	19.2	11	25	39	11	18	18	21	21	11	J1	100	59.83
N32	SA	36,37	14	13,18	13	30	23	10	11	12	14	10	11	20	25	14	19.2	11	24	39	11	18	19	21	22	11	J1	100	51.39
N33	SA	36,36	14	13,18	13	29	23	10	11	12	14	10	12	20	25	14	19.2	11	24	39	11	18	19	21	21	11	J1	100	50.29
N34	SA	36,36	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	20.2	11	24	39	11	16	19	21	21	11	J1	100	45.41
N35	SA	36,36	14	13,18	13	29	22	11	11	12	14	10	11	20	25	14	20.2	11	24	39	11	17	19	21	21	11	J1	100	46.46
N36	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	23	38	11	18	19	21	21	11	J1	100	56.63
N38	SA	37,37	14	13,19	13	30	23	12	11	12	14	10	11	20	25	14	18.2	9	24	41	12	17	17	21	21	11	J1	100	48.81
N39	SA	37,37	14	13,19	13	30	23	12	11	12	14	10	11	20	25	14	18.2	10	24	41	12	17	17	21	21	11	J1	100	48.81
N40	SA	37,37	14	12,19	13	30	22	11	11	12	14	10	14	20	25	14	19.2	10	26	40	11	18	17	22	21	11	J1	100	59.98
N41	SA	37,37	14	12,19	14	31	22	11	11	12	14	10	12	20	25	14	19.2	10	26	40	11	18	18	22	21	11	J1	100	45.64
N42	SA	37,38	15	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	23	37	11	18	18	22	21	11	J1	100	52.3
N43	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	25	38	11	17	18	23	21	11	J1	100	64.97
N44	SA	37,37	14	12,19	13	30	22	11	11	12	14	10	12	20	25	14	19.2	10	26	40	11	18	18	22	21	11	J1	100	50.52
N45	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11</																

Sample ^a	Recruitment	DYS387S1	aDYS19	DYS385a,b	DYS389I	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4	Neogen predicted haplogroup ^b	Haplogroup probability	Haplogroup fitness	
N67	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	25	38	11	17	18	21	21	11	J1	100	63.54
N69	SA	37,38	14	13,17	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	25	38	11	17	18	21	21	11	J1	100	59.14
N7	SA	38,38	14	13,18	13	29	24	10	11	12	14	10	11	20	25	14	18.2	10	24	38	11	17	17	21	21	11	J1	100	99.98
N70	SA	37,38	14	13,17	13	29	23	11	11	12	14	10	11	20	25	14	18.2	10	24	38	11	17	18	21	21	11	J1	100	53.42
N71	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	26	14	18.2	11	28	38	11	18	18	23	21	11	J1	100	55.41
N72	SA	36,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	25	39	11	18	18	21	22	11	J1	100	60.24
N73	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	25	38	11	17	18	23	21	11	J1	100	64.97
N74	SA	37,37	14	12,19	13	30	22	11	11	12	14	10	12	20	25	14	19.2	10	26	40	11	18	19	22	21	11	J1	100	48.23
N75	SA	36,39	14	13,17	14	30	23	10	11	14	14	9	11	17.4	31	16	14	11	22	42	11	17	17	21	20	11	J2	99.98	30.81
N76	SA	37,38	14	13,19	13	29	23	11	11	12	14	10	11	20	25	14	19.2	11	25	37	11	18	18	21	21	11	J1	100	59.9
N77	SA	36,37	14	13,17	13	29	23	11	11	12	14	10	11	20	25	14	18.2	11	24	39	11	18	18	19	21	11	J1	100	54.46
N78	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	19.2	11	23	37	11	18	19	22	21	11	J1	100	54.63
N79	SA	36,37	14	13,19	12	29	23	11	11	12	14	10	10	20	26	14	18.2	10	26	39	11	19	19	22	20	11	J1	100	41.42
N8	SA	37,40	13	16,17	13	32	24	10	11	13	14	10	12	20	33	15	18	10	25	39	11	18	17	19	21	11	E1b1b	100	42.24
N80	SA	37,38	14	12,19	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	23	38	11	18	18	22	21	11	J1	100	53.56
N81	SA	37,37	14	13,19	13	30	23	12	11	12	14	10	11	20	25	14	18.2	10	24	39	12	18	18	21	21	11	J1	100	51.92
N82	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	26	14	19.2	10	25	38	12	16	18	22	21	11	J1	100	52.3
N83	SA	37,37	14	12,19	13	30	22	11	11	12	14	10	14	20	25	14	19.2	10	26	39	11	18	17	22	22	11	J1	99.85	35.57
N84	SA	39,39	14	16,17	13	30	24	10	11	13	14	10	12	20	29	17	18	10	25	39	10	19	18	21	21	11	E1b1b	100	42
N85	SA	36,37	14	13,19	13	29	23	11	11	12	14	10	11	20	25	14	18.2	10	24	40	12	17	18	21	21	11	J1	100	53.72
N86	SA	39,41	14	16,16	13	31	24	10	11	13	14	10	12	20	29	15	18	10	24	39	10	19	19	18	21	11	E1b1b	100	37.98
N87	SA	36,37	13	13,16	14	30	25	9	11	14	14	10	10	20	32	15	16	11	27	40	11	22	17	18	21	12	E1b1b	100	34.99
N88	SA	37,40	16	13,14	14	30	26	10	13	14	14	12	12	19	26	15	17	11	22	39	11	20	18	21	23	13	R1b	98.92	20.8
N89	SA	37,37	14	12,19	13	29	23	11	11	12	14	10	11	20	25	14	19.2	10	25	40	13	18	18	21	21	11	J1	97.98	43.27
N90	SA	37,37	14	12,23	13	30	23	11	11	12	14	10	12	20	25	14	18.2	10	26	40	11	17	18	22	21	11	J1	100	36.79
N91	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	26	14	18.2	10	26	39	13	18	18	21	21	11	J1	100	57.21
N92	SA	37,37	14	12,23	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	26	40	11	18	18	22	21	11	J1	100	39.35
N93	SA	36,37	14	13,18	13	29	23	10	11	12	14	10	11	20	25	15	20.2	11	25	38	12	19	19	23	21	11	J1	97.85	45.6
N94	SA	37,37	14	12,23	13	30	23	11	11	12	14	10	12	20	25	14	18.2	10	26	40	11	18	18	22	21	11	J1	100	39.03
N95	SA	37,37	14	13,19	14	30	23	11	11	12	14	10	11	20	25	14	18.2	11	26	40	13	16	18	21	21	11	J1	99.79	42.42
N96	SA	36,37	14	18,20	13	30	23	10	11	13	14	10	11	21	36	16	16	10	23	36	11	18	17	21	21	13	E1b1b	65.58	22.58
N97	SA	37,37	14	12,23	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	26	40	11	18	18	22	21	11	J1	100	39.35
N98	SA	37,38	14	13,19	13	30	23	11	11	12	14	10	11	20	26	14	18.2	10	24	40	12	18	18	21	21	11	J1	100	58.87
N99	SA	36,39	15	16,16	13	31	21	10	12	13	14	11	12	21	31	14	15	10	26	40	11	18	15	22	21	11	E1b1a	100	31.76
S10	SA	38,40	14	16,19	13	31	25	11	11	13	14	10	12	21	34	15	19	10	27	36	10	19	17	16	25	11	E1b1b	99.41	17.33
S100	SA	37,37	14	13,15	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	25	39	11	18	17	21.2	20	11	J1	99.31	41.91
S101	SA	35,39	14	13,17	13	31	23	10	11	12	14	10	11	20	25	14	19.2	11	23	38	11	17	17	20	21	11	J1	99.98	44.17
S102	SA	36,36	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	25	39	11	18	17	24	21	11	J1	100	64.35
S103	SA	35,39	14	13,17	13	31	23	10	11	12	14	10	11	20	25	14	19.2	11	23	38	11	17	20	21	21	11	J1	99.98	44.17
S104	SA	36,38	15	15,18	12	29	21	10	11	14	14	11	12	18	31	15	19	12	29	39	11	20	16	20	22	11	E1b1a	100	23.68
S105	SA	36,38	14	13,18	13	30	24	11	11	12	14	10	11	19	27	14	18.2	11	24	37	11	19	16	20	21	11	J1	100	40.4
S106	SA	37,38	14	13,16	13	31	23	10	11	12	14	10	11	20	26	14	17.2	10	23	40	11	18	19	21	20	12	J1	99.97	35.3
S107	SA	37,37	14	13,18	13	30	23	10	11	13	14	10	11	19	25	14	18.2	11	25	39	11	18	16	21.2	21	11	J1	100	45.11
S108	SA	36,38	15	13,15	14	31	24	10	13	14	14	12	14	19	30	16	17	11	23	39	12	20	18	22	24	13	R1b	97.24	24.29
S109	SA	36,38	14	13,18	13	30	24	12	11	12	14	10	12	20	25	14	18.2	11	25	38	11	16	17	20	22	11	J1	100	38.02
S110	SA	36,36	15	13,19	14	31	25	12	11	13	15	10	13	20	36	15	15	10	25	42	8	21	16	21	20	13	E1b1b	100	11.08
S111	SA	37,37	14	13,17	13	30	23																						

Sample ^a	Recruitment	DYS387S1	aDYS19	DYS385a,b	DYS389I	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4	Nevgen predicted haplogroup ^b	Haplogroup probability	Haplogroup fitness	
S23	SA	36,38	14	13,18	13	30	23	11	11	12	14	10	12	20	25	15	18.2	11	25	37	11	17	17	20	21	11	J1	100	60.72
S24	SA	37,38	14	13,19	13	30	23	11	11	12	14	10	12	20	25	15	18.2	11	25	37	11	17	17	20	21	11	J1	100	53.15
S25	SA	36,38	14	13,18	13	29	24	10	11	12	14	10	11	20	25	14	18.2	11	25	38	11	19	18	20	21	11	J1	100	59.44
S26	SA	36,38	14	13,18	13	30	24	11	11	12	14	10	11	19	25	14	18.2	11	25	39	11	18	16	21.2	21	11	J1	100	54.63
S27	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	25	39	11	18	16	21.2	21	11	J1	100	52.66
S28	SA	37,37	14	13,17	13	31	23	10	11	12	14	10	11	20	25	9	17.2	10	23	40	11	18	17	20	21	12	J1	99.9	27.75
S29	SA	37,41	14	13,17	13	29	23	11	11	12	14	10	11	19	25	14	18.2	11	25	39	12	17	16	21	22	11	J1	99.98	38.31
S3	SA	37,41	14	13,17	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	25	39	12	17	16	21	22	11	J1	99.96	39.67
S30	SA	37,43.2	17	12,13	14	33	23	10	11	13	14	10	12	20	31	12	12	28	42	11	16	17	21	17	10	B	100	15.34	
S31	SA	37,38	14	13,18	13	31	23	10	11	12	14	10	11	20	26	14	19.2	12	25	39	11	19	18	21	11	J1	100	54.59	
S32	SA	38,38	14	13,20	13	30	23	11	11	12	14	10	12	19	25	15	18.2	11	24	40	12	19	17	21	11	J1	100	37.65	
S33	SA	36,36	14	13,18	13	31	23	10	11	12	14	10	11	20	25	14	18.2	11	25	40	11	18	17	22	20	11	J1	100	54.31
S34	SA	36,36	14	13,18	13	30	23	11	11	12	14	11	11	20	25	14	18.2	11	25	39	12	18	17	23	21	11	J1	100	45.43
S35	SA	37,38	14	13,17	13	30	23	10	11	13	14	10	11	20	26	15	17.2	10	24	40	11	18	20	22	20	11	J1	100	45.99
S36	SA	37,38	14	13,16	13	31	23	10	11	12	14	10	11	20	26	14	17.2	10	22	39	11	18	18	21	20	11	J1	99.96	36.73
S37	SA	36,44.2	17	12,12	14	33	23	10	12	13	14	10	12	21	29	12	12	27	42	11	19	17	20	17	10	B	100	11.5	
S38	SA	37,37	14	13,18	12	30	23	10	11	12	14	10	11	20	27	15	17.2	10	23	40	11	18	19	20	20	12	J1	100	37.65
S39	SA	36,38	14	13,18	13	30	24	11	11	12	14	10	11	20	25	15	18.2	11	25	37	11	17	16	20	21	11	J1	100	48.1
S4	SA	36,36	14	14,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	25	39	11	18	17	23	21	11	J1	100	55.99
S40	SA	37,40	14	13,17	13	31	23	10	11	12	14	10	11	20	26	15	16.2	10	23	40	11	18	18	20	20	11	J1	99.97	42
S41	SA	36,37	14	13,19	13	30	24	11	11	12	14	10	11	20	25	14	18.2	11	25	38	11	18	15	20	21	11	J1	100	49.23
S42	SA	36,40	15	11,14	14	32	25	10	11	13	14	11	10	20	33	15	17	10	23	43	12	19	19	17	24	13	R1a	100	43.68
S43	SA	34,37	14	13,17	14	32	23	12	11	12	14	10	11	20	26	14	18.2	11	25	37	11	18	18	22	22	11	J1	100	42.09
S44	SA	37,41	14	13,17	13	29	23	11	11	12	14	10	11	19	25	14	18.2	11	25	39	12	17	16	21	22	11	J1	99.98	38.31
S45	SA	37,37	14	13,17	13	31	23	10	11	12	14	10	11	20	26	15	17.2	10	23	40	11	18	19	21	20	12	J1	100	41.77
S46	SA	36,37	14	13,20	13	30	23	10	11	12	14	10	10	21	25	14	18.2	10	25	39	11	16	17	20	20	11	J1	99.17	37.06
S47	SA	36,38	14	13,18	13	30	24	11	11	12	14	10	12	20	25	14	18.2	11	24	38	11	17	17	20	21	11	J1	99.99	46.65
S48	SA	37,38	14	13,17	13	31	23	10	11	12	14	10	11	20	26	15	17.2	10	23	40	11	18	18	20	20	11	J1	100	48.05
S49	SA	37,37	14	13,17	13	29	23	11	11	12	14	10	12	20	25	14	19.2	11	25	39	11	18	18	21	20	12	J1	100	44.72
S5	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	25	39	11	18	16	21.2	21	11	J1	100	52.66
S50	SA	37,37	14	13,15	13	30	23	11	11	12	14	10	11	19	25	13	18.2	11	25	39	11	18	17	21.2	21	11	J1	99.85	38.92
S51	SA	37,39	16	16,18	14	31	22	10	11	15	14	11	11	21	31	17	16	11	24	39	12	17	16	18	21	11	E1b1a	100	29.66
S52	SA	37,37	13	14,16	12	29	23	10	11	13	14	10	12	19	30	14	11	11	24	40	11	17	16	20	21	12	E1b1b	99.94	18.63
S53	SA	35,38	14	17,17	13	29	23	11	13	15	14	9	11	19	33	16	17	10	23	36	12	16	16	17	22	11	T	99.99	22.17
S54	SA	36,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	25	39	11	16	17	22	21	10	J1	100	46.12
S55	SA	37,37	14	13,16	13	30	23	10	11	13	14	10	11	20	26	15	17.2	10	24	39	11	18	19	21	20	11	J1	100	45.91
S56	SA	37,40	14	13,17	13	31	23	10	11	12	14	10	11	20	26	15	16.2	10	23	40	11	18	18	20	20	11	J1	99.97	42
S57	SA	37,39	16	16,18	14	31	22	10	11	15	14	11	11	21	31	17	16	11	24	39	12	17	16	18	21	11	E1b1a	100	29.66
S58	SA	37,37	14	12,18	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	25	39	11	18	17	20.2	21	11	J1	99.98	45.56
S59	SA	35,38	14	17,17	13	29	23	11	13	15	14	9	11	19	33	16	17	10	24	36	12	16	16	17	22	11	T	97.25	20.21
S60	SA	37,37	14	13,15	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	25	39	11	18	17	21.2	20	11	J1	99.31	41.91
S61	SA	36,36	14	13,18	13	31	23	11	11	12	14	10	11	20	25	14	18.2	11	25	40	11	18	17	22	20	11	J1	100	56.07
S62	SA	37,38	13	13,17	13	31	23	10	11	12	14	10	12	20	27	15	17.2	11	26	39	11	18	19	17	20	11	J1	99.99	40.36
S63	SA	37,38	14	13,19	13	30	23	11	11	12	14	10	11	20	24	14	18.2	11	24	38	11	18	17	24	21	11	J1	100	61.35
S64	SA	37,40	14	13,17	13	30	23	11	11	12	14	10	11	19	25	14	18.2	11	25	40	12	17	16	21	22	11	J1	99.96	39.67
S65	SA	37,38	14	13,17	13	31	23	10	11	12	14	10	10	20	26	14	17.2	10	23	39	12	18	18	21	20	11	J1	100	37.2
S66	SA	36,37	14	12,17	14	31	23	10	11	12	14	10																	

Sample ^a	Recruitment	DYS387S1	aDYS19	DYS385a,b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4	Nevgen predicted haplogroup ^b	Haplogroup probability	Haplogroup fitness
S85	SA	38,38	14	13,17	13	33	23	10	12	14	10	11	20	26	15	19.2	10	23	40	11	18	17	21	20	11	J1	95.78	22.53	
S86	SA	36,38	14	14,18	13	30	24	10	11	12	14	10	11	20	26	15	18.2	11	25	37	11	19	17	20	21	11	J1	99.79	45.53
S87	SA	35,37	14	13,17	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	25	39	10	18	18	22	21	11	J1	100	51.84
S88	SA	38,38	14	13,17	13	32	23	10	10	12	14	10	11	20	25	15	19.2	10	23	40	11	18	17	21	20	11	J1	99.51	26.57
S89	SA	36,39	14	14,19	13	30	24	10	11	12	14	10	11	20	25	15	18.2	11	27	38	11	17	17	20	22	11	J1	63.59	38.16
S9	SA	36,38	14	12,19	13	29	23	10	11	12	14	10	12	20	25	14	19.2	11	25	39	11	19	17	21	21	12	J1	76.07	41.3
S90	SA	37,38	14	16,17	14	32	25	10	11	13	14	10	12	20	33	15	18	10	24	38	12	18	18	22	23	11	E1b1b	100	30.98
S91	SA	36,37	14	14,18	13	30	23	10	11	12	14	10	11	20	26	14	21.2	12	25	39	11	16	18	20	21	11	J1	99.99	44.18
S92	SA	37,37	14	13,18	13	29	24	10	11	12	14	10	11	20	25	14	18.2	11	25	37	11	19	18	24	21	11	J1	99.95	50.44
S93	SA	37,38	14	13,18	13	30	23	10	11	12	14	10	11	21	25	14	18.2	10	24	38	11	17	17	23	21	11	J1	100	52.26
S94	SA	37,38	14	13,18	13	30	23	10	11	12	14	10	11	20	24	14	17.2	11	25	38	11	18	18	21	21	11	J1	100	62.12
S95	SA	39,40	14	17,18	13	29	24	10	11	12	14	10	13	19	27	14	19.2	10	25	41	11	18	18	22	21	11	J1	95.1	28.96
S96	SA	35,39	14	13,17	13	31	23	10	11	12	14	10	11	20	25	14	19.2	11	23	38	11	17	17	20	21	11	J1	99.98	44.17
S97	SA	36,38	14	13,17	13	30	24	10	11	12	14	10	11	20	25	14	20.2	11	25	41	12	22	17	21	21	11	J1	98.42	31.56
S98	SA	36,38	14	13,19	13	30	23	10	11	12	14	10	11	19	25	14	18.2	10	25	40	11	18	17	19	22	11	J1	100	48.54
S99	SA	37,38	14	13,17	13	31	23	10	11	12	14	10	11	20	26	15	17.2	10	24	40	11	18	19	20	20	12	J1	100	45.38
W1	SA	37,38	14	13,20	13	30	23	10	11	13	14	10	11	20	25	14	18.2	10	26	40	12	18	18	21	21	11	J1	100	48.73
W10	SA	38,42,2	16	9,12	13	32	24	10	11	13	14	9	11	20	34	13	18	10	24	46	11	17	17	22	17	11	B	100	19.98
W100	SA	36,38	14	13,18	13	30	23	10	11	12	14	10	11	20	25	14	18.2	11	26	42	11	19	18	20	21	11	J1	100	59.3
W101	SA	37,39	16	11,15	13	29	25	11	11	13	14	11	10	20	33	15	15	11	23	44	12	18	18	18	23	12	R1a	100	52.1
W102	SA	37,38	11	16,17	13	30	24	10	12	13	14	10	11	21	36	15	13	10	24	39	12	19	17	19	21	11	E1b1b	99.85	22.5
W103	SA	34,40,2	15	9,12	12	27	23	10	11	13	14	11	12	20	32	14	16	10	23	37	11	14	17	23,2	20	11	UP	51.19	10.58
W104	SA	36,36	14	12,19	13	32	23	10	11	12	14	10	12	20	24	13	19.2	10	24	39	11	19	17	23	21	12	J1	99.93	29.9
W105	SA	36,37	14	12,20	13	30	23	11	11	12	14	10	12	20	24	14	18.2	10	26	40	11	18	18	21	21	11	J1	100	54.59
W106	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	24	39	11	18	18	22	22	11	J1	100	56.41
W107	SA	38,38	14	16,18	13	31	24	10	11	13	14	10	12	20	30	16	19	10	25	39	10	19	18	18	21	11	E1b1b	100	41.66
W108	SA	37,38	14	12,12	14	31	23	11	11	13	15	9	11	21	31	15	15	10	22	39	11	16	17	24	22	10	J2	99.48	26.92
W109	SA	38,39	14	16,17	13	31	25	10	11	13	14	10	12	20	29	16	19	10	25	39	10	19	16	18	21	12	E1b1b	100	33.24
W110	SA	36,39	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	17.2	11	25	38	11	19	17	21	21	11	J1	100	54.05
W111	SA	35,38	13	12,16	13	29	22	10	15	13	15	11	13	19	30	14	17	10	23	36	12	19	18	20	23	10	Q	100	22.47
W112	SA	37,37	14	13,19	13	30	23	11	11	13	14	10	11	20	25	14	19.2	10	24	40	12	18	18	23	21	11	J1	100	50.06
W113	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	12	20	25	14	18.2	10	25	40	12	18	18	21	21	11	J1	100	57.86
W114	SA	37,37	14	12,19	13	29	23	11	11	12	14	10	11	20	25	14	20.2	10	25	39	13	18	18	21	21	11	J1	97.98	43.27
W115	SA	39,39	14	16,17	13	31	25	10	11	13	14	10	12	20	29	16	19	10	25	39	10	19	16	18	21	12	E1b1b	100	33.24
W116	SA	37,38	14	13,17	13	30	23	10	11	12	14	10	11	20	25	14	18.2	11	25	38	11	18	18	21	21	11	J1	100	57.55
W117	SA	39,39	14	17,18	13	29	24	10	11	12	14	10	12	19	26	14	19.2	10	25	40	11	18	18	22	21	11	J1	98.6	32.78
W118	SA	38,41,2	15	11,12	13	31	24	10	11	14	15	10	11	21	31.2	13	17	12	25	43	11	18	19	22	17	11	B	99.95	20.4
W119	SA	37,37	15	13,18	13	30	22	10	11	12	14	10	12	20	28	16	17.2	10	24	40	11	19	18	22	20	11	J1	100	36.14
W120	SA	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	27	38	11	18	18	20	21	11	J1	100	52.79
W121	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	27	38	11	19	18	21	21	11	J1	100	57.93
W122	SA	36,39	13	15,16	14	30	23	11	16	13	15	9	13	19	32	16	16	11	26	38	11	18	17	21	23	11	T	100	38.25
W123	SA	36,39	13	15,15	14	30	23	11	16	13	15	9	13	19	32	15	16	11	26	38	11	17	17	21	22	11	T	100	33.94
W124	SA	36,39	13	15,16	14	32	23	11	16	13	15	9	13	19	32	16	16	11	26	38	11	18	17	21	23	11	T	100	28.8
W125	SA	36,38	14	13,18	13	30	23	10	11	12	14	10	11	20	26	14	18.2	11	25	39	11	20	17	22	21	11	J1	100	54.57
W126	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	19.2	10	25	41	13	18	18	21	21	11	J1	100	54.15
W127	SA	38,44,2	16	12,12	13	30	23	10	11	13	14	10	11	20	30	12	12	10	26	43	11	18	18	21	17	10	B	100	33.11
W128	SA</td																												

Sample ^a	Recruitment	DYS387S1	aDYS19	DYS385a,b	DYS389I	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4	Nevgen predicted haplogroup ^b	Haplogroup probability	Haplogroup fitness	
W55	SA	36,37	14	12,18	13	29	23	11	11	13	14	10	12	20	25	14	18.2	11	25	39	11	18	19	21	21	11	J1	100	53.17
W56	SA	39,39	14	17,17	13	31	24	10	11	13	14	11	12	20	29	16	19	10	26	39	10	18	18	19	21	11	E1b1b	99.98	30.81
W57	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	24	41	12	18	18	21	21	11	J1	100	56.41
W58	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	21	34	16	20	10	24	38	10	15	19	23	20	12	G	99.46	17.88
W59	SA	36,37	16	12,13	13	30	21	9	11	14	16	10	11	21	34	16	20	10	24	38	10	15	19	23	20	12	G	99.46	17.88
W60	SA	38,39	13	17,17	14	30	23	10	11	14	14	10	12	20	32	15	17	11	26	38	11	18	19	22	21	11	E1b1b	97.79	28.37
W61	SA	36,39	13	15,16	14	30	23	11	16	13	15	9	13	19	32	16	16	12	28	38	11	18	18	19	23	11	T	100	27.04
W62	SA	36,39	14	12,18	13	30	23	11	11	12	14	10	11	20	25	14	20.2	11	24	39	11	18	18	20	21	11	J1	100	50.4
W63	SA	34,40,2	15	8,12	12	27	23	10	11	13	14	11	12	20	31	14	17	10	24	37	11	17	17	22.2	20	11	UP	99.86	15.47
W64	SA	36,36	14	13,18	13	29	24	10	11	12	14	10	11	19	27	14	18.2	11	25	39	11	18	19	21	22	11	J1	99.97	40.82
W65	SA	39,44,2	15	11,11	13	31	23	11	11	13	14	10	13	20	32	13	18	11	25	38	11	17	17	21	17	11	B	100	40.87
W66	SA	36,37	14	18,19	12	29	23	10	11	13	14	10	12	20	35	15	15	11	22	37	12	20	18	21	20	12	E1b1b	100	32.71
W67	SA	39,39	14	17,18	13	29	24	10	11	12	14	10	12	19	26	14	19.2	10	25	40	11	18	18	22	21	11	J1	98.6	32.78
W68	SA	36,38	15	14,15	14	32	23	10	11	12	14	10	11	20	26	15	17.2	10	25	39	11	19	17	21	20	11	J1	94.23	29.98
W69	SA	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	24	40	12	18	18	21	21	11	J1	100	58.87
W70	SA	36,37	11	16,19	13	31	24	10	12	13	14	12	11	21	33	15	13	11	24	40	12	18	21	18	22	12	E1b1b	71.34	16.47
W71	SA	36,38	14	13,18	13	30	23	10	11	12	14	10	11	20	26	15	17.2	11	25	37	11	15	18	19	21	11	J1	100	53.98
W72	SA	37,37	14	19,19	13	31	23	11	11	12	14	10	12	20	27	14	19.2	10	27	37	11	18	18	22	21	11	J1	99.99	39.14
W73	SA	37,37	15	13,18	13	30	22	10	11	12	14	10	11	20	28	15	17.2	10	24	41	11	19	18	21	20	12	J1	100	37.29
W74	SA	39,39	14	16,17	13	31	24	10	11	14	14	10	12	20	29	15	18	10	25	39	10	19	18	18	21	11	E1b1b	100	41.51
W75	SA	38,40	14	18,19	13	29	24	10	11	12	14	10	14	19	27	14	18.2	11	25	39	11	17	20	23	21	11	J1	99.71	24.57
W76	SA	36,37	11	16,19	13	31	24	10	12	13	14	12	11	21	32	15	13	11	22	38	12	18	19	19	21	12	E1b1b	100	20.61
W78	SA	36,40	15	10,14	13	29	25	11	11	13	14	12	10	20	32	16	16	11	23	44	12	19	20	16	23	13	R1a	100	37.67
W79	SA	39,40	14	16,17	13	31	24	10	11	13	14	10	12	20	29	15	18	10	26	39	10	19	18	21	21	11	E1b1b	100	40.68
W80	SA	38,41	14	16,17	13	31	24	10	11	13	14	10	12	20	30	15	18	10	25	39	10	19	19	18	21	10	E1b1b	100	35.37
W81	SA	37,38	15	12,12	14	32	21	10	11	13	16	10	12	22	34	15	17	10	25	38	10	17	19	22	19	11	A	89.74	18.57
W82	SA	36,37	14	13,18	13	30	23	12	11	12	14	10	11	null	25	14	18.2	10	26	38	11	20	17	21	21	11	J1	100	48.15
W83	SA	39,39	15	17,17	13	31	24	10	11	13	14	11	12	20	29	16	18	10	26	39	10	18	18	21	21	11	E1b1b	96.74	30.53
W84	SA	36,39	13	15,16	14	30	23	11	16	13	15	9	13	19	32	15	16	10	26	38	11	17	18	21	21	11	T	100	39.24
W85	SA	35,36	15	14,17	13	29	23	11	15	13	14	11	11	20	26	16	17	10	28	40	11	22	19	20	22	11	Q	54.12	18.44
W86	SA	37,37	14	13,18	14	32	23	11	11	13	14	10	12	20	25	14	18.2	11	25	36	11	17	17	21	22	11	J1	100	38.74
W87	SA	39,39	14	16,17	13	31	24	10	11	13	14	10	12	20	29	15	18	10	26	39	10	20	18	21	21	11	E1b1b	100	40.24
W89	SA	37,37	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	18.2	10	25	40	11	18	19	20	21	11	J1	100	59.95
W90	SA	36,36	15	12,13	13	29	21	9	11	14	15	10	11	21	35	15	18	10	26	38	12	18	16	19	19	12	UP	50.91	18.64
W91	SA	37,37	14	13,19	13	29	23	11	11	12	14	10	11	20	25	14	18.2	10	25	38	11	18	17	20	21	11	J1	100	51.96
W92	SA	36,36	14	13,18	14	31	23	11	11	12	14	10	12	20	26	14	18.2	11	24	39	12	18	18	21	20	11	J1	100	46.02
W94	SA	37,37	14	12,18	12	29	23	11	12	14	14	10	11	20	24	14	18.2	10	26	38	12	18	17	22	21	10	J1	99.95	29.96
W95	SA	37,37	14	12,19	13	31	23	10	11	12	14	10	11	20	25	14	19.2	11	26	40	11	17	19	22	21	11	J1	99.99	49.59
W96	SA	36,38	14	13,19	13	30	23	11	11	12	14	10	11	20	25	15	18.2	11	25	39	12	18	19	21	21	11	J1	100	53.96
W97	SA	37,38	15	12,20	13	29	22	10	11	13	14	10	11	20	25	15	17.2	11	27	40	11	20.2	17	18	20	10	J1	98.51	26.13
W98	SA	39,40	14	16,16	13	31	24	10	11	13	14	10	12	20	30	15	18	10	25	39	10	19	18	22	21	11	E1b1b	100	41.08
W99	SA	37,37	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	25	40	12	19	18	22	21	12	J1	100	51.44
C1	UK	36,37	14	13,18	13	30	23	11	11	12	14	10	12	20	23	15	18.2	10	26	38	9	17	17	18	21	11	J1	100	40.45
C10	UK	39,39	14	16,16	13	31	24	10	11	13	14	10	12	20	30	17	18	11	25	39	10	17	16	18	21	12	E1b1b	100	33.46
C11	UK	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	19.2	10	25	40	12	18	18	21	21	11	J1	100	59.22
C12	UK	36,37	14	13,20	13	31	23	11	11	12	14	10	11	21	24	14	18.2	10	25	40	11	18	18	21	21	11	J1	100	48.63
C14	UK	36,37	14																										

Sample ^a	Recruitment	DYS387S1a	DYS385a,b	DYS389I	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4	Neogen predicted haplogroup ^b	Haplogroup probability	Haplogroup fitness		
E14	UK	37,37	14	13,16	13	29	23	9	11	12	14	10	11	20	26	15	17.2	10	25	36	11	19	17	20	20	11	J1	99.2	29.17
E15	UK	36,38	14	14,17	13	31	24	10	11	12	14	10	11	20	26	15	17.2	10	25	36	11	19	17	20	20	11	J1	99.95	42.7
E16	UK	36,37	14	14,17	13	31	24	10	11	12	14	10	11	20	26	15	17.2	10	26	37	11	17	16	20	20	11	J1	99.99	39.12
E17	UK	35,37	16	14,14	13	29	23	10	11	13	16	10	11	20	29	15	17	12	21	36	9	19	16	21	21	11	G	69.18	32.92
E2	UK	35,35	14	17,18	13	30	24	10	12	12	14	10	11	20	35	15	17	11	25	42	12	18	17	20	21	12	E1b1b	95.57	32.17
E3	UK	36,38	14	12,18	14	31	23	10	11	12	14	10	11	20	26	14	18.2	11	24	37	12	20	17	20	20	11	J1	99.82	35.91
E4	UK	35,37	14	17,18	13	32	24	11	11	13	14	10	13	19	31	14	14	11	23	38	12	18	19	21	23	11	E1b1b	100	22.81
E5	UK	35,36	14	17,18	13	30	24	11	12	12	14	10	11	20	35	15	17	10	25	42	12	18	17	20	23	12	E1b1b	99.64	27.22
E6	UK	37,38	15	17,18	13	30	23	10	11	14	14	10	11	20	30	16	14	11	26	37	11	18	16	21	24	13	E1b1b	98.98	22.72
E7	UK	37,37	14	13,18	13	29	23	11	11	12	14	10	11	20	25	14	18.2	11	23	37	11	17	19	21	21	11	J1	100	52.21
E8	UK	36,39	15	16,19	14	32	21	10	11	12	14	11	12	21	28	15	16	10	31	41	11	17	14	19	21	12	E1b1a	100	29.21
E9	UK	40,40	15	13,16	14	29	23	10	11	13	15	9	13	22	34	15	15	11	22	40	11	18	16	19	19	11	J2	100	41.88
N1	UK	37,38	14	14,16	13	30	23	10	11	13	14	10	11	19	27	15	17.2	11	25	38	11	17	17	20	21	11	J1	99.54	38.75
N10	UK	36,37	15	13,18	13	29	23	10	11	12	14	10	11	20	25	14	20.2	11	25	40	11	18	18	21	21	11	J1	99.86	49.37
N11	UK	36,39	14	13,17	14	30	23	10	11	14	14	9	11	17.4	31	16	14	11	22	42	11	17	17	21	20	11	J2	99.98	30.81
N2	UK	36,38	14	13,18	13	30	23	10	11	13	14	10	11	20	26	15	17.2	12	25	38	11	17	18	23	21	11	J1	100	57.17
N3	UK	36,37	14	13,18	13	29	23	10	11	12	14	10	11	20	24	14	19.2	11	24	39	11	18	18	20	21	11	J1	99.98	55.27
N4	UK	36,37	14	13,18	13	30	23	11	11	12	14	10	12	20	26	14	19.2	11	25	40	11	17	18	20	21	11	J1	100	58.21
N5	UK	38,38	15	11,16	14	29	24,3	10	14	14	10	12	21	31	14	18	11	24	40	12	20	19	21	21	10	UP	100	31.06	
N7	UK	37,38	14	13,17	12	30	23	11	11	12	14	10	11	20	25	14	18.2	11	25	39	11	18	19	22	21	11	J1	100	46.99
N8	UK	37,37	14	13,19	13	30	23	11	11	12	14	10	11	20	25	14	19.2	10	26	41	12	18	18	21	21	11	J1	100	61.41
N9	UK	37,38	15	15,16	12	29	22	10	11	13	15	10	11	20	28	17	16	11	21	38	9	17	18	20	22	11	G	100	52.04
S1	UK	37,39	16	12,14	13	30	25	10	11	13	14	11	10	20	33	15	16	11	24	43	12	19	18	17	23	13	R1a	100	48.81
S10	UK	36,37	14	13,17	13	29	22	11	11	12	14	10	11	20	25	13	17.2	11	25	38	11	18	18	20	21	11	J1	100	42.44
S11	UK	37,39	14	13,18	13	31	24	10	11	12	14	10	11	20	25	14	18.2	10	25	38	11	17	18	22	21	11	J1	100	51.76
S12	UK	37,39	15	13,16	13	29	23	9	11	12	15	9	11	21	30	15	14	10	22	39	12	19	16	20	21	12	J2	99.84	52.94
S13	UK	38,38	16	11,13	13	30	25	10	11	13	14	11	11	19	31	14	15	11	23	41	12	17	17	23	13	R1a	100	38.64	
S14	UK	36,38	14	13,18	13	30	23	10	11	12	14	10	11	20	24	14	19.2	11	25	36	11	18	20	22	21	11	J1	100	54.22
S15	UK	37,37	14	13,18	13	31	24	10	11	12	14	10	11	20	26	15	17.2	10	23	40	11	19	20	21	20	11	J1	100	37.71
S16	UK	38,38	14	17,18	14	30	24	10	11	12	14	10	13	19	27	15	18.2	10	25	41	11	18	19	21	21	11	J1	78.42	27.28
S17	UK	38,38	14	13,18	13	32	23	10	10	12	14	10	11	19	26	15	17.2	10	23	40	11	18	17	21	20	11	J1	100	29.65
S18	UK	36,37	14	14,18	13	30	23	11	11	12	14	10	13	20	26	14	19.2	11	25	39	11	19	18	21	21	11	J1	100	42.9
S19	UK	38,39	14	13,16	13	29	22	9	11	12	14	9	11	21	32	15	16	10	23	37	10	16	18	20	23	11	J2	99.86	45.84
S2	UK	37,39	15	13,19	13	30	22	10	11	12	14	10	11	20	28	15	18.2	10	24	41	11	19	18	21	20	12	J1	99.99	35.07
S21	UK	38,40	13	16,18	13	31	24	8	11	13	14	10	12	20	31	15	16	10	24	40	10	18	16	20	23	9	E1b1b	100	26.41
S22	UK	35,36,43	15	14,16	13	29	23	11	15	13	14	11	11	20	26	16	18	10	27	40	11	21	19	20	22	11	Q	57.03	19.11
S23	UK	37,41	14	13,17	14	31	23	11	11	12	14	10	11	19	25	14	18.2	11	25	40	12	17	16	22	22	11	J1	99.77	35.84
S24	UK	37,39	15	17,18	13	31	21	11	11	13	14	10	12	21	27	15	16	10	26	38	11	20	16	21	21	12	E1b1a	99.97	32.11
S25	UK	36,38	14	13,19	13	30	23	10	11	12	14	10	11	19	25	14	18.2	11	25	40	11	18	17	21	21	11	J1	100	53.62
S26	UK	37,38	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	11	25	37	11	19	17	21	21	11	J1	100	60.72
S27	UK	36,39	14	14,16	14	32	24	10	13	13	14	9	11	19	35	17	16	10	24	36	12	18	16	20	22	11	T	100	39.12
S3	UK	35,1,36	14	13,17	13	29	24	11	11	12	14	10	11	20	26	13	18.2	11	25	39	11	18	18	21	21	11	J1	100	46.41
S4	UK	37,38	15	13,13	13	30	23	11	11	12	14	10	11	20	26	15	19.2	11	26	39	11	18	18	21	21	11	J1	100	45.48
S5	UK	37,37	16	13,16	13	29	23	10	13	13	14	9	11	19	35	15	17	10	22	36	11	17	16	20	21	11	T	100	41.97
S6	UK	37,37	14	13,18	13	30	23	11	11	12	14	10	11	19	25	14	18.2	10	24	40	12	18	19	21	21	11	J1	100	57.82
S7	UK	36,39	14	13,13	13	31	23	10	11	12	14	10	11	20	24	14	18.2	11	25	40	11	18	20	21	21	11	J1	100	50.25
S8	UK	36,37	13	13,14	14	30	24	10	11	13	14	10	11	20	32	16	17	11											

Sample ^a	Recruitment	Nevgen predicted haplogroup ^b																				Haplogroup probability	Haplogroup fitness						
		DYF387S1	aDYS19	DYS385a,b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS449	DYS456	DYS458	DYS460	DYS481	DYS518	DYS533	DYS570	DYS576	DYS627	DYS635	YGATAH4			
W5	UK	37,37	14	13,18	13	30	23	11	11	12	14	10	11	20	25	14	18.2	10	24	40	12	18	19	21	11	J1	100	57.82	
W6	UK	39,40	14	17,17	13	30	24	10	11	13	14	10	13	20	29	15	18	10	25	39	10	19	18	21	11	E1b1b	99.99	38.75	
W7	UK	35,37	14	13,19	13	29	23	11	11	12	14	10	11	20	25	14	19.2	11	25	40	11	20	18	20	21	11	J1	99.98	52.49
W8	UK	36,38	14	14,18	14	32	23	10	11	12	14	10	11	20	26	14	18.2	11	26	40	11	19	18	20	21	11	J1	100	44.2
W9	UK	37,37	14	13,19	13	30	23	11	11	12	14	10	11	21	25	14	18.2	10	26	38	12	18	18	22	21	11	J1	100	54.35

^aC: Central; E: Eastern; N: Northern; S: Southern; W: Western. ^bUP: unpredicted.

Table S2: Diversity summary statistics for Y-STR haplotypes, considering country of recruitment.

pop	n	Yfiler® plus						Yfiler® comparison			
		No. unique hts	No. pair hts	No. trio hts	HMP	Haplotype Diversity	% unique hts DC	HMP	Haplotype Diversity	% unique hts DC	
All	597	541	25	2	0.0018	0.9998	90.6	95.3	0.0048	0.9969	63.0
All-SA	503	454	23	1	0.0022	0.9998	90.3	95.0	0.0052	0.9967	61.4
All-UK	94	92	1		0.0109	0.9998	97.9	98.9	0.0129	0.9977	88.3
C	125	117	4		0.0085	0.9995	93.6	96.8	0.0245	0.9834	62.4
C-SA	107	101	3		0.0099	0.9995	94.4	97.2	0.0263	0.9829	59.8
C-UK	18	18			0.0556	1.0000	100.0	100.0	0.0617	0.9935	88.9
E	110	110			0.0091	1.0000	100.0	100.0	0.0099	0.9992	90.9
E-SA	93	93			0.0108	1.0000	100.0	100.0	0.0117	0.9991	91.4
E-UK	17	17			0.0588	1.0000	100.0	100.0	0.0657	0.9926	88.2
N	106	92	7		0.0107	0.9987	86.8	93.4	0.0176	0.9917	57.5
N-SA	96	84	6		0.0117	0.9987	87.5	93.8	0.0202	0.9901	55.2
N-UK	10	10			0.1000	1.0000	100.0	100.0	0.1000	1.0000	100.0
S	140	125	6	1	0.0081	0.9991	89.3	94.3	0.0106	0.9965	72.1
S-SA	114	99	6	1	0.0102	0.9986	86.8	93.0	0.0132	0.9955	67.5
S-UK	26	26			0.0385	1.0000	100.0	100.0	0.0414	0.9969	92.3
W	116	106	5		0.0094	0.9993	91.4	95.7	0.0113	0.9973	79.3
W-SA	93	85	4		0.0119	0.9988	91.4	94.6	0.0135	0.9972	77.4
W-UK	23	23			0.0435	1.0000	100.0	100.0	0.0435	1.0000	100.0

pop: population; n: number of individuals; ht: haplotype; HMP: haplotype match probability; DC: discrimination capacity; C: central; E: eastern; N: northern; S: southern; W: western.

SA: Saudi Arabia-recruited; UK: UK-recruited. Yfiler® comparison: lists statistics considering only those STRs included in the 17-STR Yfiler® kit

Table S3: Predicted haplogroup distributions and diversities, considering country of recruitment.

pop	n	Predicted haplogroup															$h \pm s.d.$
		A	B	E1b1a	E1b1b	G	H	J1	J2	L	Q	R1a	R1b	T	UP		
All	597	5	10	9	66	8	2	424	16	6	8	14	5	15	9	0.481 ± 0.024	
All-SA	503	4	10	6	48	5	2	374	9	5	7	10	5	13	5	0.436 ± 0.027	
All-UK	94	1		3	18	3		50	7	1	1	4		2	4	0.676 ± 0.046	
C	125	1			7	1		107	2	1	1	1	1	1	2	0.265 ± 0.052	
C-SA	107	1			3	1		96	1	1	1		1	1	1	0.195 ± 0.052	
C-UK	18				4			11	1			1			1	0.601 ± 0.113	
E	110	2	3	3	20	3	2	51	7	4	3	7	2	3		0.745 ± 0.037	
E-SA	93	2	3	2	15	2	2	43	5	4	3	7	2	3		0.752 ± 0.041	
E-UK	17			1	5	1		8	2							0.713 ± 0.083	
N	106			1	7	2		91	2				1		2	0.260 ± 0.056	
N-SA	96			1	7	1		84	1				1		1	0.231 ± 0.056	
N-UK	10					1			7	1					1	0.533 ± 0.180	
S	140		2	4	9			113	3		1	3	1	4		0.344 ± 0.052	
S-SA	114		2	3	7			97	1			1	1	2		0.273 ± 0.054	
S-UK	26			1	2			16	2		1	2		2		0.618 ± 0.106	
W	116	2	5	1	23	2		62	2	1	3	3		7	5	0.671 ± 0.041	
W-SA	93	1	5		16	1		54	1		3	2		7	3	0.629 ± 0.050	
W-UK	23	1		1	7	1		8	1			1			3	0.802 ± 0.060	
Abu-Amero et al. (2009) ^a	157	0	3	12	12	5	3	63	25	3	4	8	3	8	8 ^b	0.796 ± 0.026	

pop: population; n: number of individuals; ht: haplotype; UP: unpredicted haplogroup; C: central; E: eastern; N: northern; S: southern; W: western.
 SA: Saudi Arabia-recruited; UK: UK-recruited.

^a: Abu-Amero et al. (2009) Saudi Arabian Y-chromosome diversity and its relationship with nearby regions. *BMC Genet* 10: 59.

^b: Haplogroups are known from SNP typing for the 8 samples of Abu-Amero placed here in the 'UP' category, so not truly 'unpredicted'.