



# Y-Chromosome and mtDNA Genetics Reveal Significant Contrasts in Affinities of Modern Middle Eastern Populations with European and African Populations

## Citation

Badro, Danielle A., Bouchra Douaihy, Marc Haber, Sonia C. Youhanna, Angélique Salloum, Michella Ghassibe-Sabbagh, Brian Johnsrud, et al. 2013. Y-chromosome and mtDNA genetics reveal significant contrasts in affinities of modern Middle Eastern populations with European and African populations. PLoS ONE 8(1): e54616.

## Published Version

doi:10.1371/journal.pone.0054616

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:10613666>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Y-Chromosome and mtDNA Genetics Reveal Significant Contrasts in Affinities of Modern Middle Eastern Populations with European and African Populations

Danielle A. Badro<sup>1</sup>, Bouchra Douaihy<sup>1,9</sup>, Marc Haber<sup>1,2,9</sup>, Sonia C. Youhanna<sup>1</sup>, Angélique Salloum<sup>1</sup>, Michella Ghassibe-Sabbagh<sup>1</sup>, Brian Johnsrud<sup>3</sup>, Georges Khazen<sup>1</sup>, Elizabeth Matisoo-Smith<sup>4</sup>, David F. Soria-Hernanz<sup>2,5</sup>, R. Spencer Wells<sup>5</sup>, Chris Tyler-Smith<sup>6</sup>, Daniel E. Platt<sup>7</sup>, Pierre A. Zalloua<sup>1,8\*</sup>, The Genographic Consortium<sup>1</sup>

**1** The Lebanese American University, Chouran, Beirut, Lebanon, **2** Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona, Spain, **3** Modern Thought and Literature, Stanford University, Stanford, California, United States of America, **4** Allan Wilson Centre for Molecular Ecology and Evolution, University of Otago, Dunedin, New Zealand, **5** The Genographic Project, National Geographic Society, Washington, DC, United States of America, **6** The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom, **7** Computational Biology Centre, IBM TJ Watson Research Centre, Yorktown Heights, New York, United States of America, **8** Harvard School of Public Health, Boston, Massachusetts, United States of America

## Abstract

The Middle East was a funnel of human expansion out of Africa, a staging area for the Neolithic Agricultural Revolution, and the home to some of the earliest world empires. Post LGM expansions into the region and subsequent population movements created a striking genetic mosaic with distinct sex-based genetic differentiation. While prior studies have examined the mtDNA and Y-chromosome contrast in focal populations in the Middle East, none have undertaken a broad-spectrum survey including North and sub-Saharan Africa, Europe, and Middle Eastern populations. In this study 5,174 mtDNA and 4,658 Y-chromosome samples were investigated using PCA, MDS, mean-linkage clustering, AMOVA, and Fisher exact tests of  $F_{ST}$ 's,  $R_{ST}$ 's, and haplogroup frequencies. Geographic differentiation in affinities of Middle Eastern populations with Africa and Europe showed distinct contrasts between mtDNA and Y-chromosome data. Specifically, Lebanon's mtDNA shows a very strong association to Europe, while Yemen shows very strong affinity with Egypt and North and East Africa. Previous Y-chromosome results showed a Levantine coastal-inland contrast marked by J1 and J2, and a very strong North African component was evident throughout the Middle East. Neither of these patterns were observed in the mtDNA. While J2 has penetrated into Europe, the pattern of Y-chromosome diversity in Lebanon does not show the widespread affinities with Europe indicated by the mtDNA data. Lastly, while each population shows evidence of connections with expansions that now define the Middle East, Africa, and Europe, many of the populations in the Middle East show distinctive mtDNA and Y-haplogroup characteristics that indicate long standing settlement with relatively little impact from and movement into other populations.

**Citation:** Badro DA, Douaihy B, Haber M, Youhanna SC, Salloum A, et al. (2013) Y-Chromosome and mtDNA Genetics Reveal Significant Contrasts in Affinities of Modern Middle Eastern Populations with European and African Populations. PLoS ONE 8(1): e54616. doi:10.1371/journal.pone.0054616

**Editor:** David Caramelli, University of Florence, Italy

**Received:** June 21, 2012; **Accepted:** December 13, 2012; **Published:** January 30, 2013

**Copyright:** © 2013 Badro et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The Genographic Project is supported by funding from the National Geographic Society, IBM and the Waitt Family Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** IBM has no marketed products associated with the contents of this paper. IBM has no income derived from contracts related to consultancy or capital or intellectual assets in this research. The affiliation of an author (DP) with IBM does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: pierre.zalloua@lau.edu.lb

These authors contributed equally to this work.

Membership of the The Genographic Consortium is provided in the Acknowledgments section.

## Introduction

As a crossroad between Africa, Arabia, Asia, and Europe, the Levant has been a primary historical stepping stone in the first modern human expansions out of Africa and for later migrations into and out of Europe, Asia, and Africa [1–7]. As such, it has also become a land of remarkable human diversity. The earliest fossil and archaeological evidence of modern humans outside of the African continent are from the Levant, presumably indicating a migration via the northern route, and date to 125–95 kya [8,9]. Additionally, genetic studies suggest that the initial peopling of

Eurasia occurred through the northern Levantine (modern day Lebanon and Syria) route [10–12]. Two proposed routes chart the dispersal of anatomically modern humans out of the African continent: (1) a northern route, reaching west and central Asia through the Sinai Peninsula and the Levant, and (2) a southern route via the Bab el-Mandeb Strait and along the south Asian coast, ultimately reaching Australia [13–15].

While the out-of-Africa migrations have been major determining factors, other migratory events have strongly influenced genetic marker distributions throughout the Levant and the surrounding geographical areas. During the last glacial maximum

(LGM, 26.5–19 kya), most of the Levant was an uninhabitable desert, with forested hills in Levantine Mediterranean coastal areas [7]. The genetics of the modern Levant were largely determined by subsequent repopulation (especially during the Neolithic agricultural revolution) and mass movements associated with empire building. Neolithic expansions in particular, beginning around 10 kya, induced gene flow between the Fertile Crescent and Europe, which shaped the genetic structure of both regions [16–22].

Most genetic studies of the Levant as a geographical area have focused exclusively on either Y-chromosome [23–27] or mitochondrial markers [28]. Further, contrasts between Y-chromosome and mtDNA data provide distinct insights into human expansions unavailable to somatic genome analyses [29]. While comparative analyses among the two marker types have been undertaken in the Middle East and Africa [30–34], none of these studies have explored the contrasting relationships of expansions throughout Europe, North Africa, the Levant and Arabian Peninsula after the LGM. Building on a previous study that reported phylogeographic characteristics of Y-chromosome markers in the Levantine region [25], we now compare and contrast Y and mtDNA phylogeographic distributions in the Levant and investigate the affinities of Middle Eastern populations with European and African populations.

## Materials and Methods

### Ethics statement

The samples were collected from donors after they had given their written informed consent to the project and to the data analysis, which was approved by the IRB of the Lebanese American University.

### mtDNA data

3,663 mtDNA records collected from the literature represented populations from Burkina-Faso [35], Cyprus [36], Egypt [37], Ethiopia [38], France [39,40], Greece [36], Iraq [31], Jordan [30], Kenya [41], Libyan Sahara [42], Mali [35], Morocco [43,44], Niger [35], Saudi Arabia [45], Slovakia [46], Tunisia [44,47], and Yemen [38,48]. In addition to this data, we added 1,511 new samples from Lebanon, Libya, Jordan, Palestine, and Syria. Samples were collected from unrelated blood donors from five countries. Surname repetitions were avoided and used as a criterion for absence of relatedness among volunteers, appropriate for Y-chromosome analysis. All demographic data were provided by self-assignment.

Given the broad cultural and genetic diversity in the region, terms such as “Middle East” may be problematical. Historically, the term evolved during the era of European Imperialism, and included all lands between Arabia and India, but came to include Turkey through Saudi Arabia, extending east through Afghanistan and Pakistan. In this report, “Middle Easterners” refers to Iraqis, Jordanians, Lebanese, Palestinians, Saudis, Syrians, and Yemenis. The Greek data represent the Southeastern Europe region for mtDNA analyses, and is labelled “Southeastern Europe” in the rest of this report. France represents Western Europe mtDNA, and is labelled “Western Europe” throughout the rest of this report. These are reported in Table S1. All haplogroups were reduced to the most informative sets for the purpose of homogeneous representation and comparative analyses. mtDNA haplogroup frequencies are displayed in Table 1 and shown as pie charts in Figure 1.

### Y-Chromosome data

1,774 previously published Y-chromosome records were obtained from the literature representing populations from the Balkans [49], Burkina-Faso [50], Ethiopia [50], Italy [51], Kenya [50], Saudi Arabia [52], Slovakia [53], and Yemen [52]. In addition 2,884 previously published data from our laboratory representing populations from Cyprus, Egypt, Lebanon, Libya, Jordan, Morocco, Palestine, Syria and Tunisia were added to this study. The Italian Y-chromosome samples represent Western Europe in this study, and are labelled “Western European” through the rest of this report. The Balkan samples represent the Southeastern Europe region, and are labelled “Southeastern European” in the rest of this report. These are reported in Table S2, with haplogroup frequencies reported in Table S3. The geographical haplogroup frequency distributions are displayed in Figure S1. Haplogroups of the Saudi, Yemeni, and Slovak populations were not available, thus we have predicted those haplogroups using the populations haplotypes and the online haplogroup prediction tool [54,55]: [www.hprg.com/hapest5/hapest5a/hapest5.htm](http://www.hprg.com/hapest5/hapest5a/hapest5.htm). We have also computed STR-predicted Y haplogroups across populations that had been SNP defined to ascertain STR-based haplogroup assignment accuracy, and identify geographically correlated trends in assignment error rates that may impact our conclusions.

### mtDNA sequencing and in silico prediction of haplogroups

Total DNA was extracted from the peripheral leukocyte fraction of whole blood drawn in EDTA anticoagulant or cheek swab samples using a standard phenol/chloroform extraction procedure. The hypervariable region I (HVS-I) was amplified using primers designed by Maca-Meyer et al. [12]. Amplified HVS-I products were sequenced using a forward primer at position 15876 and a reverse primer at position 639 with ABI Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and analysed on an Applied Biosystems 3130 xl Genetic Analyser.

Mutations in the HVS-I region were defined by aligning and comparing the sequences to the revised Cambridge Reference Sequence (rCRS) using the SeqScape software.

mtDNA haplogroups were predicted using the Genographic Project’s online haplogroup prediction tool: [nnhgttool.nationalgeographic.com](http://nnhgttool.nationalgeographic.com).

### mtDNA Genotyping of samples

Haplogroup affiliations were confirmed using the Taqman approach with customized primers and probe sets to identify the SNPs listed in Table S4 (Applied Biosystems). Samples with incompatible prediction and Taqman results were excluded from the study. Mitochondrial nomenclature was assigned according to prior studies since established as standards [38,56–59]. Data archiving was manually organized and edited.

### Reduction to most informative derived set of Haplogroups

MtDNA Haplogroups reported in the literature were updated and reconciled to the 2009 phylogeny reported by van Oven et al. [59]. Construction of the most informative derived set was achieved by identifying the maximum level of resolution shared across all included studies. If some subhaplogroup markers were not typed in any given study, but no samples in that study resolved to a less-derived paralog, then the most derived resolution was retained for the constructed most informative derived set.

**Table 1.** mtDNA Haplogroup frequencies of 1509 newly sequenced Levantine samples and 3665 samples collected from the literature.

Country	n	L*	L3*	M*	D	N*	R*	RO	HV	H	U	References
<b>Levant</b>												
Lebanon	979	0.6	1.4	1.4	0.2	10.5	21.2	5.0	8.5	29.9	21.1	This study
Syria	234	3.4	2.6	0.9	0.0	11.1	25.6	2.6	9.0	24.4	20.5	This study
Palestine	120	5.0	2.5	3.3	0.0	8.3	27.5	5.8	8.3	25.0	14.2	This study
Cyprus	79	3.8	0.0	2.5	0.0	12.7	17.7	0.0	0.0	22.8	40.5	Irwin et al. [36]
<b>Europe</b>												
France	871	0.6	0.3	0.1	0.0	5.4	18.0	0.5	6.9	45.4	22.8	Dubut et al. [39] Richard et al. [40]
Greece	372	0.0	0.0	1.1	0.0	9.1	19.4	1.9	5.9	42.2	20.4	Irwin et al. [36]
Slovakia	200	1.0	0.0	1.0	0.0	10.0	20.0	0.0	4.0	47.0	17.0	Malyarchuk et al. [46]
<b>Arabian Peninsula</b>												
Jordan	290	4.1	5.9	1.4	0.3	10.7	17.9	3.4	6.9	26.2	23.1	This study Gonzales et al. [30]
Iraq	52	0.0	0.0	1.9	0.0	0.0	34.6	5.8	11.5	17.3	28.8	Al-Zahery et al. [31]
Saudi Arabia	539	6.5	3.5	7.1	0.0	12.2	28.0	18.2	0.7	8.7	15.0	Abu-Amero et al. [45]
Yemen	300	22.0	16.3	5.3	0.0	6.7	21.7	9.3	2.3	4.7	11.7	Kivisild et al. [38] Cerny et al. [48]
<b>North Africa</b>												
Egypt	278	8.3	12.2	6.8	0.0	10.4	21.6	17.27	6.1	0.0	13.7	Saunier et al. [37]
Libya	31	6.5	25.8	0.0	0.0	3.2	12.9	6.5	9.7	25.8	9.7	This study
Libyan Sahara	129	21.7	11.6	1.6	0.0	0.0	0.0	0.0	3.9	61.2	0.0	Ottoni et al. [42]
Morocco	137	18.2	10.9	7.3	0.0	2.2	10.2	3.6	7.3	23.4	16.8	Turchi et al. [44] Harich et al. [43]
Tunisia	160	20.6	14.4	1.9	0.0	0.6	7.5	0.0	7.5	31.3	16.3	Cherni et al. [47] Turchi et al. [44]
<b>East sub Saharan</b>												
Ethiopia	232	36.6	18.1	14.7	0.0	3.9	5.2	11.6	2.6	0.9	6.5	Kivisild et al. [38]
Kenya	84	47.6	44.0	4.8	0.0	0.0	1.2	1.2	0.0	0.0	1.2	Brandstätter et al. [41]
<b>West sub Saharan</b>												
Burkina-Faso	40	35.0	5.0	17.5	0.0	0.0	0.0	0.0	20.0	22.5	0.0	Pereira et al. [35]
Mali	21	28.6	0.0	19.0	0.0	0.0	0.0	0.0	0.0	52.4	0.0	Pereira et al. [35]
Niger	25	44.0	32.0	4.0	0.0	0.0	0.0	0.0	12.0	4.0	4.0	Pereira et al. [35]

doi:10.1371/journal.pone.0054616.t001

Further, HVS-I regions reported by different sources varied. The range representing the largest common subset of HVS-I SNPs reported included 16090 through 16365. Identified SNPs are reported in Table S5.

### Statistical analyses

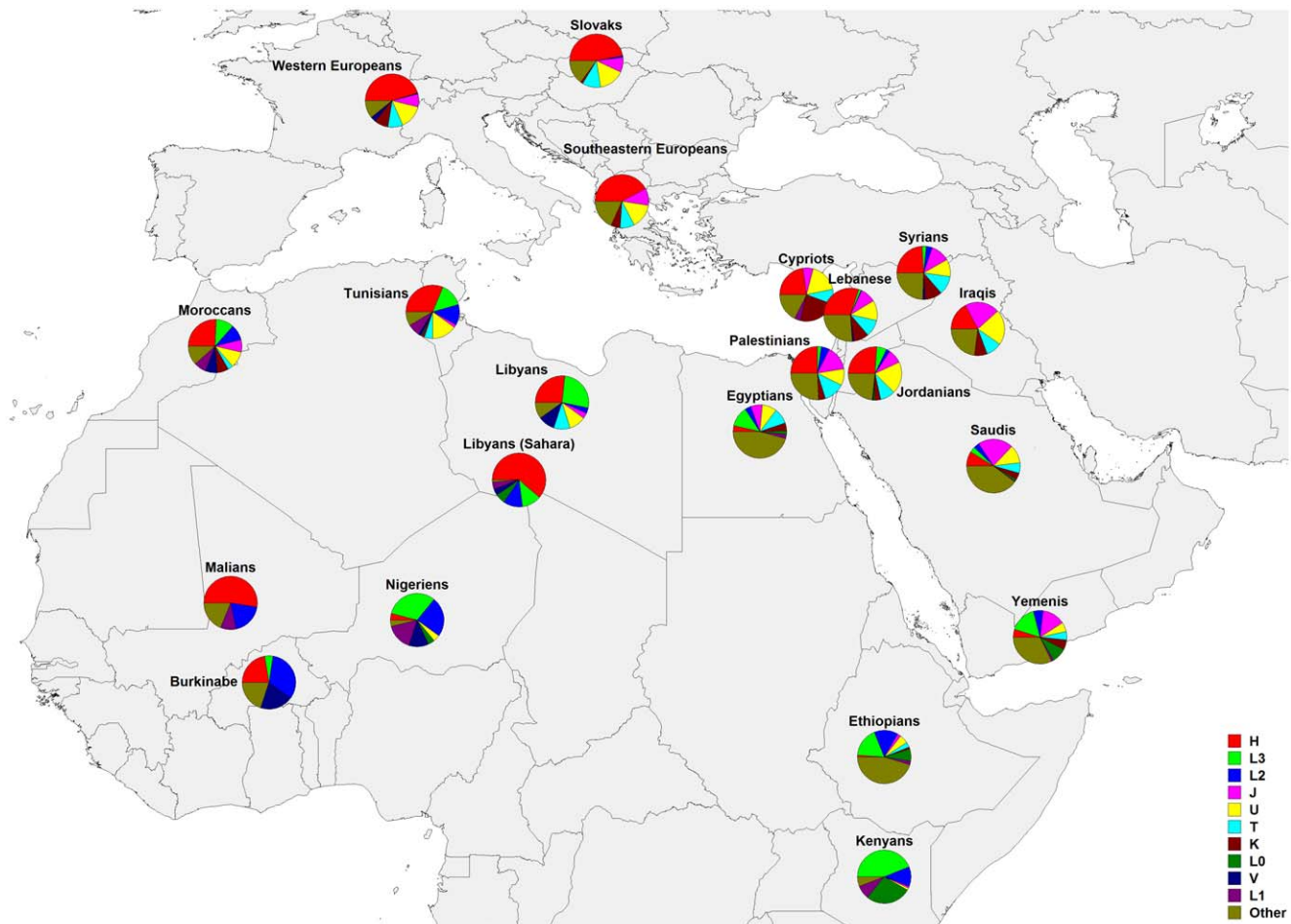
**Fisher Exact Tests.** Fisher exact tests were performed for haplogroup frequencies within populations. These tests were performed against a background of all populations (Table S6), as well as among Middle Eastern populations (from Iraq, Jordan, Lebanon, Palestine, Saudi Arabia, Syria, and Yemen) only (Table S7), with very low-power tests excluded.

**PCA.** Numbers of samples bearing mtDNA and Y-Chromosome reduced haplogroups within each population, and relative haplogroup frequencies within populations, were computed using R [60]. Principal Component Analysis was computed using prcomp in R [61]. Results were displayed with principal component contributions from each haplogroup using biplot. Agglomerative clustering with mean linkage (UPGMA) was

applied to Euclidean distances computed between relative frequency vectors for each population using agnes and displayed in Figure 2 for mtDNA Haplogroups and Figure S2 for Y-Haplogroups. These dendrograms should not be taken as population histories, but rather provide a repeatable description of population similarities also visible in the PCA.

**MDS.** HVS-I SNPs were constructed against CRS [62] as revised rCRS [63], and the subrange common to all publications was selected. ARLEQUIN [64] was employed to compute  $F_{ST}$ 's [65], which were used as distances for non-metric MDS analysis [66], as implemented in isoMDS [61] in R. Agglomerative clustering with mean linkage was applied to the  $F_{ST}$  distances in the same way that they were applied to the Euclidean distances as described in the PCA section. An identical MDS and clustering were applied to Slatkin's  $R_{ST}$  distances [67] obtained from Y-chromosome samples. These results were displayed in Figures 3.

**Heatmap.** Relative comparisons between mtDNA  $F_{ST}$  and Y-chromosome  $R_{ST}$  distances were constructed using a heatmap based on the normalized ratio of the Y-chromosome  $R_{ST}$  distance with respect to the total distance ( $R_{ST}/(R_{ST}+F_{ST})$ ) for mtDNA



**Figure 1. Geographic distribution of mtDNA haplogroups.** Frequencies distribution from the current study and from the published data [30,31,35–48] as reported in Table 1. doi:10.1371/journal.pone.0054616.g001

HVS-1  $F_{ST}$  and Y-chromosome  $R_{ST}$  distances (Figure 4). The dendrograms are obtained using complete linkage hierarchical clustering with Euclidean distances between Y/(mtDNA HVS-I+Y) scores.

**AMOVA.** The agglomerative clusters reflecting the results from the MDS analysis were used to identify groups of populations representing affinity of Middle Eastern populations with European and African populations observed in Y-chromosome and mtDNA genetics. AMOVA [68] was applied to the mtDNA and Y sets for each of the mtDNA and Y affinity sets, yielding a 2 by 2 measure of the differences between mtDNA and Y affinities, reported in Table 2.

## Results

### Phylogeographic distribution of mtDNA haplogroups

A total of 185 distinct HVS-I SNPs were identified across all populations (Table S5). The distribution of mtDNA haplogroups shows systematic variation with geography.

The haplogroups' geographical distribution shows affinity between the Northern Levant (modern day Lebanon and Syria) and Europe with clear distinctions between the Levant and the Arabian Peninsula with regards to Africa (Fig. 1, Table 1). The main mtDNA haplogroups for both Europe and the Northern Levant are H and R\*. The subhaplogroup H is more frequent in

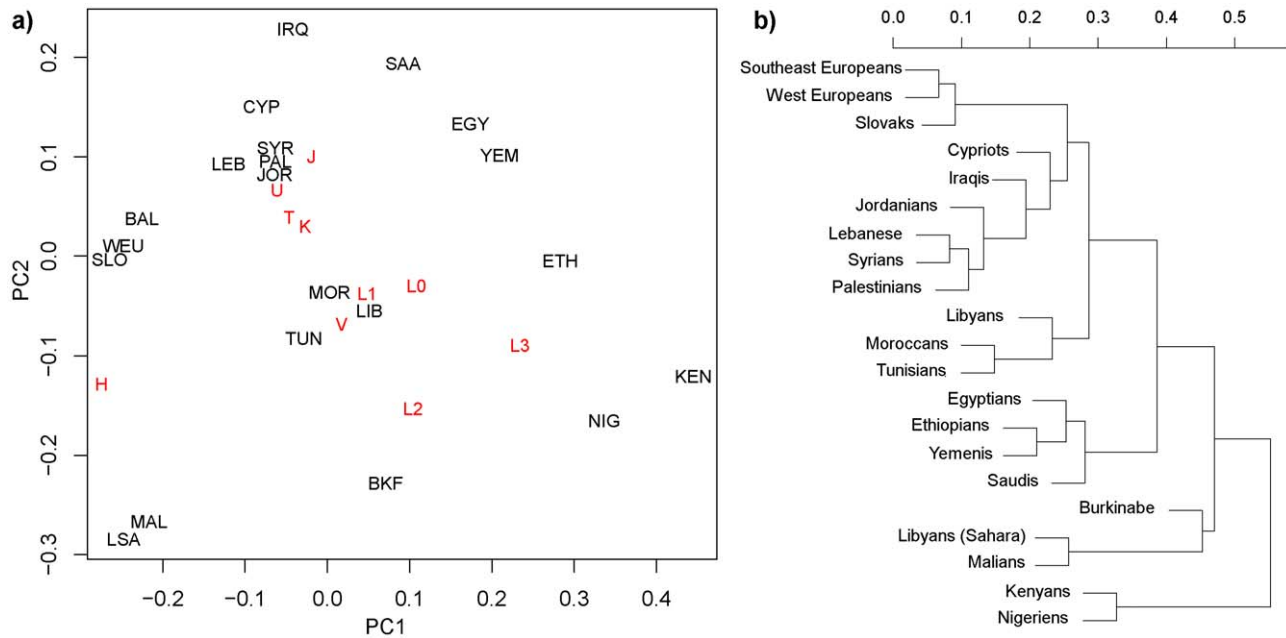
Europe (45%) than in the Levant (25%). Among the Levantine populations, only the Lebanese share Western Europe's overrepresentation of H.

Fisher exact tests were applied to determine when haplogroup frequency differences among populations over both Pan-Mediterranean tests (Table S6), and regional Middle Eastern tests (Table S7), were significant. They reveal patterns of significant over- and under- representation of haplogroups marking regional affinities.

In Lebanese, haplogroups H, HV, T, and K are overrepresented, while Syrians are overrepresented in haplogroups T and K. Western Europeans show overrepresentation of haplogroups H, K.

By contrast, haplogroups J, R0, and M are significantly overrepresented in Saudis, and underrepresented in Western Europeans. Haplogroup J was also significantly overrepresented in Iraqis, among Palestinians, and Yemenis.

The African haplogroups L\* and L3\* are very rare (frequencies less than 1%) and underrepresented in Europe, noting that rarity reduces the power of these tests. In the Levant, Lebanese have the lowest frequency for these haplogroups with generally highly significant underrepresentation, The L haplogroups show rather broad penetration into Yemen, with most being significantly overrepresented, with Yemenis being the only population with an overrepresentation of L6. We have not found haplogroup L6 in our Lebanese (N = 980), Syrian (N = 234), and Jordanian samples



**Figure 2. Populations comparison based on mtDNA haplogroups.** a) Principal Component Analysis of relative frequencies of haplogroups within populations, b) with mean-linkage (UPGMA) dendrogram determined from Euclidean distances. doi:10.1371/journal.pone.0054616.g002

( $N=290$ ). Further, they are absent from Abu-Amero et al's samples from Saudi Arabia, as well as other published results included in his study. A lower bound on the relative frequency of L6 in non-Yemeni Middle East of  $f \geq 8.29 \cdot 10^{-4}$  would guarantee that at least one or more would be observed at least 95% of the time out of the 3614 samples collected across the data on which we are reporting. Assuming independent sampling following a binomial test, there is 5% or less chance of seeing zero L6 by chance with a relative frequency of  $f$  or higher, establishing  $8.29 \cdot 10^{-4}$  as an upper bound to the relative frequency of L6 with a 95% confidence.

We note that subtypes HV0, HV1 and HV2 are generally too weakly represented among our populations to yield tests with adequate power. HV0 and HV1 show sufficient power when pooled into regions. We note that HV0 appears primarily in Europe (34 out of 42 HV0 samples are European, with  $p < 0.0001$ ), HV1 is primarily non-European (4 out of 60 were European, with 20 among the African sample, and 36 among Middle Eastern samples,  $p < 0.0001$ ). HV2 were very rare, with no significant  $p$ -values.

Some U subhaplogroups show regional localization, but none of them rose to sufficient frequency to make any significant contribution to the PCA. U3 appears most frequently in Jordan (Fisher's test:  $p = 2.39e-8$ ), with representation throughout the Middle East. U4 ( $p = 1.49e-7$ ) and U5 ( $p = 2.2e-16$ ) appear to be more heavily European.

The two leading principal components displayed in Figure 2 capture 47.9% and 26.9% of the variance showing a well-defined separation between Mediterranean African populations and sub-Saharan populations (Fig 2a). There is a clear cluster of North African populations comprised of Libyans, Moroccans, and Tunisians. The Nile River marks another boundary of mtDNA differentiation within Africa, linking Egypt, Ethiopia and Kenya but also extending through to Yemen. Yemenis and Saudis both associate strongly with Egyptians, whereas the Jordanian, Lebanese, Palestinian, and Syrian populations clustered together.

Thus, the Arabian Peninsula population clusters were relatively differentiated from the more northern Levantine populations.

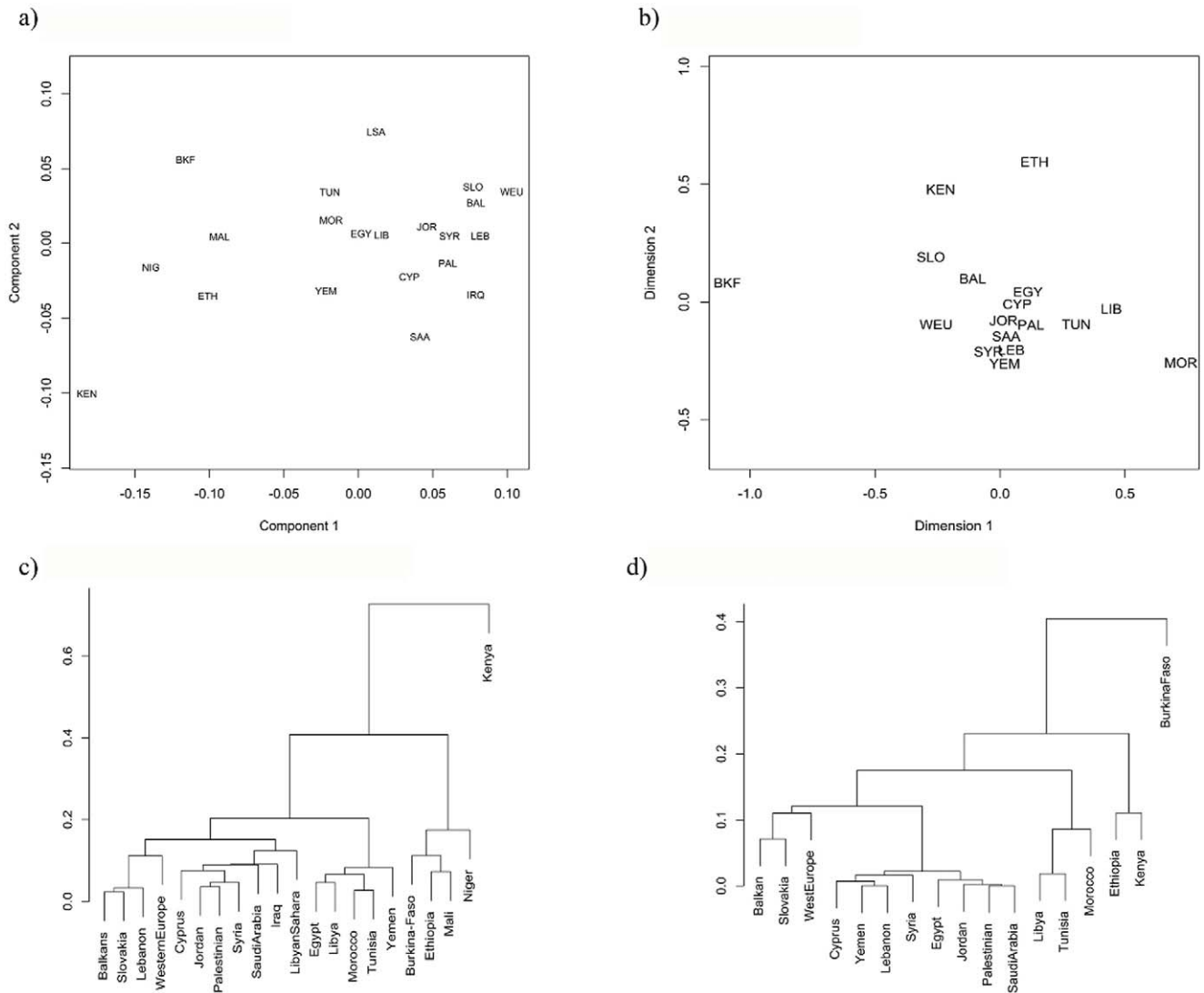
Mitochondrial DNA Haplogroups showing significant contributions to the principal components include H, L3, L2, L0, V, L1, M, J, U, T, K, HV, and R0. The principal vectors for HV, T, K, J, and U point almost directly at the Levantine cluster (Fig 2a). H marks Western Europe and is a significant contributor to Libyan Sahara and Mali mtDNA diversity. L2 and L3 frequencies distinguish the populations of Kenya, Niger, Burkina Faso, Mali, Tunisia, and Libyan Sahara, with a decrease in frequencies of L haplotypes from Kenya through Saudi Arabia.

The dendrogram based on mtDNA haplogroup frequencies (Fig 2b) reveals the strongest differentiation across the Sahara, showing the northern populations differentiated from the southern ones (with Nigeria, Kenya, Mali, Libyan South Sahara, and Burkina-Faso). Egyptian, Yemeni, Saudi Arabian, and Ethiopian populations form a cluster that is distinct from the rest of North Africa, the remaining parts of the Middle East, and Europe. Among these, Libyans, Moroccans, and Tunisians, form a cluster.

UPGMA and PCA showed Yemenis and Saudis (two of the STR predicted Hg populations) closely associated, forming a clear outlier to clusters identifying more northerly Middle Eastern populations and Europe. Slovaks (the third predicted population) also formed a distinct outlier to all of these. Africans were partitioned into northern African populations and Sahel populations, and distinct from the other populations. Burkinabe formed a very distinct outlier to every other population.

MDS analyses were performed using mitochondrial HVS-I based  $F_{ST}$  (Table S8) and Y-chromosome STR  $R_{ST}$  (Table S9) data (Figure 3a & 3b). The  $F_{ST}$ 's computed with mtDNA HVS-I data and  $R_{ST}$ 's computed from Y-STRs of Ethiopian and Levantine populations tended to be less than 1/3, with  $Nm > 1$  [69], roughly corresponding to gene flow between populations over the course of time [5,20,38,70,71].

For the mtDNA HVS-I  $F_{ST}$  MDS analysis, the European populations formed a clear cluster very close to the Cypriots,



**Figure 3. Nonmetric Multidimensional Scaling.** a) mtDNA  $F_{ST}$  and b) Y-STR  $R_{ST}$  distances with c) mtDNA  $F_{ST}$  and d) Y-STR  $R_{ST}$  mean-linkage dendrogram.  
doi:10.1371/journal.pone.0054616.g003

Jordanians, Lebanese, Palestinians, and Syrians. Egyptian, Libyan, Moroccan, and Tunisian populations form a clear cluster. Significantly, Yemenis are on the far side of North Africans, distinct from the Levantine populations and the Libyan Sahara population stands significantly separated from the North African group. The sub-Saharan populations are clearly distinguished from the Mediterranean populations and show significant distances between them in comparison to the Mediterranean populations. The mtDNA HVS-I MDS and dendrogram show most of the Levantine and Arabian Peninsula populations clustering together. Significantly, Yemenis do not seem to cluster with proximal African populations or with Saudis. The entire Levant population seems to cluster with Western Europeans, Southeastern Europeans, and Slovaks.

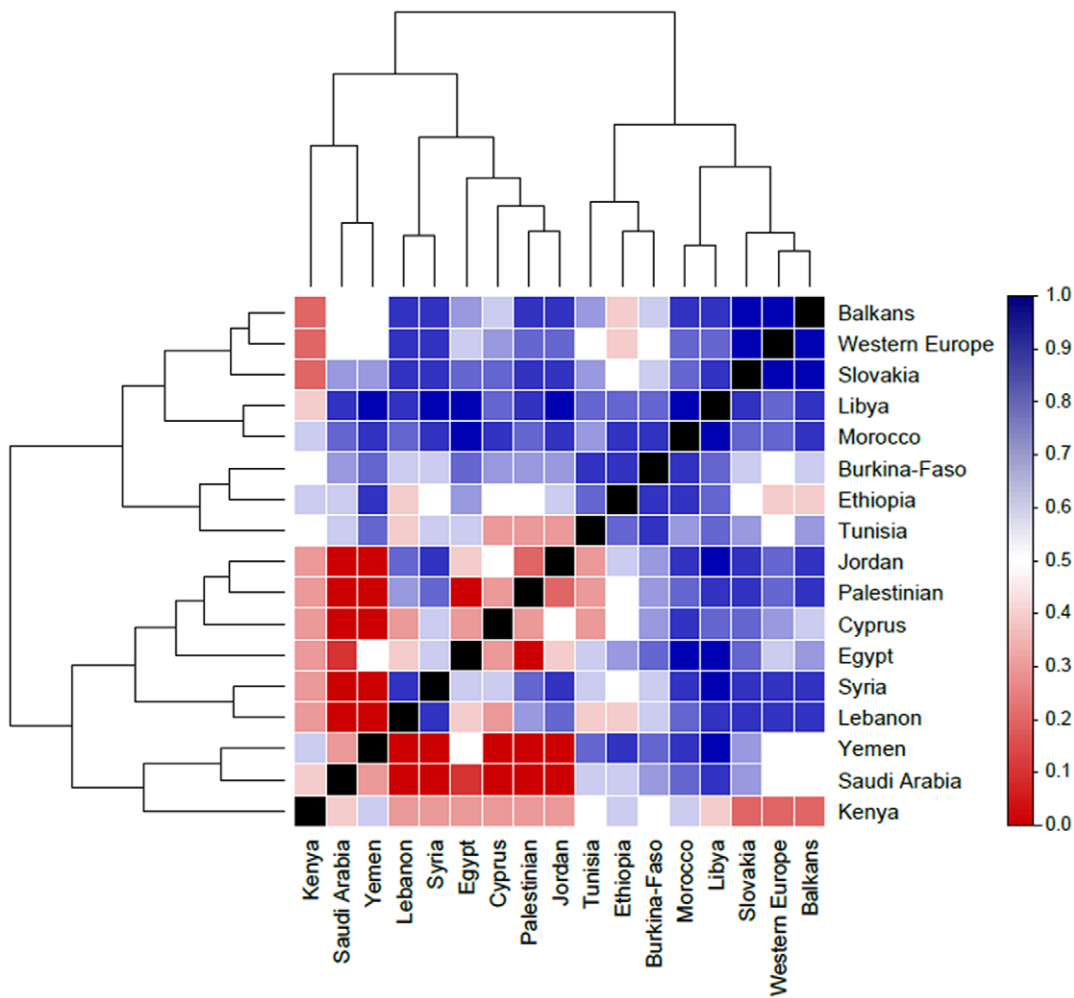
In contrast to mtDNA, the Y-STR-based MDS shows a tight cluster of Cypriots, Egyptians, Jordanians, Lebanese, Palestinians, Saudis, Syrians, and Yemenis, though Libyans, Tunisians, and Moroccans extend away from this cluster. The Southeastern Europeans, Slovaks, and Western Europeans lie in the opposite direction. The dendrogram shows a European cluster closer to the

Levant/Arabian Peninsula cluster and the North African cluster acting as out-group to those.

In general, the MDS plots for mtDNA and Y-STRs show general agreement of European populations extending from the Levant in one direction and North Africans tending to extend in another direction. This places the Levant as a middle ground, either by averaging of in-migration, as a source feeding both North African populations and European populations, or both. The Y and mtDNA MDS plots differ in identifying affinities of Lebanese with Europeans and Yemenis with Egyptians.

#### Comparative analyses of paternal and maternal lineages in the Levant

The relative distance heatmap plot (Figure 4) shows proportion of genetic distances of mtDNA vs. Y. Red colors indicate greater distance of mtDNA vs. Y, while blue colors indicate greater distance of Y vs. mtDNA. Hierarchical clustering organizes the plot relating populations showing similar profiles of Y vs. mtDNA isolation. Most striking is that Saudis, Kenyans, and Yemenis cluster together away from Lebanese, Syrians, Palestinians,



**Figure 4. Heatmap of the Y(Y+mtDNA) distances.** The heatmap shows the normalized ratio of the Y  $F_{ST}$  distance with respect to the total distance ( $Y_{R_{ST}+mtDNA_{F_{ST}}}$  distances). The dendrograms are obtained using complete linkage hierarchical clustering with the Euclidean distance measure. doi:10.1371/journal.pone.0054616.g004

Cypriots and Jordanians in terms of showing relatively high differentiation of mtDNA vs. Y-chromosome genetics. Dendrograms provide a consistent description of the organization of data that may be easily compared with PCA or MDS plots. The application of mean-linkage dendrograms to Y STR data, mtDNA HVS-I data, and mtDNA haplogroup frequency data provides a

consistent basis of comparison. Application of AMOVA to clustering results provides an independent test characterized by p-values and percent variances between vs. within groups. We are not inferring relationships of heritage among populations by application of mean-linkage clustering.

**Table 2. mtDNA vs. Y-chromosome AMOVA results contrasting mtDNA and Y dendrogram-based classifications.**

		mtDNA dendrogram based affinities				Y-Chromosome dendrogram based affinities			
		Variation	% Variation	F	p-value	Variation	% Variation	F	p-value
mtDNA AMOVA	Within Populations	2.43	97.15	0.029	$<10^{-4}$	2.43	97.59	0.024	$<10^{-4}$
	Within Groups	0.028	1.11	0.014	$<10^{-4}$	0.045	1.84	0.018	$<10^{-4}$
	Between Groups	0.043	1.74	0.017	$3.6 \times 10^{-3}$	0.014	0.57	0.006	$1.1 \times 10^{-3}$
Y-Chromosome AMOVA	Within Populations	6.56	87.05	0.129	$<10^{-4}$	6.03	89.04	0.11	$<10^{-4}$
	Within Groups	0.456	6.05	0.065	$<10^{-4}$	0.329	4.86	0.052	$<10^{-4}$
	Between Groups	0.519	6.89	0.068	$6.3 \times 10^{-3}$	0.414	6.11	0.061	$<10^{-4}$

doi:10.1371/journal.pone.0054616.t002



In order to preserve normalization, common subsets comprised of the 11 populations in common in both dendrograms were included. Each of the candidate partitions marking mtDNA affinities and Y affinities formed three groups. The groups representing mtDNA affinities were: (1) Southeastern Europeans, Lebanese, Slovaks, and Western Europeans vs. (2) Cypriots, Jordanians, Palestinians, Saudis, and Syrians, vs. (3) Egyptians and Yemenis. The groups representing Y affinities were: (1) Southeast Europeans, Slovaks, and Western Europeans vs. (2) Cypriots, Lebanese, Syrians, and Yemenis, vs. (3) Egyptians, Jordanians, Palestinians, and Saudis. These two affinity groupings were applied to both the Y and the mtDNA data, yielding results presented in Table 2. Both Y and mtDNA tend to cluster African, European, and Middle Eastern populations separately, and all combinations showed highly significant between-group vs. within-group variations. This reflects the dominating clustering distinguishing Africa, Europe, and the Middle East populations that mean-linkage clustering is picking up. Affinities of Lebanese and the Levantine populations with Europeans vs. Africans depend on comparisons of AMOVA variations within and between groups. Notably, the mtDNA affinity grouping increased AMOVA between-group variation of mtDNA HVS-I data by a factor of 3.05 compared to the result obtained applying the Y affinity grouping to the mtDNA HVS-I data, and decreased AMOVA within-group variation by a factor of 1.66. However, application of the Y affinity grouping reduced AMOVA between-group variation in Y STR data by a factor of 1.13 while reduced AMOVA within-group variations in the Y STR data by a factor of nearly 1.2 compared to the mtDNA affinity grouping. These factors are relatively neutral in contrasting Lebanese Y-chromosome affinity with Europe vs. North Africa, and actually place Lebanese Y-chromosome organization closer to Europeans than Africans. It is expected that mean-linkage clustering would minimize AMOVA within-groups variation, leading to larger AMOVA between-groups variation. Observation did not meet expectation. Instead, the AMOVA within groups' variations for the Y-chromosomes were reduced using the mtDNA clustering compared to Y clustering, suggesting reduced discrimination using the Y clustering for Y-STR genetics.

### Limitations

Y Chromosome haplogroup frequency analyses are limited by a relatively high misclassification rate, with more than half of the populations showing more than 10% misclassification, and Ethiopia showing nearly 50%. Since PCA is a non-linear computation which folds in all populations, the apparent locations of any two populations may shift relative to each other when a third population is added or distorted.

### Discussion

Here we present mitochondrial characteristics of a large group of newly typed samples from five populations (Lebanese, Libyans, Jordanians, Palestinians, and Syrians) and compare their geographical affinity, distribution, and frequency with those of Y-chromosome markers from populations across the broader region of Africa, Europe and the Arabian Peninsula.

The Y-chromosome results of the current study are in agreement with previous studies, suggesting a Middle Eastern gene pool with greater affinity to Africa. Maternal lineages of the Levantine populations studied here, however, reveal stronger European genetic affinities, while not showing Arabian peninsular influences.

### The contrast between the two lineages

Our results show a contrast of mtDNA affinities with previous Y-DNA results [25]. While our Y-DNA MDS and mean-linkage clustering showed a much greater proportion of East African and Near East Y-chromosomes in the Levant, evidence of much less mtDNA affinity, however, was found between the Levant and its southern neighbours.

European mtDNA affinity with the Levant was established in haplogroup frequency data through Fisher exact tests, PCA, and mean-linkage clustering based on Euclidean distances, and in HVS-I derived  $F_{ST}$  distances via MDS and mean-linkage cluster analysis. The mtDNA results are distinct from the Y-STR  $R_{ST}$ -based mean-linkage cluster analysis that showed closer affinity of the Levant populations with Cypriots, North Africans, and Yemenis, than to Europeans.

This cluster analysis suggests that the position of Lebanese relative to European Y-chromosome genetics represented in STR haplotype data is also much more ambiguous than suggested entirely by frequency analysis, revealing otherwise cryptic relationships between Lebanese's Y-STR structure and that of Europeans. Cluster analysis of Y-chromosome frequency based data shows similar partitioning of Europe, Africa, and Middle East, with the Levant much more strongly associated with the Middle East than Europe. As with mtDNA, African Y-chromosome haplogroup data also shows a clear partition between Northern populations and Sahel populations. Due to uncertainties in haplogroup inference from STRs, affinities of Yemenis with Ethiopians vs. Egyptians are uncertain, as are the relationships of Saudi Arabian haplogroups both similar to Yemenis or differentiated from Yemenis in affinity with African populations.

### The Levant and Europe

Beyond the associations noted above, Lebanese show affinity with Europeans for mtDNA haplogroups H, HV, T, K, J, and U, all of which have been identified as markers of agricultural expansions from the Fertile Crescent into Europe [17].

Colonization of West Eurasia by modern humans is believed to have been a consequence of the Out-of-Africa dispersal and to have occurred via the Levant [6]. Indeed, migrating modern humans are believed to have settled near the Arabian Sea until climate changes allowed them to reach the Levant and then Europe [17,72–74]. The LGM, followed by re-expansions from smaller LGM communities relatively isolated by widespread arid conditions, further impacted the coastal-inland contrast of Y-chromosome genetics [25,75]. The significant overrepresentation of mtDNA haplogroup HV among Levantine populations compared to their southern neighbours has suggested these lineages were most likely derived from a single maternal Levantine source population [17].

### Arabian genetic expansions: Arabia East Africa, and North Africa

From the 7<sup>th</sup> millennium B.C.E., empire expansions and trade, including the slave trade, heavily influenced genetic migration between Yemen and East Africa. Alternatively, known trade networks linking Egypt with Yemen included those for obsidian, and later through Aksum, spices, incense and other precious materials, as well as slaves [76–78]. It is particularly clear from prior mtDNA studies of this region that East African migration into Arab populations involved females to an extensive degree [79]. While Ethiopian and other East African populations may appear to be better candidates for the origins of modern Yemeni populations, our PCA and MDS analyses, and their associated

mean-linkage clustering of Yemen's mtDNA, show greater affinity between Yemenis, Egyptians and North Africans. They share in common haplogroups J, L0, L2, and N1. Comparison of mtDNA HVS-I  $F_{ST}$  distances also suggest that Yemen appears more similar to Egypt than Ethiopia.

Two haplogroups in this region show significant evidence of relative isolation. First, mtDNA patterns for haplogroup J reflect relatively moderate genetic outflow from Saudi Arabia, and haplogroup L6 is strongly localized within Yemen. Haplogroup J is evenly distributed throughout the Middle East, except in Saudi Arabia where it is significantly overrepresented.

It is likely the pattern of Hg J's significant penetration, and the shared underrepresentation of Hg H, tips the balance for Yemenis' mtDNA affinity with Egyptians. Given the significant underrepresentation of Hg J in East Africa, while not being significantly uncommon in Egypt, it is therefore plausible that Arabian female gene flow followed well established trade routes on the Red Sea with Egypt and North Africa while avoiding assimilation of Yemeni L6's on the way.

The most striking feature of the heat map (Fig 4) is the relative isolation of mtDNA genetics of Yemenis and Saudis from the other populations in the Middle East in comparison to Y-chromosome variation. While Yemenis appear to share overrepresented haplogroups that characterize each of its neighbouring populations, none of the African populations have become dominated by Saudi Arabian J's, nor have Middle Eastern populations been differentially dominated by the in-migration of African L's the way Yemenis have.

The expansion of trade through the Red Sea and into the Indian Ocean basin starting in Classical times has provided the largest opportunities for genetic transfers from Africa into Yemen, being dominated by the Red Sea superpower: Egypt. The distribution of mtDNA haplogroup L6 provides a measure of the limited impact of genetic outflow from Yemen, and this flow seems to have been primarily unidirectional. This establishes the upper limits for Yemeni female-mediated gene flow during the Muslim Expansions, as well as identifying possible routes for the expansions.

Whether considering haplogroup composition revealed in Fisher tests, PCA, or  $F_{ST}$  based MDS analysis of HVS-I data, mtDNA shows a much stronger affinity between Levantine populations and Europeans compared with the rest of the Middle Eastern populations, or with North Africans. While Lebanese and Yemeni mtDNA epitomize very distinct affinities to different populations and regions well outside of the Middle East, Saudi Arabia seems to display strong local over-representation haplogroup J, while Yemen is even more localized in its L6. Further, these large-scale differences in affinity between mtDNA genetics appear in sharp contrast to regional affinities seen in their Y-chromosomal counterparts. While the mtDNA signal is sharp and clear in its affinities, the Y-chromosome results show somewhat more ambiguous associations in  $R_{ST}$ -based analyses, with Lebanese showing less within-group variation when organized consistently with mtDNA and demonstrating associations closer to Europeans than Africans. This would suggest that while male migrants accompanied female migrants, especially to Europe, females did not always accompany male migrants, especially into North Africa. This leaves a more ambiguous signal for male compared to female migrations.

The historical and archaeological record reveals how trade and labour, colonization and settlement events, and military expansions all contributed to the immigration and displacement of individuals throughout these regions. As a distinct crossroad between geographic regions and civilizations, the Levant and the

Near East harbour unique genetic affinities which are revealed most clearly through the comparison of Y-chromosome and mtDNA data.

Due to uncertainties in haplogroup inference from STRs, specific questions regarding affinities of Yemen with Ethiopia vs. Egypt are inaccessible, as are questions regarding the relationship of Saudi Arabian haplogroups both similar to Yemenis or differentiated from Yemenis in affinity with African populations.

## Supporting Information

**Figure S1 Geographic distribution of Y haplogroups.** Frequencies from published data as reported in Table S3. (TIF)

**Figure S2 Populations comparison based on Y haplogroups** a) Principal Component Analysis of relative frequencies of Y haplogroups within populations, b) with mean-linkage (UPGMA) dendrogram determined from Euclidean distances. (TIF)

**Table S1 mtDNA haplotypes analyzed in this study.** (XLS)

**Table S2 Y chromosome STR haplotypes and haplogroups employed in this study.** (XLS)

**Table S3 Y chromosome Haplogroup frequencies of populations used in this study.** (XLS)

**Table S4 List of the mtDNA Haplogroup marker SNPs typed for this study.** (XLS)

**Table S5 List of HVS-I mtDNA SNPs identified across populations used in this study.** (XLS)

**Table S6 Fisher exact tests for haplogroup frequencies vs. population across all study populations.** (XLS)

**Table S7 Fisher exact tests for haplogroup frequencies vs. population within the Middle East.** (XLS)

**Table S8 mtDNA  $F_{ST}$  distances between populations.** (XLS)

**Table S9 Y STR  $R_{ST}$  distances between populations.** (XLS)

## Acknowledgments

We thank the sample donors for taking part in this study. We would like to thank Professor Colin Renfrew for his insights, critical comments, and suggestions which helped us improve this manuscript significantly.

The Genographic Consortium includes: Janet S. Ziegler (Applied Biosystems, Foster City, California, United States); Li Jin & Shilin Li (Fudan University, Shanghai, China); Pandikumar Swamikrishnan (IBM, Somers, New York, United States); Asif Javed, Laxmi Parida & Ajay K. Royyuru (IBM, Yorktown Heights, New York, United States); Lluís Quintana-Murci (Institut Pasteur, Paris, France); R. John Mitchell (La Trobe University, Melbourne, Victoria, Australia); Syama Adhikarla, ArunKumar GaneshPrasad, Ramasamy Pitchappan & Arun Varatharajan Santhakumari (Madurai Kamaraj University, Madurai, Tamil Nadu, India); Angela Hobbs & Himla Soodyall (National Health Laboratory Service, Johannesburg, South Africa); Elena Balanovska & Oleg Balanovsky (Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia); Daniela R. Lacerda & Fabrício R.

Santos (Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil); Pedro Paulo Vieira (Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil); Jaume Bertranpetit, David Comas, Begonia Martínez-Cruz & Marta Melé (Universitat Pompeu Fabra, Barcelona, Spain); Christina J. Adler, Alan Cooper, Clio S. I. Der Sarkissian & Wolfgang Haak (University of Adelaide, South Australia, Australia); Matthew E. Kaplan & Nirav C. Merchant (University of Arizona, Tucson, Arizona, United States); Colin Renfrew (University of Cambridge, Cambridge, United Kingdom); Andrew C. Clarke & Elizabeth A. Matisoo-Smith (University of Otago, Dunedin, New Zealand); Matthew

C. Dulik, Jill B. Gaijeski, Amanda C. Owings, Theodore G. Schurr & Miguel G. Vilar (University of Pennsylvania, Philadelphia, Pennsylvania, United States).

## Author Contributions

Conceived and designed the experiments: PZ CTS RSW EMS. Performed the experiments: DB BD SY AS. Analyzed the data: DP MH DSH GK. Contributed reagents/materials/analysis tools: BJ MGS. Wrote the paper: PZ DP DB.

## References

- Stringer CB, Grun R, Schwarcz HP, Goldberg P (1989) ESR dates for the hominid burial site of Es Skhul in Israel. *Nature* 338: 756–758.
- Bar-Yosef O (1992) The role of western Asia in modern human origins. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 337: 193–200.
- Lahr MM, Foley RA (1998) Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *American journal of physical anthropology Suppl* 27: 137–176.
- Tchernov E (1994) New comments on the biostratigraphy of the Middle and Upper Pleistocene of the southern Levant. In: Ben-Yosef O, Kra RS, editors. *Late Quaternary Chronology and Paleoclimates of the Eastern Mediterranean: Radiocarbon*. pp. 333–350.
- Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, et al. (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *American journal of human genetics* 74: 532–544.
- Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, et al. (2006) The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* 314: 1767–1770.
- Bar-Yosef O (1998) The Natufian Culture in the Levant, Threshold to the Origins of Agriculture. *Evolutionary Anthropology* 6: 159–177.
- Valladas H, Reyss JL, Joron JL, Valladas G, Bar-Yosef O, et al. (1988) Thermoluminescence data of Mousterian Troto-Cro-Magnon remains from Israel and the origin of modern man. *Nature* 331: 614–616.
- Mercier N, Valladas H, Bar-Yosef O, Stringer CB, Joron JL (1993) Thermoluminescence dates for the Mousterian Burial Site of Es-Skhul, Mt. Carmel. *Journal of Archaeological Science* 20: 169–174.
- Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325: 31–36.
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253: 1503–1507.
- Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC genetics* 2: 13.
- Nei M, Roychoudhury AK (1993) Evolutionary Relationships of Human Populations on a Global Scale. *Molecular biology and evolution* 10: 927–943.
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.
- Foley RA, Lahr MM (1997) Mode 3 technologies and the evolution of modern humans. *Cambridge Archaeological Journal* 7: 3–36.
- Cavalli-Sforza LL (1997) Genes, peoples, and languages. *Proceedings of the National Academy of Sciences of the United States of America* 94: 7719–7724.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *American journal of human genetics* 67: 1251–1276.
- Richards M, Macaulay V, Torroni A, Bandelt HJ (2002) In search of geographical patterns in European mitochondrial DNA. *American journal of human genetics* 71: 1168–1174.
- Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000) Geographic patterns of mtDNA diversity in Europe. *American journal of human genetics* 66: 262–278.
- Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, et al. (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *American journal of human genetics* 74: 1023–1034.
- Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, et al. (2003) Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Annals of human genetics* 67: 312–328.
- Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, et al. (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Human immunology* 62: 871–884.
- Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, et al. (2004) Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Human genetics* 115: 357–371.
- Flores C, Maca-Meyer N, Gonzalez AM, Oefner PJ, Shen P, et al. (2004) Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *European journal of human genetics* : EHG 12: 855–863.
- El-Sibai M, Platt DE, Haber M, Xue Y, Youhanna SC, et al. (2009) Geographical structure of the Y-chromosomal genetic landscape of the Levant: a coastal-inland contrast. *Annals of human genetics* 73: 568–581.
- Zalloua PA, Platt DE, El Sibai M, Khalife J, Makhoul N, et al. (2008) Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *American journal of human genetics* 83: 633–642.
- Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ (2008) Y-chromosome diversity characterizes the Gulf of Oman. *European journal of human genetics* : EHG 16: 374–386.
- Rowold DJ, Luis JR, Terreros MC, Herrera RJ (2007) Mitochondrial DNA gene flow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. *Journal of human genetics* 52: 436–447.
- Underhill PA, Kivisild T (2007) Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual review of genetics* 41: 539–564.
- Gonzalez AM, Karadsheh N, Maca-Meyer N, Flores C, Cabrera VM, et al. (2008) Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations. *Annals of human biology* 35: 212–231.
- Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, et al. (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Molecular phylogenetics and evolution* 28: 458–472.
- Pilkington MM, Wilder JA, Mendez FL, Cox MP, Woerner A, et al. (2008) Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa. *Molecular biology and evolution* 25: 517–525.
- Coelho M, Sequeira F, Luiselli D, Belez S, Rocha J (2009) On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC evolutionary biology* 9: 80.
- Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, et al. (2005) Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *European journal of human genetics* : EHG 13: 867–876.
- Pereira L, Cerny V, Cerezo M, Silva NM, Hajek M, et al. (2010) Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. *European journal of human genetics* : EHG 18: 915–923.
- Irwin J, Saunier J, Strouss K, Paintner C, Diegoli T, et al. (2008) Mitochondrial control region sequences from northern Greece and Greek Cypriots. *International journal of legal medicine* 122: 87–89.
- Saunier JL, Irwin JA, Strouss KM, Ragab H, Sturk KA, et al. (2009) Mitochondrial control region sequences from an Egyptian population sample. *Forensic science international Genetics* 3: e97–103.
- Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, et al. (2004) Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *American journal of human genetics* 75: 752–770.
- Dubut V, Chollet L, Murail P, Cartault F, Beraud-Colomb E, et al. (2004) mtDNA polymorphisms in five French groups: importance of regional sampling. *European journal of human genetics* : EHG 12: 293–300.
- Richard C, Pennarun E, Kivisild T, Tambets K, Tolk HV, et al. (2007) An mtDNA perspective of French genetic variation. *Annals of human biology* 34: 68–79.
- Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koeh DK, et al. (2004) Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *International journal of legal medicine* 118: 294–306.
- Otoni C, Martínez-Labarga C, Loogvali EL, Pennarun E, Achilli A, et al. (2009) First genetic insight into Libyan Tuaregs: a maternal perspective. *Annals of human genetics* 73: 438–448.
- Harich N, Costa M, Fernandes V, Kandil M, Pereira J, et al. (2010) The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. *BMC evolutionary biology* 10: 138.
- Turchi C, Buscemi L, Giacchino E, Onofri V, Fendt L, et al. (2009) Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enrichment of forensic mtDNA databases with Northern Africa data. *Forensic science international Genetics* 3: 166–172.
- Abu-Amro KK, Larruga JM, Cabrera VM, Gonzalez AM (2008) Mitochondrial DNA structure in the Arabian Peninsula. *BMC evolutionary biology* 8: 45.

46. Malyarchuk B, Grzybowski T, Derenko M, Perkova M, Vanecek T, et al. (2008) Mitochondrial DNA phylogeny in Eastern and Western Slavs. *Molecular biology and evolution* 25: 1651–1658.
47. Cherni L, Loueslati BY, Pereira L, Enmafaa H, Amorim A, et al. (2005) Female gene pools of Berber and Arab neighboring communities in central Tunisia: microstructure of mtDNA variation in North Africa. *Human biology* 77: 61–70.
48. Cerny V, Mulligan CJ, Ridl J, Zaloudkova M, Edens CM, et al. (2008) Regional differences in the distribution of the sub-Saharan, West Eurasian, and South Asian mtDNA lineages in Yemen. *American journal of physical anthropology* 136: 128–137.
49. Bosch E, Calafell F, Gonzalez-Neira A, Flaiz C, Mateu E, et al. (2006) Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Annals of human genetics* 70: 459–487.
50. de Filippo C, Barbieri C, Whitten M, Mpoloka SW, Gunnarsdottir ED, et al. (2011) Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Molecular biology and evolution* 28: 1255–1269.
51. Ferri G, Ceccardi S, Lugaresi F, Bini C, Ingravallo F, et al. (2008) Male haplotypes and haplogroups differences between urban (Rimini) and rural area (Valmarecchia) in Romagna region (North Italy). *Forensic science international* 175: 250–255.
52. Alshamali F, Pereira L, Budowle B, Poloni ES, Currat M (2009) Local population structure in Arabian Peninsula revealed by Y-STR diversity. *Hum Hered* 68: 45–54.
53. Petrejčikova E, Sotak M, Bernasovska J, Bernasovsky I, Rebala K, et al. (2011) Allele frequencies and population data for 11 Y-chromosome STRs in samples from Eastern Slovakia. *Forensic science international Genetics* 5: e53–62.
54. Athey T (2005) Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach. *Journal of Genetic Genealogy* 1: 1–7.
55. Athey T (2006) Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach. *Journal of Genetic Genealogy* 2: 34–39.
56. Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, et al. (2002) The making of the African mtDNA landscape. *American journal of human genetics* 71: 1082–1111.
57. Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, et al. (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. *American journal of human genetics* 74: 454–465.
58. Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, et al. (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS biology* 3: e247.
59. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human mutation* 30: E386–394.
60. R Development Core Team (2011) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
61. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. New York, NY: Springer-Verlag.
62. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457–465.
63. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics* 23: 147.
64. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* 10: 564–567.
65. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105: 767–779.
66. Cox TF, Cox MAA (2001) *Multidimensional Scaling*, Second Edition. New York, NY: Chapman and Hall.
67. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457–462.
68. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
69. Wright S (1951) The genetical structure of populations. *Ann Eugenics* 15: 323–354.
70. Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA (2002) Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *American journal of human genetics* 70: 265–268.
71. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, et al. (2000) Y chromosome sequence variation and the history of human populations. *Nature genetics* 26: 358–361.
72. Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, et al. (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308: 1034–1036.
73. Mellars P (2006) Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proceedings of the National Academy of Sciences of the United States of America* 103: 9381–9386.
74. Van Andel TH, Tzedakis PC (1996) Paleolithic landscapes of Europe and environs: 150,000–25,000 years ago: an overview. *Quaternary Science Reviews* 15: 481–500.
75. Chiaroni J, King R, Underhill PA (2008) Correlation of annual precipitation with human Y-chromosome diversity and the emergence of Neolithic agricultural and pastoral economies in the Fertile Crescent. *Antiquity* 82: 281–289.
76. Fattovich R (1997) The Near East and Eastern Africa: Their Interaction. In: Vogel JO, editor. *Encyclopedia of precolonial Africa*. Walnut Creek: AltaMira Press. pp. 479–484.
77. Segal R (2001) *Islam's Black Slaves: The Other Black Diaspora*. New York: Farrar, Straus, and Giroux.
78. Lewis B (1990) *Race and slavery in the Middle East: an historical enquiry*. New York: Oxford University Press.
79. Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, et al. (2003) Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *American journal of human genetics* 72: 1058–1064.