The Matrilineal Ancestry of Ashkenazi Jewry: Portrait of a Recent Founder Event

Doron M. Behar,[1] Ene Metspalu,[2] Toomas Kivisild,[2] Alessandro Achilli,[3] Yarin Hadid,[1] Shay Tzur,[1] Luisa Pereira,[4] Antonio Amorim,[4,5] Lluís Quintana-Murci,[6] Kari Majamaa,[7] Corinna Herrnstadt,[8] Neil Howell,[8] Oleg Balanovsky,[2,9] Ildus Kutuev,[2,10] Andrey Pshenichnov,[2,9] David Gurwitz,[11] Batsheva Bonne-Tamir,[11] Antonio Torroni,[3] Richard Villems,[2] and Karl Skorecki[1]

[1]Rappaport Faculty of Medicine and Research Institute, Technion and Rambam Medical Center, Haifa, Israel; [2]Department of Evolutionary Biology, University of Tartu and Estonian Biocentre, Tartu, Estonia; [3]Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy; [4]Instituto de Patologia e Immunologia Molecular da Universidade do Porto (IPATIMUP) and [5]Faculdade de Ciências, Universidade do Porto, Porto, Portugal; [6]Centre National de la Recherche Scientifique (CNRS) FRE 2849, Unit of Molecular Prevention and Therapy of Human Diseases, Institut Pasteur, Paris; [7]Department of Neurology, University of Turku, Turku, Finland; [8]MitoKor, Inc.,* San Diego, CA; [9]Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow; [10]Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences, Ufa, Russia; and [11]Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Both the extent and location of the maternal ancestral deme from which the Ashkenazi Jewry arose remain obscure. Here, using complete sequences of the maternally inherited mitochondrial DNA (mtDNA), we show that close to one-half of Ashkenazi Jews, estimated at 8,000,000 people, can be traced back to only 4 women carrying distinct mtDNAs that are virtually absent in other populations, with the important exception of low frequencies among non-Ashkenazi Jews. We conclude that four founding mtDNAs, likely of Near Eastern ancestry, underwent major expansion(s) in Europe within the past millennium.

The founder effect was originally defined as "the establishment of a new population by a few original founders (in an extreme case, by a single fertilized female) which carry only a small fraction of the total genetic variation of the parental population" (Mayr 1963). Since then, DNA variation studies in human populations—particularly those employing the mtDNA and the male-specific portion of the Y chromosome (MSY)—have proven invaluable for generating models of the evolution of modern humans (Cavalli-Sforza and Feldman 2003). Recent progress in the analysis of complete mtDNA genomes, combined with the growing number of samples in existing databases, provides a sophisticated tool to dissect microevolutionary processes, such as founding events, in individual haplogroups (Hgs) and populations (Ingman et al. 2000; Achilli et al. 2004, 2005; Loogväli et al. 2004; Palanichamy et al. 2004; Macaulay et al. 2005; Thangaraj et al. 2005).

The term "Ashkenazi" refers to Jews of mainly central and eastern European ancestry, in contrast to those of Iberian (Sephardic), Near Eastern, or North African origin (Ostrer 2001). Most historical records indicate that the founding of the Ashkenazi Jewry took place in the Rhine Basin, followed by a dramatic expansion into eastern Europe. However, both the origin and size of the maternal ancestral deme remain obscure. Two features have made the Ashkenazi Jewish population an excellent candidate for genetic studies. First, its unique, well-documented overall demography is consistent with several founding events, repeated bottlenecks, and dramatic expansions, from an estimated number of ~25,000 in 1300 A.D. to >8,500,000 around the turn of the 20th century (DellaPergola 2001; Ostrer 2001). Second, the unusual accumulation of >20 recessive disorders restricted to this population and the proposed possible explanations for this phenomenon have fueled a long-standing and lively debate between theories favoring heterozygote advantage and theories postulating drift of rare disease-causing

Received September 28, 2005; accepted for publication December 6, 2005; electronically published January 11, 2006.
Address for correspondence and reprints: Dr. Richard Villems, Department of Evolutionary Biology, University of Tartu and Estonian Biocentre, Riia 23, Tartu 51010, Estonia. E-mail: rvillems@ebc.ee
* Now Migenix.
Am. J. Hum. Genet. 2006;78:000–000. © 2006 by The American Society of Human Genetics. All rights reserved. 0002-9297/2006/7803-00XX$15.00

alleles during the rapid population expansion that followed the founding event (Diamond 1994; Risch et al. 1995; Ostrer 2001). The published genetic data addressing the question of a founding event in the maternal history of Ashkenazi Jews (Torroni et al. 1996; Thomas et al. 2002; Behar et al. 2004) are partially discordant, with some studies detecting a strong founder event (Torroni et al. 1996; Behar et al. 2004) and others reaching the conclusion that there is little evidence for such an event (Thomas et al. 2002). In particular, our recent analysis of HVS-I sequences (Behar et al. 2004), extended herein to a larger fraction of the control-region (16024–00300) variation in Ashkenazi Jews, has yielded a broad range of Hgs well known to be prevalent and shared between Europe and the Near East and, therefore, not informative in determining the geographic origin of the population ancestral to contemporary Ashkenazi Jews (tables 1 and 2). Despite the presence of the entire range of west Eurasian Hgs in the Ashkenazi mtDNA pool, Hg frequencies clearly deviated from those reported elsewhere for west Eurasian populations. This deviation was due to a striking overrepresentation of Hgs K and N1b. However, since common Hgs cover large geographic areas and comprise numerous lineages that usually coalesced tens of thousands of years ago, monitoring their general frequencies does not effectively allow determination of the real number of ancestral maternal lineages that gave rise to the present-day diversity in a population. Therefore, although in previous studies the elevated frequencies of mtDNA Hgs K and N1b in Ashkenazi Jews suggested a maternal founding event (Torroni et al. 1996; Behar et al. 2004), the number of lineages within these Hgs, their putative origin, and their level of restriction to Ashkenazi Jews remain to be resolved. To this end, in the current study, using complete mtDNA sequence analysis, we identified the founding lineages of Ashkenazi Hgs K and N1b and contrasted their variation against a global set of mtDNAs belonging to the same Hgs. On the basis of these data, we infer the actual number, as well as the temporal and geographic origin, of these maternal founders.

We initially generated a maximum parsimony tree of 121 complete mtDNA sequences belonging to Hg K. The tree encompassed 28 novel and 93 previously reported mtDNAs (fig. 1 and table 3). The sequencing procedure and phylogeny construction were performed as described in appendix A. Of the 28 novel samples, 13 were from Ashkenazi Jews, and 15 were selected from non-Ashkenazi Jews and non-Jewish Near Eastern populations. Samples for complete mtDNA sequencing were chosen to include the widest possible range of Hg K internal variation, on the basis of sequence analysis of the mtDNA control region. Figure 1 shows that Hg K splits at its root into two primary branches, K1 and K2. All 789 Hg K mtDNAs reported herein (table 4) were designated as either "K1" or "K2" (89% and 11%, respectively), revealing no additional branching at the root of K. Subhaplogroup (subHg) K2 is subdivided into two subsequent major branches labeled "K2a" and "K2b," whereas K1 splits into three branches—K1a, K1b, and K1c—that are defined by positions 497, 5913, and 498del, respectively. All but one of the K1 mtDNAs reported in this study belong to one of these three branches. SubHg K1a encompassed most subbranches and is also the dominant branch in our sample set, encompassing 80% of all Hg K mtDNAs. The 13 Ashkenazi complete mtDNAs clustered into three distinct branches in the phylogeny: K1a1b1a, K1a9, and K2a2a (fig. 1).

Table 1

Control-Region Sequences and Hg Affiliation of Ashkenazi mtDNAs

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

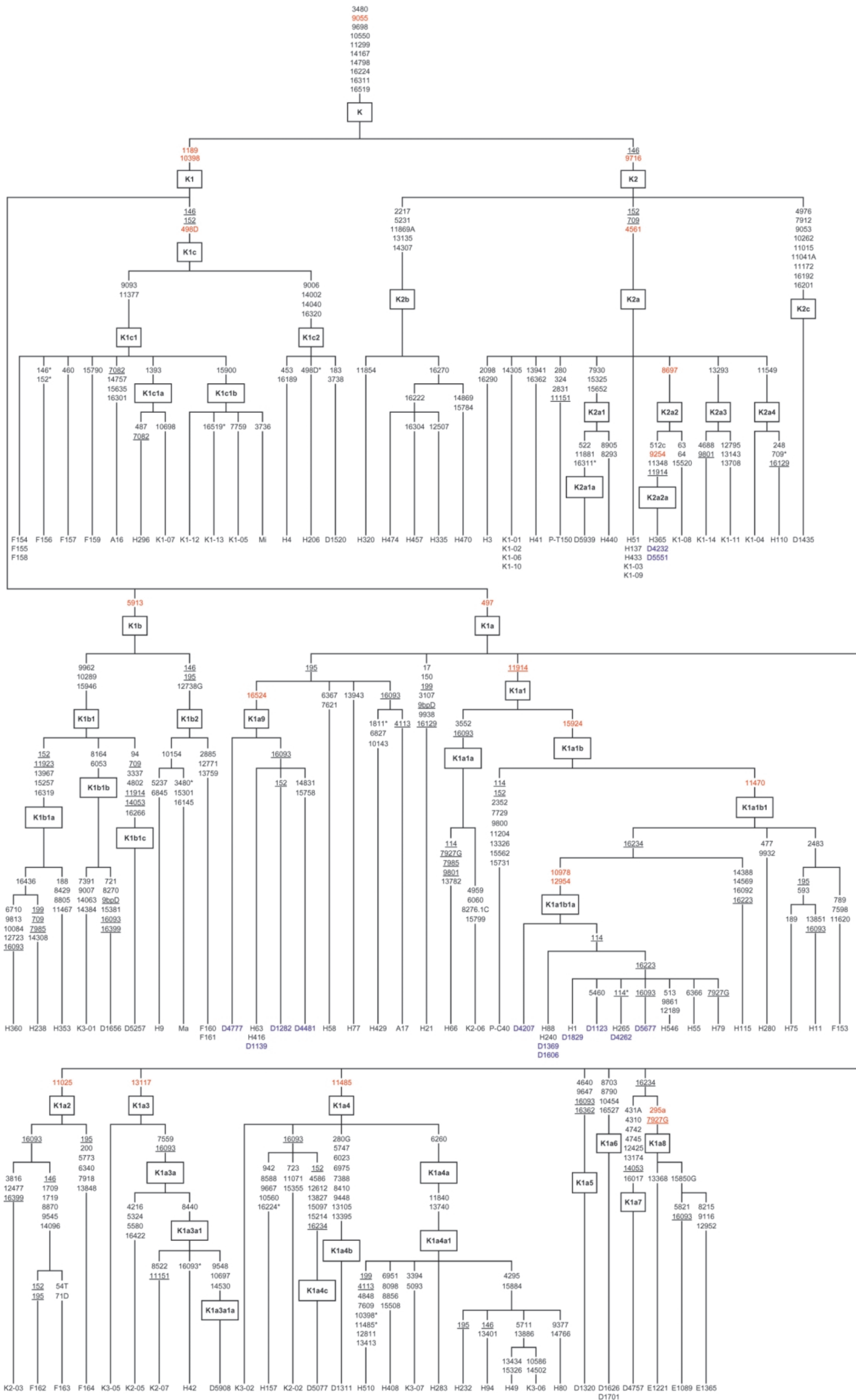
Table 2

Control-Region Haplotypes of Ashkenazi Hg K mtDNAs

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

K1a1b1a is marked by two coding-region transitions, 10978 and 12954, and includes 14 of the 121 complete sequences. Seven of these are reported for the first time herein and are from Ashkenazi subjects, whereas the other seven were reported elsewhere as forming a specific cluster termed "K1a" (Herrnstadt et al. 2002). The ethnicities or religious affiliations of these seven subjects are not available, but they were all collected in the United States and shared the control-region mutations with the Ashkenazi samples. Since the majority of contemporary Ashkenazi Jews reside in the United States, it is possible that this cluster represents a sample set of Ashkenazi Jews, though we have no way to confirm or refute this.

K1a9 mtDNAs are marked only by the control-region transition 16524 but lack the diagnostic mutations of the other K1a2 through K1a8 subHgs (fig. 1). Six samples belong to this lineage—four of which, from Ashkenazi Jews, are reported herein for the first time. The other two mtDNAs were reported elsewhere, and the same considerations noted for subHg K1a1b1a, with respect to the ethnic or religious affiliation, also apply (Herrnstadt et al. 2002). K2a2a mtDNAs are marked by three coding-region transitions: 9254, 11348, and 11914. Three mtDNAs belong to this lineage, two of which are from Ashkenazi individuals and are reported



3480
9055
9698
10550
11299
14167
14798
16224
16311
16519
K
1189
10398
K1
146
9716
K2
146
152
498D
K1c
2217
5231
11869A
13135
14307
K2b
152
709
4561
K2a
4976
7912
9053
10262
11015
11041A
11172
16192
16201
K2c
9093
11377
K1c1
9006
14002
14040
16320
K1c2
146*
152*
460
15790
7082
14757
15635
16301
1393
K1c1a
487
7082
10698
15900
K1c1b
16519*
7759
3736
453
16189
498D*
183
3738
11854
16270
16222
16304
12507
14869
15784
2098
16290
14305
13941
16362
280
324
2831
11151
7930
15325
15652
K2a1
522
11881
16311*
K2a1a
8905
8293
8697
K2a2
512c
9254
11348
11914
K2a2a
63
64
15520
13293
K2a3
4688
9801
12795
13143
13708
11549
K2a4
248
709*
16129
F154
F155
F158
F156
F157
F159
A16
H296
K1-07
K1-12
K1-13
K1-05
Mi
H4
H206
D1520
H320
H474
H457
H335
H470
H3
K1-01
K1-02
K1-06
K1-10
H41
P-T150
D5939
H440
H51
H137
H433
K1-03
K1-09
H365
D4232
D5551
K1-08
K1-14
K1-11
K1-04
H110
D1435
5913
K1b
497
K1a
9962
10289
15946
K1b1
146
195
12738G
K1b2
195
17
150
199
3107
9bpD
9938
16129
11914
K1a1
152
11923
13967
15257
16319
K1b1a
8164
6053
K1b1b
94
709
3337
4802
11914
14053
16266
K1b1c
10154
5237
6845
3480*
15301
16145
2885
12771
13759
16524
K1a9
16093
152
14831
15758
6367
7621
13943
16093
1811*
6827
10143
4113
3552
16093
K1a1a
15924
K1a1b
114
152
2352
7729
9800
11204
13326
15562
15731
11470
K1a1b1
114
7927G
7985
9801
13782
4959
6060
8276.1C
15799
16234
10978
12954
K1a1b1a
114
16223
5460
114*
16093
513
9861
12189
6366
7927G
14388
14569
16092
16223
477
9932
2483
195
593
189
13851
16093
789
7598
11620
16436
6710
9813
10084
12723
16093
199
709
7985
14308
188
8429
8805
11467
7391
9007
14063
14384
721
8270
9bpD
15381
16093
16399
H360
H238
H353
K3-01
D1656
D5257
H9
Ma
F160
F161
D4777
H63
H416
D1139
D1282
D4481
H58
H77
H429
A17
H21
H66
K2-06
P-C40
D4207
H88
H240
D1369
D1606
H1
D1829
D1123
H265
D4262
D5677
H546
H55
H79
H115
H280
H75
H11
F153
11025
K1a2
13117
K1a3
11485
K1a4
4640
9647
16093
16362
K1a5
8703
8790
10454
16527
K1a6
16234
431A
4310
4742
4745
12425
13174
14053
16017
K1a7
295a
7927G
K1a8
13368
15850G
5821
16093
8215
9116
12952
16093
3816
12477
16399
146
1709
1719
8870
9545
14096
152
195
54T
71D
195
200
5773
6340
7918
13848
7559
16093
K1a3a
4216
5324
5580
16422
8440
K1a3a1
8522
11151
16093*
9548
10697
14530
K1a3a1a
16093
942
8588
9667
10560
16224*
723
11071
15355
152
4586
12612
13827
15097
15214
16234
K1a4c
280G
5747
6023
6975
7388
8410
9448
13105
13395
K1a4b
6260
K1a4a
11840
13740
K1a4a1
199
4113
4848
7609
10398*
11485*
12811
13413
6951
8098
8856
15508
3394
5093
4295
15884
195
146
13401
5711
13886
13434
15326
10586
14502
9377
14766
K2-03
F162
F163
F164
K3-05
K2-05
K2-07
H42
D5908
K3-02
H157
K2-02
D5077
D1311
H510
H408
K3-07
H283
H232
H94
H49
K3-06
H80
D1320
D1626
D1701
D4757
E1221
E1089
E1365

herein for the first time. The third mtDNA was reported elsewhere in the same U.S. data set (Herrnstadt et al. 2002).

After the delineation of the Hg K topology and in a search for clues regarding the possible origin of the Ashkenazi lineages, 789 Hg K mtDNAs of a global set of 13,359 samples (tables 5 and 6) were hierarchically screened for polymorphisms relevant to include or exclude the samples from the three Ashkenazi K subbranches—K1a1b1a, K1a9, and K2a2a. The position of the three dominant Ashkenazi founding sequences within the phylogenetic tree of 789 Hg K genomes is illustrated in figure 2. Of the 182 Ashkenazi Hg K mtDNAs, 179 could be readily assigned to K1a1b1a, K1a9, or K2a2a (table 6), demonstrating that virtually all Hg K genomes present in the Ashkenazi Jews belong to these three distinct monophyletic clades and, in turn, comprise 30% of all Ashkenazi maternal lineages. Of 123 K1a1b1a mtDNAs (fig. 2 and table 6), 122 were from Jews—113 of Ashkenazi and 9 of Spanish-exile ancestry (6 Bulgarian, 2 Italian, and 1 Turkish). The only non-Jewish K1a1b1a mtDNA that shared the HVS-I haplotype 16223-16224-16234-16311 with the Ashkenazi Jews was found in a subject from Hmelnitski, a Ukrainian town with a major Jewish settlement until the Second World War. As for K1a9, 48 of the 789 K mtDNAs were members of this subHg (fig. 2 and table 6), and 47 were from Jews—41 Ashkenazi, 4 Spanish exile (2 Bulgarian, 1 former Yugoslavian, and 1 from Turkey), 1 from Iraq, and 1 from Syria. A subHg K1a9 mtDNA was found in one Hungarian of unidentified ethnic or religious affiliation. Finally, 28 (25 Ashkenazi Jews, 1 Bulgarian Jew, 1 Georgian Jew, and 1 Azerbaijani Jew) of the 789 K samples belonged to subHg K2a2a (fig. 2 and table 6). This subHg and its parental Hg were not found in any of 11,452 non-Jewish samples.

The same principles were followed to investigate the high incidence of Hg N1b in Ashkenazi Jews. Hg N1b

Table 3

Source and Ethnic Origin of the 121 Complete mtDNA Sequences

Sample[a]	Source	Population
D5939	Present article	Algerian Jew
D1123	Present article	Ashkenazi Jew
D1139	Present article	Ashkenazi Jew
D1282	Present article	Ashkenazi Jew
D1369	Present article	Ashkenazi Jew
D1606	Present article	Ashkenazi Jew
D1829	Present article	Ashkenazi Jew
D4207	Present article	Ashkenazi Jew
D4232	Present article	Ashkenazi Jew
D4262	Present article	Ashkenazi Jew
D4481	Present article	Ashkenazi Jew
D4777	Present article	Ashkenazi Jew
D5551	Present article	Ashkenazi Jew
D5677	Present article	Ashkenazi Jew
D5077	Present article	Cherkes
D1311	Present article	Druze
D1320	Present article	Druze
D1626	Present article	Druze
D1701	Present article	Druze
D1520	Present article	Moroccan Jew
D5908	Present article	Moroccan Jew
D5257	Present article	Palestinian
D1435	Present article	Yemenite Jew
D1656	Present article	Yemenite Jew
D4757	Present article	Yemenite Jew
E1089	Present article	Lebanese
E1221	Present article	Saudi
E1365	Present article	Syrian
A	Achilli et al. 2005	Italian
H	Herrnstadt et al. 2002	Unknown
F	Finnila et al. 2001	Finnish
K	Coble et al. 2004	Unknown
Ma	Maca-Meyer et al. 2001	Iberian
Mi	Mishmar et al. 2003	Unknown
P	Palanichamy et al. 2004	Indian

[a] Serial numbers of mtDNAs from the literature were not changed.

Figure 1 Most parsimonious tree of complete Hg K mtDNA sequences. The tree is rooted in Hg U* and includes 121 mtDNAs, of which 28 are novel and 93 were reported elsewhere (12 sequences from Finnila et al. [2001], 1 from Maca-Meyer et al. [2001], 47 from Herrnstadt et al. [2002] [including the control-region information that was not reported], 1 from Mishmar et al. [2003], 28 from Coble et al. [2004], 2 from Palanichamy et al. [2004], and 2 from Achilli et al. [2005]). The genotyping information from Finnila et al. (2001) included herein corrects several inaccuracies that were reported elsewhere for the control-region phylogeny. Mutations are shown on the branches and are transitions unless the base change is explicitly indicated. Deletions are indicated by a "D" following the deleted nucleotide position. Insertions are indicated by a dot followed by the number and type of inserted nucleotide(s). Underlined nucleotide positions occur at least twice in the tree. An asterisk (*) at the end of a nucleotide position denotes a reversion. The information of the reported samples is presented in table 3. Samples in blue are from Ashkenazi Jews. Nucleotide positions in red were assayed in the entire set of 789 mtDNAs belonging to K. To resolve three possible reticulations, reversions at nucleotide 7082 in sample A01 and nucleotides 10398 and 11485 in sample H510 and a homoplasy at nucleotide 7927 for subHg K1a1a and K1a8 were assumed. The tree was drawn by hand from networks constructed using program Network 4.1.0.0. We applied the reduced-median algorithm ($\rho = 2$) followed by the median-joining algorithm ($\epsilon = 2$), as described at the Fluxus Engineering Web site. The control-region positions were twofold down-weighted, with respect to coding-region positions. The hypervariable nucleotides 16182 and 16183 in HVS-I and the indels at nucleotides 309, 315, and 524 in HVS-II were excluded. Coding-region mutations at nucleotides 750, 1438, 1811, 2706, 4769, 7028, 8860, 11467, 11719, and 12308 are not shown unless reverted. Ethnic designation of the sequences was available for the sequences reported by Finnila et al. (2001), Maca-Meyer et al. (2001), Palanichamy et al. (2004), and Achilli et al. (2005), and for the novel samples reported herein (table 3).

Table 4

Hg K Database

The table is available in its entirety in the online edition of *The American Journal of Human Genetics.*

is virtually absent in Europeans but appears at frequencies of ~3% or higher in those from Levant, Arabia, and Egypt (Richards et al. 2003; Kivisild et al. 2004; unpublished results of Tartu and Haifa groups). This Hg is defined by the transversion C16176G, relative to the revised Cambridge Reference Sequence (rCRS) (Andrews et al. 1999), and is reported in all non-Jewish Near Eastern N1b mtDNAs. However, all but one of the Ashkenazi N1b mtDNAs were found to harbor a C→A transversion at nucleotide position 16176. To assess whether this was another Ashkenazi founding lineage, we followed the same approach applied to Hg K. We randomly chose two mtDNAs for complete sequencing and identified several shared mutations that were absent in a previously reported N1b complete sequence (Maca-Meyer et al. 2001) (fig. 3). We then examined nucleotide positions 11928 and 12092—sites of two of the private mutations—in 82 N1b mtDNAs that were available to us (table 7). Fifty-six of the 57 Ashkenazi Jews, the Spanish-exile Jews, and the Moroccan Jew, who shared 16176A, could be assigned to this same lineage. The single Ashkenazi and all other mtDNAs with 16176G did not harbor the mutations at 11928 and 12092. Curiously, the 16176A transversion probably occurred twice in the phylogeny of N1b. Indeed, we have found lineages with 16176A in Slavic-speaking populations both in the Balkans and in Ukraine, but these possessed HVS-I mutations different from those present among the Jews, and they did not harbor the mutations at 11928 and 12092; thus, they are clearly phylogenetically distinct from N1b genomes of the Ashkenazi Jews.

In total, we have identified four Ashkenazi founding lineages, three within Hg K and one in Hg N1b, deriving from only four ancestral women and accounting for fully 40% of the mtDNAs of the current Ashkenazi population (~8,000,000 people). The most dominant of these lineages, K1a1b1a, encompasses 62% of the Ashkenazi K mtDNAs, which translates into 19.4% of contemporary Ashkenazi Jews, or ~1,700,000 people. The second most common lineage is within Hg N1b and corresponds to an additional 800,000 people. We compared the pattern of lineage distribution seen in Ashkenazi Jews with a global database of ~30,000 mtDNAs, 13,359 of which are from populations in which Hg K and N1b are present (table 6), and we could not detect anything similar. For instance, in European populations, the closest the existing literature offers is the database of 192 complete Finnish mtDNA sequences (Finnila et al. 2001). Though these mtDNAs were nonrandomly selected from a larger control-region database of nearly 500 individuals living in a north-central region of Finland, a typical, frequently derived, phylogenetically recent lineage comprises merely 3%–4% of the total. In the study of Coble et al. (2004), the entire mtDNA sequences of 241 individuals matching 1 of the 18 most common control-region haplotypes in European Caucasian populations were determined. The frequency of these 18 haplotypes was found to account for only 20.8% of the European Caucasian mtDNAs—a very significant difference compared with the Ashkenazi Jews, for which four complete sequence haplotypes comprise 42% of the mtDNAs. Furthermore, even mtDNAs with the same control-region motif were rarely found to completely match at the coding-region level, with an average coding-region mismatch of 6.2 mutations observed within the 241 completely sequenced mtDNAs. This would correspond to an approximate average date of 15,000–16,000 years ago for the most recent common ancestor, under the same assumptions used to calculate the coalescence of the Ashkenazi lineages (see below). Therefore, in contrast to Thomas et al. (2002), we conclude that a significant founding event is, indeed, readily evident in the maternal history of Ashkenazi Jews. Thomas et al. (2002) based their analysis on a portion of the HVS-I region (nucleotide positions 16093–16383) from 78 subjects and, in contrast to Torroni et al. (1996) and Behar et al. (2004), reported that the most frequent sequence found in their Ashkenazi sample (their "modal haplotype") is identical to the HVS-I sequence of the so-called rCRS, with a frequency not significantly different from that observed among their European host populations. However, it is now well established that the rCRS sequence in HVS-I can be found in several different Hgs—H, HV, pre-HV, U, and a paraphyletic group R*. Thus, although the inference by Thomas et al. (2002) soundly emanates from the results of their limited sample size and subfragment analysis, the conclusion reached is not borne out with the use of control-region analysis in a larger sample set. Moreover, even under the improbable assumption that all rCRS mtDNAs studied by Thomas et al. (2002) belonged to Hg H, additional genealogical resolution is not provided, since such a motif can be found in a diverse variety of subclades of Hg H, many of which diverged from a common ancestor >10,000 years ago (Finnila et al. 2001; Herrnstadt et al. 2002; Achilli et al. 2004; Loogväli et al. 2004; Pereira et al. 2005), a time frame uninformative for studying recent ancestries in general and the founding of the Ashkenazi

Table 5

Description of the Populations

The table is available in its entirety in the online edition of *The American Journal of Human Genetics.*

Table 6

SubHg Affiliation of the 789 mtDNAs Belonging to K

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

population in particular. Complete sequencing of Ashkenazi mtDNA genomes with the rCRS sequence motif in their HVS-I region, as demonstrated for Hg K and N1b in the current study, would be a straightforward approach to reveal additional maternal founder lineages among such mtDNAs.

The coalescence time for each of the four lineages was calculated independently in two ways: from HVS-I and from the entire coding-region sequence (table 8). When we analyzed the coding-region data, we used only the information obtained from the novel complete mtDNA genomes in which Ashkenazi ancestry was firmly established (table 1). Historical records suggest the establishment of the Ashkenazi population during the 7th and 8th centuries in the Rhine valley by a few migrating families arriving from northern Italy (Ostrer 2001). The estimated number of Ashkenazi Jews in the 12th and 13th centuries is 25,000, a number sufficient to keep allelic frequencies largely in balance against random genetic drift within ∼30–40 generations. Therefore, the demographic history suggests that the crucial founder events are likely to have occurred sometime before the 12th century, whereas the expansion phase of Ashkenazi Jewry in Europe, with its ups and downs, has lasted more than a millennium. Our coalescence analysis is in agreement with this assumption, since the expansion time calculations for the four Ashkenazi lineages point to the past 20 centuries and are close to the historical founding period of the Ashkenazi population. It should be noted that, despite relatively large SDs, the coalescence time estimates for the four maternal founder Ashkenazi lineages are at least an order of magnitude lower than those obtained for the corresponding parental clades, which clearly signals a recent beginning of their expansion.

There are two fundamental questions with respect to the geographic origin of the Ashkenazi founding lineages. First, were these lineages a part of the mtDNA pool of a population ancestral to Ashkenazi Jews in the Near East, or were they established within the Ashkenazi Jews later in Europe, as a result of introgression from European or Eurasian groups? Second, where did these lineages expand? The observed global pattern of distribution renders very unlikely the possibility that the four aforementioned founder lineages entered the Ashkenazi mtDNA pool via gene flow from a European host population. For example, in databases of HVS-I sequences of British, Irish, German, French, or Italian subjects, these Ashkenazi sample founder lineage sequences were not observed (Baasner et al. 1998; Lutz et al. 1998; Pfeiffer et al. 2001). Furthermore, the non-Ashkenazi Jewish populations sharing the Ashkenazi mtDNA Hg K lineages turn out to be from Jewish communities that trace their origins to the expulsion from Spain in 1492. Either a shared ancestral origin of the two groups or, alternatively, a postexile admixture between neighboring Ashkenazi and Spanish-exile Jewish populations may explain the sharing of these maternal lineages. However, the very presence of the Ashkenazi founding lineages, albeit at low frequencies, in North African, Near Eastern, and Caucasian Jews, supports a common Levantine ancestry. The maternal subclade from which the Ashkenazi mtDNA lineage K2a2a arose was not found in any other of the populations reported herein (table 6). The Ashkenazi K1a9 and K1a1b1a lineages were not found in non-Jews, with the exception of the former in a single Hungarian and the latter in a single Ukrainian, both of unknown ethnicity. However, it is of interest that K1a1b1a sister lineages, which share with it a common ancestry at the internal nodal level of subclade K1a1b1 (fig. 2), can be found in Portugal, Italy, France, Morocco, and Tunisia (table 6). This reveals that this particular limb of the Hg K phylogenetic tree is of a wider Mediterranean presence and origin. Likewise, the distribution of Hg N1b in southwestern Asia and North Africa (Rando et al. 1998; Richards et al. 2000) supports a Near Eastern, rather than a European, origin for this Hg. It is noteworthy that our extensive sample set from the Caucasus (table 5) does not offer any hint that the four dominant Ashkenazi mtDNA lineages might have arrived from this region. However, it can be concluded that, irrespective of where exactly the mutations defining these Ashkenazi lineages arose, their expansion clearly took place during the time period of the sojourn of the Ashkenazi population in Europe.

It is important to note that, although our findings clarify the restricted nature of mtDNA lineage diversity carried by Ashkenazi Jewry and provide evidence of a founder event specifically for the matrilineal ancestry, some questions remain unsolved and could be the focus of future studies. First, our findings are not sufficient to answer questions about the extent and location of the ancestral deme from which Ashkenazi Jewry, as a population, arose. It is possible that, for the MSY and autosomal loci, different patterns might be observed. Second, our findings cannot provide a quantitative assessment of the Ashkenazi population bottleneck, be-

Table 7

N1b Haplotypes in Jews and Near Eastern Non-Jews

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

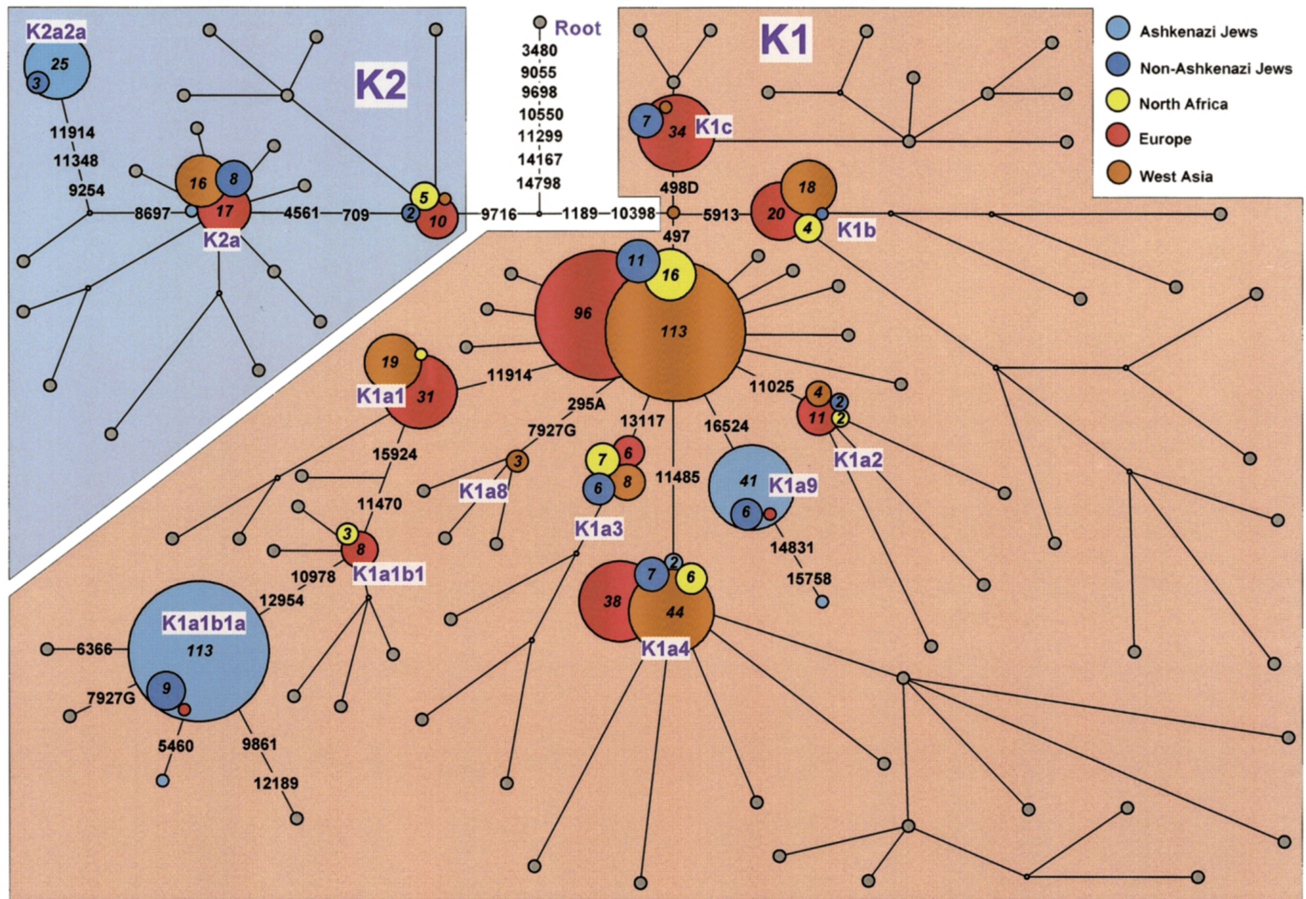


Figure 2 The location of the three Ashkenazi lineages (*light blue*) belonging to Hg K in a global set of complete K sequences. For the network construction, first, a topology network encompassing the 91 branches found in the total set of 121 complete K sequences was drawn using the same principles described in figure 1. For simplicity, we used only one complete mtDNA sequence per node, and we indicate only the coding-region nucleotide positions that are relevant in the context of the three Ashkenazi lineages K1a1b1a, K1a9, and K2a2a, with the exception of positions 497, 498D, and 16524, which define K1a, K1c, and K1a9, respectively. Branch lengths were sometimes distorted to increase legibility. Then, the topology network was compared with a second phylogenetic tree of the entire set of 789 K mtDNAs analyzed hierarchically for the diagnostic mutations (*red*). All Jews of North African, Caucasian, Near Eastern, and Spanish-exile ancestry were aggregated into the category of "non-Ashkenazi Jews." Likewise, all non-Jewish samples from Anatolia, the Caucasus, central and southwestern Asia, and the Near East were aggregated into the category of "West Asia." Circle sizes are proportional to the haplotype frequency in the sample.

cause the fraction of the ancestral maternal deme represented by the four particular founding lineages that we have identified cannot be determined. Third, the effects on the nuclear genome of the maternal founding event, whose mtDNA population genetic imprint we have observed, is influenced by other parameters, such as admixture, recombination events, and paternal contribution, that have occurred through the generations. The quantitative contribution of each of these parameters is not known, and these are likely to result in a different pattern for the nuclear genome, compared with that of mtDNA. Nevertheless, the analysis of mtDNA sequence variation enables the detailed and quantitative elucidation of a maternal founder event, which cannot be inferred from analysis of other genomic regions.

In conclusion, the present study highlights the importance of a combined phylogenetic/phylogeographic strategy that includes complete mtDNA sequence analysis to accurately portray maternal founding events and to infer conclusions relevant to both shared ancestries and population-level effects that shaped the mtDNA gene pool in a given population. In the Ashkenazi Jews, this approach enabled us to reconstruct a detailed phylogenetic tree for the major Ashkenazi Hgs K and N1b, allowing the detection of a small set of only four individual female ancestors, likely from a Hebrew/Levantine mtDNA pool, whose descendants lived in Europe and carried forward their particular mtDNA variants to 3,500,000 individuals in a time frame of <2 millennia. This founding event(s), established here as a dominant

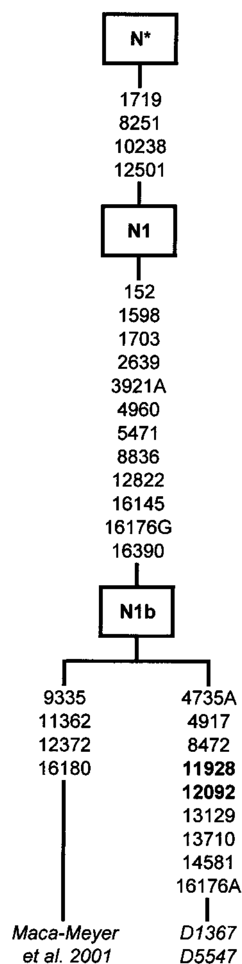


Figure 3 Most parsimonious tree of complete Hg N1b mtDNA sequences. The tree encompasses the two complete Ashkenazi N1b sequences and one previously published sequence (Maca-Meyer et al. 2001). The same considerations detailed in figure 1 are relevant here. Nucleotide positions 11928 and 12092 (*bold*) were examined in all N1b mtDNAs.

mechanism in the genetic maternal history of the Ashkenazi Jews, is a vivid example of the founder effect originally described by Mayr (1963) 4 decades ago.

Acknowledgments

We thank the individuals who provided samples for this study, the National Laboratory for the Genetics of Israeli Populations, which also provided samples, and Vered Friedman and Guennady Yudkovsky for technical assistance. This research was supported in part by Israeli Science Foundation grants (to K.S.), the Annie Chutick Endowment and Technion (to K.S.), the Estonian Science Foundation (to E.M. and T.K.), European Union Framework Programme Genemill and Genera grants (to R.V.), the Consiglio Nazionale della Ricerche–Ministro dell'Istruzione dell'Univerita e della Ricerca (CNR-MIUR) Genomica Funzionale-Legge 449/97 (to A.T.), the Fondo Investimenti Ricerca di Base 2001 (to A.T.), the Progetti Ricerca Interesse Nazionale 2005 (to A.T.), the National Science Foundation (to N.H.), and Programa Operacional Ciência, Tecnologia e Inovação (to A. Amorim).

Table 8

Coalescence Ages of the Four Major Ashkenazi mtDNA Lineages

		Age[a]		SD[b]	
Hg and SubHg	*N*	ρ	Years	σ	Years
HVS-I:					
K1a1b1a:					
16224 16234 16311	46	.09	1,755	.04	877
16223 16224 16234 16311	67	.16	3,313	.07	1,506
K1a9:					
16224 16311	5	0	0	.20	4,036
16093 16224 16311	36	0	0	.03	561
K2a2a:					
16224 16311	25	.08	1,614	.06	1,142
N1b:					
16145 16176A 16223	56	.04	721	.03	510
Coding region:					
K1a1b1a:					
Entire branch	14	.36	1,835	.16	821
Ashkenazi	7	.14	742	.14	742
K1a9:					
Entire branch	6	.33	1,713	.24	1,211
Ashkenazi	4	.50	2,569	.35	1,816
K2a2a:					
Ashkenazi	2	0	0	.50	2,569
N1b:					
Ashkenazi	2	0	0	.50	2,569

[a] Coalescence time was calculated by considering one transition between nucleotides 16090–16365 (HVS-I) equal to 20,180 years (Forster et al. 1996) and one base substitution between nucleotides 577–16023 (coding region) equal to 5,138 years (Mishmar et al. 2003).

[b] SD of the ρ estimate (σ) was calculated as reported elsewhere (Saillard et al. 2000).

Appendix A

Methods

Sampling

A total of 583 Ashkenazi Jewish and 1,111 non-Ashkenazi Jewish samples were analyzed for mtDNA Hg-specific markers, identifying 186 and 63 samples, respectively, belonging to Hg K (HVS-I data for 565 of the Ashkenazi samples have been reported elsewhere [Behar et al. 2004]). Of the 186 samples found among Ashkenazi Jews and the 63 Hg K samples found among non-Ashkenazi Jews, 182 and 62, respectively, were available for further genotyping. Next, we identified Hg K samples in all population sample collections available from the laboratories in Haifa, Tartu, Pavia, Porto, and Paris, and we reported only populations in which Hg K samples were found and available for further genotyping. We

evaluated a total of 11,452 samples from 67 populations and identified 636 Hg K samples, 545 of which were available for further genotyping. Table 5 details the total number of samples available from each population, the number of Hg K samples within each population, the number of Hg K samples from each population that were technically available for genotyping in this study, the reporting laboratory, and the first study in which the samples were reported (Bermisheva et al. 2002; Behar et al. 2004; Pereira et al. 2004; Quintana-Murci et al. 2004). We then examined the same data set of Jewish samples described above for the presence of Hg N1b samples and identified a total of 57 and 11 Hg N1b samples in Ashkenazi and non-Ashkenazi Jews, respectively. We also reported the results of 14 Hg N1b samples found in the same Druze, Palestinian, and Bedouin samples included in the sample set described above (table 5), and we compared the Ashkenazi N1b Hg mtDNAs with the unpublished database of 8,644 Caucasian, European, Near and Middle Eastern, and North African subjects included in this study and available in Tartu (table 5). All samples reported herein were derived from blood, buccal swab, or blood cell samples that were collected with informed consent, in accordance with procedures approved by institutional human subjects review committees in their respective locations. All subjects reported the birthplace of their mothers, grandmothers, and, in most cases, great-grandmothers.

Control-Region Sequencing

Sequences of the control region were determined from position 16024 to 00300, by use of the ABI Prism Dye Terminator cycle-sequencing protocols developed by Applied Biosystems (Perkin-Elmer). Control-region sequence data were used to define haplotypes within the Hgs. The control-region information reported from Haifa and Paris extends from 16024 to 00300. The control-region information reported from Tartu spans from 16024 to 16400, corresponding to HVS-I. The control-region information reported from Pavia extends from 16024 to at least 00200 and, in most cases, up to 00300. The control-region information reported from Porto extends from 16024 to 16365 and from 00072 to 00340. The hypervariable positions 16182 and 16183 in HVS-I and the indels at positions 00309 and 00315 in HVS-II were excluded from the analysis (tables 1 and 4).

Complete mtDNA Sequencing

DNA was amplified using 18 primers to yield nine overlapping fragments, as reported elsewhere (Taylor et al. 2001). After purification, the nine fragments were sequenced by means of 56 internal primers to obtain the complete mtDNA genome. Sequencing was performed on a 3100 DNA Analyzer (Applied Biosystems), and the resulting sequences were analyzed with the SEQUENCHER software. Mutations were scored relative to the rCRS (Andrews et al. 1999). The novel 28 Hg K and 2 N1b complete mtDNA sequences reported herein have been submitted to GenBank (accession numbers DQ301789–DQ301818). Quality control was assured as follows: first, each base pair was determined once with a forward and once with a reverse primer; second, any ambiguous base call was tested by additional and independent PCR and sequencing reactions; and third, all sequences were examined by two independent investigators.

Hg Labeling

All Hg K samples reported herein harbored the diagnostic marker G9055A, as confirmed by the RFLP reaction −9052*Hae*II. Diagnostic markers for genotyping subclades of Hg K were selected on the basis of the analyses of the complete mtDNA sequences. Positions 295, 497, 498, 1189, 4561, 7927, 8697, 8703, 8790, 9254, 9647, 9716, 10398, 11025, 11914, 13117, 11470, 11485, 15924, 10978, and 12954 were obtained by direct sequencing. Positions 1189, 5913, 9716, 10398, and 11914 were obtained by direct sequencing or by means of RFLPs +1186*Rsa*I, +5913*Hae*III, +9714*Bsa*I, +10398*Dde*I, and −11914*Tai*I or −11911*Hpy*CH4IV, respectively. All Hg N1b samples reported herein harbored the diagnostic marker T10238C, as confirmed by +10237*Hph*I. Two diagnostic positions, 11928 and 12092, derived from the complete mtDNA sequences analysis of the two Hg N1b samples, were examined by direct sequencing in all Jewish, Druze, Palestinian, and Bedouin samples.

Nomenclature

We followed a consensus Hg nomenclature scheme (Richards et al. 1998). Numbers 1–16569 refer to the position of the mutation in the rCRS (Andrews et al. 1999). We use the term “lineage” to denote a cluster of related, evolving haplotypes within an Hg. Note that haplotypes and lineages can relate to HVS-I, to the control region, or to the complete mtDNA sequence data. Nomenclature within Hg K has been the subject of some ambiguity because of the recycling of the designation “K1a” (Herrnstadt et al. 2002; Palanichamy et al. 2004). We followed the definitions of both publications for “K1” and “K2.” We followed Palanichamy et al. (2004) for the definitions of “K1a,” “K1a1,” “K1a2,” “K1b,” “K1c,” and “K2a” and corrected in this article a few inaccuracies reported for the samples labeled “SF#153” and “CH#365.” Herrnstadt et al. (2002) used the designation of “K1a” for one subHg of K. To avoid con-

fusion and because this lineage was found to be highly important in the current study, we specifically note that we assigned to this lineage the more detailed designation "K1a1b1a."

Web Resources

Accession numbers and URLs for data presented herein are as follows:

Fluxus Engineering Web site, http://www.fluxus-engineering.com/
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for the complete mtDNA sequences [accession numbers DQ301789–DQ301818])

References

Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti AS, Bandelt HJ, Semino O, Torroni A (2005) Saami and Berbers: an unexpected mitochondrial DNA link. Am J Hum Genet 76:883–886

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet 75:910–918

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Baasner A, Schafer C, Junge A, Madea B (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. Forensic Sci Int 98:169–178

Behar DM, Hammer MF, Garrigan D, Villems R, Bonne-Tamir B, Richards M, Gurwitz D, Rosengarten D, Kaplan M, Della Pergola S, Quintana-Murci L, Skorecki K (2004) MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. Eur J Hum Genet 12:355–364

Bermisheva M, Tambets K, Villems R, Khusnutdinova E (2002) Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia. Mol Biol (Mosk) 36:990–1001

Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. Nat Genet Suppl 33:266–275

Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons TJ (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int J Legal Med 118:137–146

DellaPergola S (2001) Jewish demography 2001. In: DellaPergola S, Even S (eds) Papers in Jewish demography. The Hebrew University of Jerusalem Press, Jerusalem, pp 11–33

Diamond JM (1994) Human genetics: Jewish lysosomes. Nature 368: 291–292

Finnila S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. Am J Hum Genet 68:1475–1484

Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70:1152–1171

Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R (2004) Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75:752–770

Loogväli EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, et al (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol 21:2012–2021

Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 111:67–77

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2:13

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308:1034–1036

Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge, pp 162

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R (2004) Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet 5:26

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100:171–176

Ostrer H (2001) A genetic profile of contemporary Jewish populations. Nat Rev Genet 2:891–898

Palanichamy M, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri T, Palla V, Zhang YP (2004) Phylogeny of mtDNA macrohaplogroup N in India based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75: 966–978

Pereira L, Cunha C, Amorim A (2004) Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. Int J Legal Med 118:132–136

Pereira L, Richards M, Goios A, Alonso A, Albarran C, Garcia O, Behar DM, Golge M, Hatina J, Al-Gazali L, Bradley DG, Macaulay V, Amorim A (2005) High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Res 15:19–24

Pfeiffer H, Forster P, Ortmann C, Brinkmann B (2001) The results of an mtDNA study of 1,200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic casework. Int J Legal Med 114:169–172

Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K (2004) Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. Am J Hum Genet 74:827–845

Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, Bandelt HJ (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. Ann Hum Genet 62:531–550

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo

C, et al (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276

Richards M, Rengo C, Cruciani F, Gratrix F, Wilson J, Scozzari R, Macaulay V, Torroni A (2003) Extensive female-mediated gene flow from sub-Saharan Africa into Near Eastern Arab populations. Am J Hum Genet 72:1058–1064

Richards MB, Macaulay VA, Bandelt HJ, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241–260

Risch N, de Leon D, Ozelius L, Kramer P, Almasy L, Singer B, Fahn S, Breakefield X, Bressman S (1995) Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. Nat Genet 9:152–159

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67:718–726

Taylor RW, Taylor GA, Durham SE, Turnbull DM (2001) The determination of complete human mitochondrial DNA sequences in single cells: implications for the study of somatic mitochondrial DNA point mutations. Nucleic Acids Res 29:E74

Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L (2005) Reconstructing the origin of Andaman Islanders. Science 308:996

Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, Goldstein DB (2002) Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. Am J Hum Genet 70:1411–1420

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850