MUTATION IN BRIEF

Molecular Dissection of the Y Chromosome Haplogroup E-M78 (E3b1a): A Posteriori Evaluation of a Microsatellite-Network-Based Approach Through Six New Biallelic Markers

Fulvio Cruciani[1], Roberta La Fratta[1], Antonio Torroni[2], Peter A. Underhill[3], and Rosaria Scozzari[1,4]*

[1]*Dipartimento di Genetica e Biologia Molecolare, Università "La Sapienza", Rome, Italy;* [2]*Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy;* [3]*Department of Genetics, Stanford University, Stanford, California;* [4]*Istituto di Biologia e Patologia Molecolari, Consiglio Nazionale delle Ricerche, Rome, Italy*

*Correspondence to: Rosaria Scozzari, Dipartimento di Genetica e Biologia Molecolare, Università "La Sapienza", Piazzale A. Moro 5, 00185 Rome, Italy; Tel.: (office) ++39 06 49912826; Tel.: (lab) ++39 06 49912924; Fax: ++39 06 4456866; E-mail: rosaria.scozzari@uniroma1.it

Communicated by Peter Oefner

The human Y chromosome haplogroup E-M78 (E3b1a) occurs commonly and is distributed in northern and eastern Africa, western Asia, and all of Europe. Previously, only two rarely observed internal biallelic markers (UEPs) were known within the E-M78 clade. Here we report the identification of six novel UEPs that significantly refine the phylogeny of this haplogroup. Then, we evaluate the correspondence between the newly defined sub-haplogroups and the E-M78 haplotype clusters previously identified by an 11-microsatellite loci-based network encompassing 232 chromosomes (Cruciani et al., 2004). We observed considerable correspondence between the trees generated by the two types of markers, but also noted important discrepancies between microsatellite and UEP findings. Overall, this analysis reveals that the currently visible terminal branches of the Y tree still contain a large amount of information, in terms of undiscovered biallelic markers, and that caution is needed when using the microsatellite alleles as surrogates of unique event polymorphisms. © 2006 Wiley-Liss, Inc.

KEY WORDS: Y-chromosome; microsatellite; polymorphism; haplogroup; network; human evolution

INTRODUCTION

The high genealogical resolution of the Y chromosome (The Y Chromosome Consortium, 2002; Jobling and Tyler-Smith, 2003), established by recent advances in mutation-detection technology, and the phylogeographic methodology (Avise, 2000; Underhill et al., 2000) have provided new opportunities for the study of modern human evolution. As a consequence of the continuing refinement of the Y chromosome genealogy, most of the terminal haplogroups of the tree are found to be restricted to a specific population and/or geographic region. By contrast, haplogroup E-M78 is characterized by a broad geographic distribution, which includes northern and eastern Africa, western Asia and all of Europe (Underhill et al., 2000; Bosch et al., 2001; Cruciani et al., 2002, 2004; Arredi et al.,

2004; Semino et al., 2004; Peričić et al., 2005), suggesting that it encompasses a collection of sub-haplogroups with very different evolutionary histories. The internal diversification of the E-M78 haplogroup may reveal further population structuring, allowing the evaluation of multiple temporal and spatial aspects of modern human evolutionary history, such as population range expansion, gene flow and admixture episodes.

Previously, an 11-microsatellite loci-based network of E-M78 haplotypes showed four clusters, α, β, γ and δ, with corresponding geographic structure (Cruciani et al., 2004). Three clusters (α, β and γ) were each defined by unusual Y-microsatellite alleles at $\leq$ 2 loci that have complex structure and/or likely reduced mutagenicity. The fourth cluster, δ, contained the residual haplotypes. As for unique event polymorphisms (UEPs), only two (M148 and M224) have been identified within E-M78 up to now, but both are rare and rather uninformative (Underhill et al., 2000, 2001; Bosch et al., 2001; Cruciani et al., 2002; Arredi et al., 2004; Shen et al., 2004).

To achieve a better refinement of the molecular diversity within the Y chromosome haplogroup E-M78, we analyzed about 60 kb of the non-recombining portion of the Y (NRY) for each of six E-M78 chromosomes, selected to include the widest range of internal microsatellite variation. This approach allowed the identification of six novel single nucleotide substitutions, and six new haplogroups/paragroups within the E-M78 clade. Based on these findings, we explored the issue of whether the molecular dissection of the E-M78 haplogroup into the microsatellite-based clusters previously described (Cruciani et al., 2004) is supported by its subdivision into the newly discovered binary haplogroups.

MATERIALS AND METHODS

Mutation detection

Two chromosomes from each of E-M78 microsatellite clusters α and δ, and one chromosome from each of E-M78 microsatellite clusters β and γ were analyzed by both DNA sequencing and denaturing high performance liquid chromatography (DHPLC).

Overall 60.2 kb from the transducin beta-like 1Y (TBL1Y, MIM# 400033) gene (Yp11.2) (exons 1-18 and portions of flanking intronic sequences) were analyzed for each individual, either by sequencing (44.4 kb) or by DHPLC (15.8 kb). PCR primers were designed on the basis of the TBL1Y sequence reported in Genome Browser web site (april 2003 assembly of the human genome; http://genome.ucsc.edu/) using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (detailed methods and primer sequences available on request). For sequencing, PCR primers were used to amplify dsDNA fragments 0.8-1.2 kb in length as templates. Following DNA amplification, PCR products were purified from primers, nucleotides, and salts using the QIAquick PCR purification kit (Qiagen, Hilden, Germany. http://www1.qiagen.com). Cycle sequencing was performed using the BigDye Terminator Cycle Sequencing Kit with Amplitaq DNA polymerase (Applied Biosystems, Foster City, CA, USA. http://www.appliedbiosystems.com) and an internal or PCR primer. Cycle sequencing products were purified by ethanol precipitation and run on an ABI Prism 3730XL DNA sequencer (Applied Biosystems). Chromatograms from the six sequenced individuals were aligned and analyzed for mutations using Sequencher 3.0 (Gene Codes Corporation, Ann Arbor, MI, USA. http://www.genecodes.com).

For DHPLC analysis, PCR amplified products (420-450 bases long) were analyzed for the presence of heteroduplex as described by Underhill et al. (2001).

Resequencing was performed to confirm the new mutations identified by sequencing or DHPLC analysis.

SNP genotyping

Variant (V) sites V12 (AC012068.5:g.5370A>G), V13 (AC012068.5:g.24534G>A), V22 (AC012068.5:g.42228T>C), V27 (AC012068.5:g.78322A>T) and V32 (AC012068.5:g.115092G>C) have been genotyped as RFLPs, while V19 (AC010889.3:g.134549T>C) has been analyzed by DHPLC. Summary information about the genotyping protocols for the variable sites identified in the present study is reported in table 1 (amplicon sequences available on request).

PCRs were performed in a 50 μl volume with 50 ng DNA, 200 μM each dNTP, 2.5 mM $MgCl_2$, 1 unit of Taq polymerase, and 10 pmoles of each primer. A touchdown PCR program was used with an annealing temperature decreasing from 63°C to 56°C over 14 cycles, followed by 20 cycles with an annealing temperature of 56°C. Twenty microliters of PCR product were digested with the appropriate restriction enzyme following manufacturer´s instructions. After digestion, restriction fragments were separated on a 3% (w/v) agarose gel (3:1 NuSieve:agarose) and visualized by UV.

The variable sites M148 and M224 were analyzed by DHPLC using the method and primers reported in Underhill et al. (2001).

All UEPs were surveyed in a sample of 232 E-M78 Y chromosomes previously genotyped for 11 microsatellites (Cruciani et al., 2004).

Table 1. Description of the new Y chromosome polymorphisms within E-M78 clade

Marker	Mutation[a]	Primers	Product length	Genotyping
V12	g.5370A>G	V12 FOR: CAAAGTTTATTTTCAAAGGGGAGA	439 bp	RFLP (BsgI)
		V12 REV: CCATAAAGTTGGGTTGAAGGAG		
V13	g.24534G>A	V13 FOR: GGATGCTGCTAAACATCCTACA	235 bp	RFLP (AciI)
		V13 REV: ATCCCATCTCAATCCCTTAACA		
V19	g.134549T>C	M148 FOR: AACAGAATTATCAGGAAAAGGTTT	568 bp	DHPLC (54°C)
		M148 REV: TTTTACTTGTTCGTGTACTTTCAA		
V22	g.42228T>C	V22 FOR: AATGCCTCAACTTACAGAAATGG	289 bp	RFLP (MmeI)
		V22 REV: CACTGACCAGAAACAGCATGAG		
V27	g.78322A>T	V27 FOR: CTCCTCAAGCACCTGTACTGTC	360 bp	RFLP (PvuII)
		V27 REV: GCTCGGTGACTCTGGAGAAC		
V32	g.115092G>C	V32 FOR: GCAAAATCCCAGAACATCATT	355 bp	RFLP (MnlI)
		V32 REV: TCATTGACCCAAAGCAGACA		

[a]DNA mutation numbering is based on gDNA reference sequences AC012068.5 (markers V12, V13, V22, V27, and V32) and AC010889.3 (V19).

Data analysis

Analysis of molecular variance (AMOVA; Excoffier et al., 1992) was performed using Arlequin, version 3.0 (Excoffier et al., 2005; http://lgb.unige.ch/arlequin/). Molecular distances between binary sub-haplogroups/paragroups or microsatellite haplotypes within haplogroup E-M78 were taken into account resulting in estimates of Φ_{st} and R_{st} values, respectively.

Nonmetric multidimensional scaling (MDS) analyses (software newMDSX, http://www.newMDSX.com) based either on pairwise Φ_{st} or pairwise R_{st} were used to analyze the relationship among groups of populations from 9 geographic regions: western Europe (N=15), central/eastern Europe (N=16), southern Europe (N=58), Balkans (N=33), Anatolia (N=16), Middle East (N=11), north-western Africa (N=30), north-eastern Africa (N=12), and eastern Africa (N=39). Two Y chromosomes from sub-Saharan Africa were omitted from this analysis.

Statistical significance of the correlation between Φ_{st} and R_{st} matrices of pairwise distances among populations was assessed through Mantel test using the program Arlequin version 3.0.

Fisher's exact test was performed with the Metropolis algorithm and the software RxC (http://www.marksgeneticsoftware.net).

RESULTS AND DISCUSSION

E-M78 microsatellite diversity

Previous work has shown that Y chromosomes belonging to haplogroup E-M78 can be subdivided on the basis of microsatellite variation and network analysis into four well-defined clusters (Cruciani et al., 2004). An analysis of these data by AMOVA generates an R_{st} estimate among microsatellite clusters as high as 0.58 (P<10^{-3}). This is a rather unexpectedly large value for a terminal node of the Y chromosome tree, consistent with highly divergent founding haplotypes for the different clusters. Clusters α, β and γ are typically restricted to specific geographic areas (cluster α in Europe, β in northern Africa and γ in eastern Africa), while the fourth cluster - δ - is widespread throughout Europe, west Asia and Africa, albeit at relatively low frequencies (Fig. 1A). Accordingly, the portion of total genetic variance among four geographic regions (Europe, west Asia, northern Africa and eastern Africa) explained by differences between microsatellite haplotypes is relatively low (R_{st}= 0.27; P< 10^{-3}) as compared with that obtained when the haplotypes belonging to cluster δ are omitted from the analysis (R_{st}= 0.52; P< 10^{-3}).
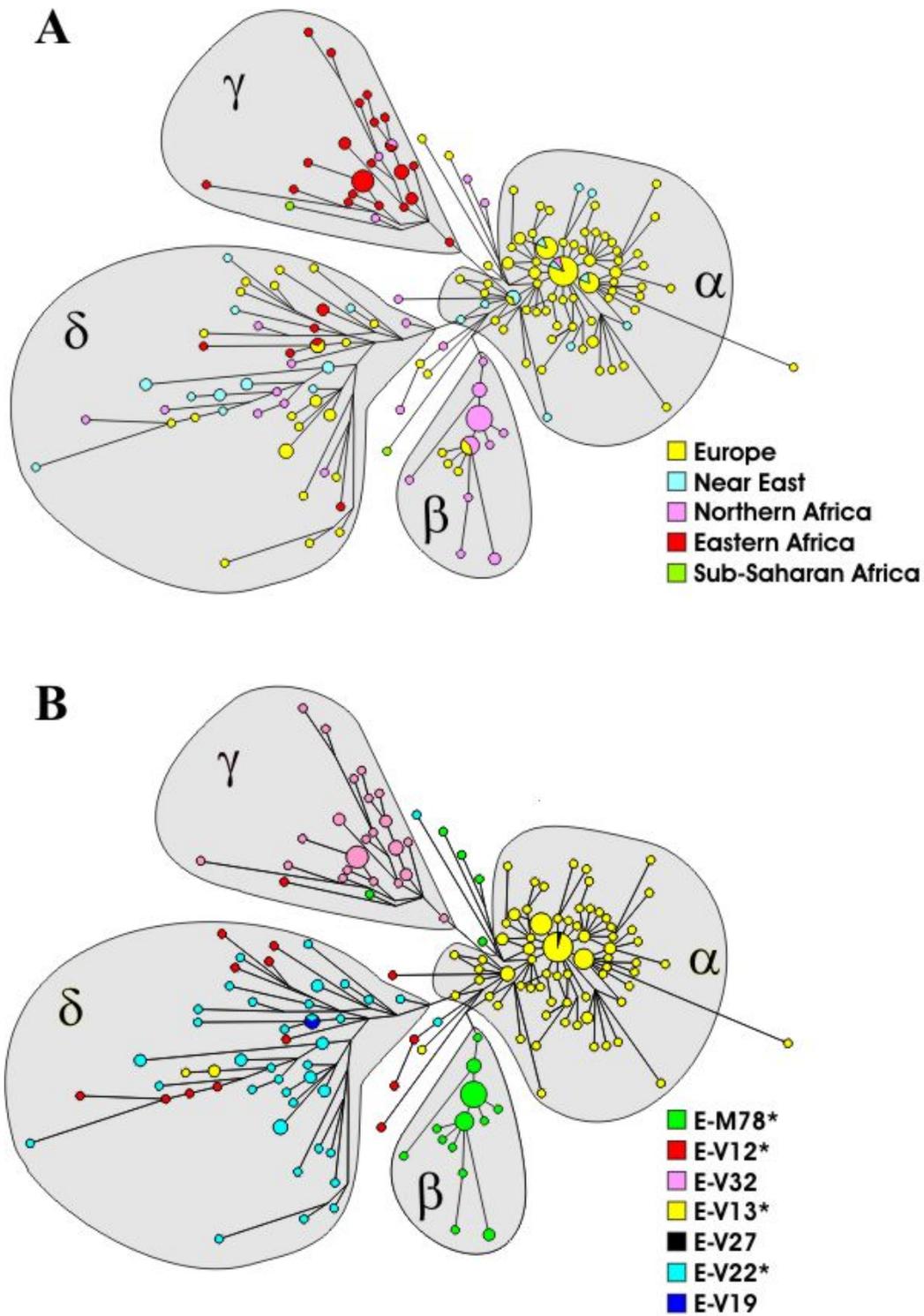
Figure 1. Eleven-microsatellite network of E-M78 chromosomes (modified from Fig. 2B of Cruciani et al., 2004). Microsatellite haplotypes are represented by circles filled with colors corresponding to geographic regions (**A**) and binary haplogroups/paragroups (**B**).

New UEPs within E-M78

The analysis of about 60 kb of the NRY in each of six E-M78 chromosomes, selected to include the widest range of microsatellite diversity, lead to the identification of five novel single nucleotide substitutions, while a sixth one was identified by DHPLC analysis performed on the same amplicon containing the M148 marker (Underhill et al., 2000). The molecular survey of selected E-M78 chromosomes permitted us to reconstruct the phylogenetic relationships among the six new and the two previously described UEPs (Fig. 2): the UEPs V12, V13 and V22 define first order branches of haplogroup E-M78, while M148, M224, V32, V27 and V19 define terminal nodes. Finally, to test the accuracy of the global E-M78 Y-chromosomal tree based on microsatellites, a total of 232 Y chromosomes, among those previously reported to belong to E-M78 and typed for 11 microsatellites (Cruciani et al., 2004), were screened for the six UEPs detected through sequence and DHPLC analysis and for the two UEPs previously described. M148 and M224, previously identified in one southern Asian and one western Asian individual, respectively (Underhill et al., 2000; Shen et al., 2004), were not observed, V27 and V19 proved to be private variants (one and two subjects, respectively), while V12, V13, V22 and V32 were relatively frequent in the global sample.
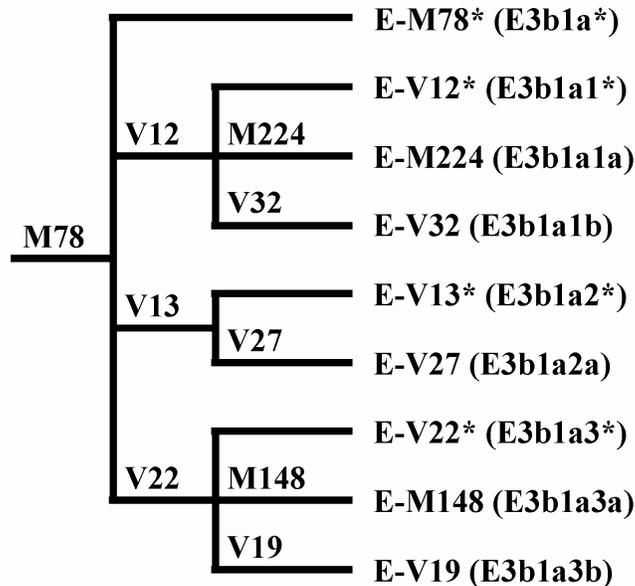

Figure 2. Phylogenetic tree of haplogroup E-M78, based on eight UEPs, six of which are reported here for the first time (V12, V13, V19, V22, V27, and V32). Haplogroup nomenclature is by mutation and by lineage (within brackets). The latter takes into account the revised topology of the E3b haplogroup as reported by Cruciani et al. (2004).

Congruence between microsatellite clusters and haplogroups/paragroups

The superposition of the new biallelic markers on the microsatellite network is shown in Fig. 1B. The relationships between microsatellite-based clusters and binary haplogroup affiliations are shown in Table 2. There is a strong correspondence between the microsatellite and UEP findings. Genetic differentiation among UEP-defined groups of microsatellite-haplotypes is quite high (R_{st} = 0.59; P<10^{-3}), and very similar to that, discussed above, obtained by the grouping of the haplotypes into clusters. We also explored the correspondence between the two types of markers at a geographical level. There is a high and significant correlation observed between pairwise binary-haplogroup-based and microsatellite-haplotype-based distances among nine geographic regions (r = 0.91; P< 10^{-3}). The two-dimensional plots from MDS show similar clusters of populations (Fig. 3), with slight differences possibly due to differential sensitivity of Φ_{st} and R_{st} to different mutational processes. North-western Africa, eastern Africa and the Balkans are well separated from each other and also from the central cluster in both

diagrams. The differences are explained to a large extent by: the haplotypes associated with paragroup E-M78* (predominantly observed in north-western Africa), haplogroup E-V32 (found almost exclusively in eastern Africa) and haplogroup E-V13 (the only subset of E-M78 observed in the Balkans). E-V13 is also commonly found in the populations of the central cluster, where, however, other E-M78 sub-haplogroups are also present.

Table 2. Relationships between microsatellite-based clusters (columns) and binary haplogroup/paragroup affiliations (rows) within haplogroup E-M78

Haplogroup	Microsatellite clusters					
	Cluster α	Cluster β	Cluster γ	Cluster δ	Not classified	Total
E-M78*	0	29	1	0	4	34
E-V12*	0	0	1	9	4	14
E-V32	0	0	34	0	0	34
E-V13	98	0	0	3	2	103
E-V22	0	0	0	45	2	47
Total	98	29	36	57	12	232[a]

[a]Number of E-M78 chromosomes identified in a global random sample of 3401 Y chromosomes (Cruciani et al., 2004). Fisher's exact test, $P<10^{-3}$.
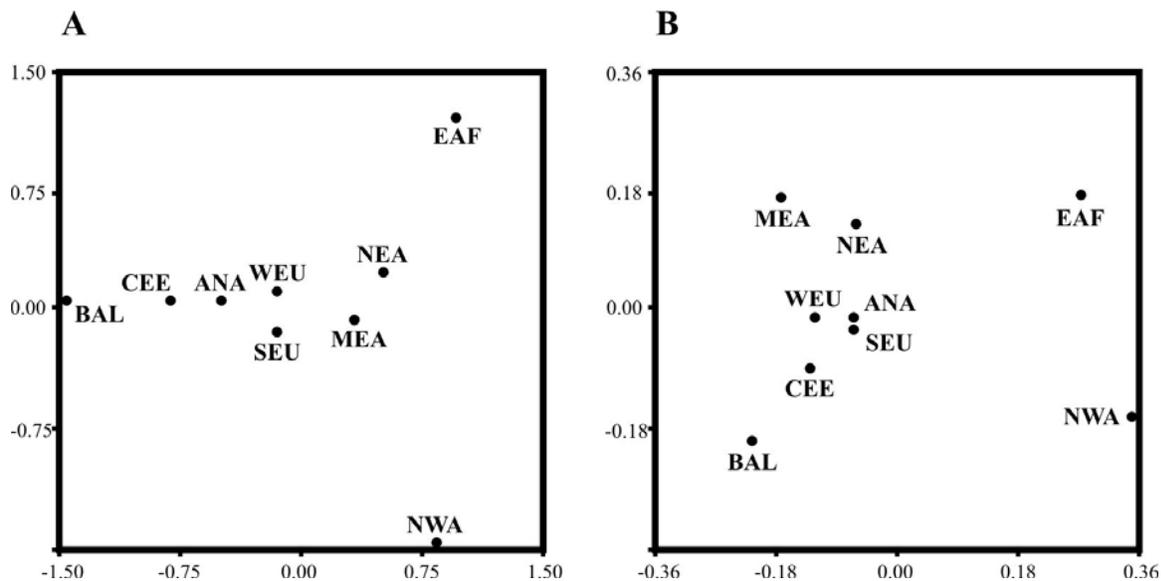

Figure 3. MDS analyses based on Φ_{st} (A) and R_{st} (B) pairwise molecular distances. Molecular distances were computed as described in Materials and Methods. Population codes are as follows: western Europe (WEU), central/eastern Europe (CEE), southern Europe (SEU), Balkans (BAL), Anatolia (ANA), Middle East (MEA), north-western Africa (NWA), north-eastern Africa (NEA), and eastern Africa (EAF).

A similar congruence between microsatellite-based clusters of haplotypes and haplogroups defined by slow-evolving UEPs has also been observed in previous studies (Forster et al., 2000; Zegura et al., 2004; Kayser et al., 2005; Xue et al., 2005; Mitchell et al., 2006). However, such a correspondence is not expected, since recurrent/parallel mutations at rapidly evolving multiallelic markers can obscure phylogenetic relationships among chromosomes. Microsatellite loci differ widely in their mutation rates, thus the use of a weighting scheme that takes into account these differences can help to identify minimally reticulated networks, approaching the true trees (Forster et al., 2000). Moreover, multi-step mutational events producing unusually short microsatellite alleles are expected to behave as pure UEPs, as compared to other alleles of the same locus, because of low allele-specific mutation rates (Brinkmann et al., 1998). For example, cluster γ in Fig. 1 is characterized by the short 11-repeat allele at the DYS19 locus, which is most likely generated by a multiple-repeat deletion event. Similar associations

between sub-networks defined by microsatellite short alleles and UEP-defined haplogroups were also recognized in previous studies (Zei et al., 2003; Di Giacomo et al., 2004). Finally, the use of a very large set of simple microsatellites (Kayser et al., 2004), could help to define clusters of closely related chromosomes sharing a common origin (Xue et al., 2005). A highly statistically significant positive correlation between Y-UEP-based and Y-microsatellite-based genetic distance matrices, similar to that observed in the present study, was found by Kayser et al. (2005) in a study of Polish and German populations. Based on such a finding and in consideration of the relatively high mutation rate of microsatellite loci, the significant genetic differentiation observed between the two neighboring populations was interpreted as being to a large extent the consequence of very recent historical events (Kayser et al., 2005). Coalescence ages for the well-defined E-M78 clusters α, β and γ here analyzed have been estimated and reported to be relatively old (in the range of 3.2-12.9 ky, including 95% CIs) (Cruciani et al., 2004). This implies that the differentiation among clusters, and their correspondence with binary haplogroups, can be retained for a long time if, for each cluster, founder events involve very different haplotypes in term of microsatellite composition. Under these circumstances, the Y-UEP-based and Y-microsatellite-based power of discrimination would be equivalent, as observed, for example, with the two ancient European binary haplogroups R1a and P*(xR1a) and their associated microsatellite haplotypes (Kayser et al., 2005; Roewer et al., 2005).

Discrepancies between microsatellite clusters and haplogroups/paragroups

Despite the major congruence between E-M78 microsatellite clusters and binary haplogroups, there are also important discrepancies between the trees generated by the two types of markers. First, the large majority of the δ chromosomes belong to the clades E-V22 and E-V12*, but a few representatives of both clades are found outside the δ cluster. Within δ, E-V22 and E-V12* chromosomes are intermingled and not clearly differentiated by their microsatellite haplotypes. Second, all of the cluster β chromosomes belong to paragroup E-M78*. However, E-M78* also includes some non-β chromosomes which are highly differentiated in their microsatellite haplotypes. Third, there is a striking correspondence between the microsatellite clusters α and γ and the binary haplogroups E-V13 and E-V32, respectively (Fig. 1B). However, while all cluster α chromosomes belong to E-V13, some of the E-V13 chromosomes are not contained in such a cluster. Conversely, all of the E-V32 chromosomes fall within cluster γ (defined by the rare DYS19 11-repeat allele), but two γ chromosomes are members of paragroups E-V12* and E-M78*.

Taking into account the above data, the previously described European cluster α and the northern African cluster β are indeed confirmed as monophyletic groups of chromosomes, that, very likely, have their own defining binary marker yet to be discovered. Cluster α chromosomes constitute a major branch of the binary haplogroup V13, which, in turn, includes also a few, highly differentiated chromosomes - previously classified either in cluster δ or unclassified. All 29 chromosomes within cluster β belong to the paragroup E-M78*, which is relatively rare and almost exclusively restricted to a single geographic region (i.e. northern Africa), thus a common origin for at least a large part of these is likely.

Different scenarios characterize clusters δ and γ . The presence of three E-V13 chromosomes within cluster δ and the exclusion of some E-V12 and E-V22 chromosomes demonstrate that cluster δ cannot be regarded as a monophyletic unit. As for cluster γ, we have established close phylogenetic relationships of its members - now classified as E-V32 chromosomes - with those belonging to E-V12* within the E-V12 clade (Fig. 2). These relationships go undetected through the microsatellite network (red and pink chromosomes in Fig. 1B), most likely due to recurrent mutations at microsatellite loci. An alternative explanation would be that V12 has mutated twice independently in the E-M78 lineage. The latter possibility is less likely, given that V12 is within a terminal branch of the Y chromosome tree; moreover, it was never found by sequencing 18 Y*(xM78) chromosomes representing deep branches of the Y chromosome phylogeny (data not shown). Thus, the new markers we have detected now offer the opportunity to explore in a better defined phylogenetic context the origin and distribution of the chromosomes belonging to haplogroup E-M78. Also, it is worth noting that twelve chromosomes that we were unable to assign to any cluster in the previous network analysis are now classified within four different haplogroups/paragroups (see Table 2 and Fig. 1B) as highly divergent microsatellite haplotypes. Thus, even though they represent only 5% of the total E-M78 chromosomes analyzed, their inclusion into the respective haplogroups/paragroups heavily affects inferences about time and place of origin of these haplogroups/paragroups.

Finally, although there is a strong correspondence between cluster γ – defined by the rare DYS19 11-repeat allele - and haplogroup E-V32, the presence, in cluster γ, of haplotypes belonging to the binary paragroups E-

M78* and E-V12* can only be explained by admitting either a paraphyletic or a polyphyletic origin for the chromosomes in the cluster.

Overall these findings indicate that caution is needed when using the microsatellite alleles as surrogates of UEPs (e.g. Malaspina et al., 2001; Cruciani et al., 2004; Di Giacomo et al., 2004; Semino et al., 2004; Sanchez et al., 2005).

Final considerations

Recent studies have shown that the molecular dissection of haplogroups from the mitochondrial DNA (mtDNA) – the other uniparental genetic system - can provide important inferences about various aspects of the evolutionary history of human populations (Achilli et al., 2004, 2005; Behar et al., 2006). In contrast to the mtDNA, the dissection of binary haplogroups of the Y tree, due to the relatively low degree of sequence variation of the Y chromosome, has been so far performed through the use of high mutation rate markers such as microsatellites (Scozzari et al., 1999, 2001; Malaspina et al., 2001; Cruciani et al., 2002, 2004; Lell et al., 2002; Zerjal et al., 2003; Semino et al., 2004; Xue et al., 2005). The data shown here represent a successful attempt to identify evolutionarily stable markers within a previously visualized terminal branch of the NRY tree. Our results suggest by inference that, even though a large number of NRY UEPs has been reported in the past few years (Underhill et al., 2000, 2001; The Y Chromosome Consortium 2002; Cinnioğlu et al., 2004; Shen et al., 2000, 2004; Wilder et al., 2004), the apparent terminal branches of the Y tree still contain a large amount of information, in terms of undiscovered biallelic markers, which can be relevant for the detection of spatial patterns attributable to both ancient dispersals and more recent events of gene flow.

REFERENCES

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon J-M, Roostalu U, Loogväli E-L, Kivisild T, Bandelt H-J, 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.

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

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, Tyler-Smith C. 2004. A predominantly Neolithic origin for Y-chromosomal DNA variation in North Africa. Am J Hum Genet 75:338-345.

Avise JC. 2000. Phylogeography: the history and formation of species. Cambridge, Massachusetts: Harvard University Press.

Behar DM, Metspalu E, Kivisild T, Achilli A, Hadid Y, Tzur S, Pereira L, Amorim A, Quintana-Murci L, Majamaa K, Herrnstadt C, Howell N, Balanovsky O, Kutuev I, Pshenichnov A, Gurwitz D, Bonne-Tamir B, Torroni A, Villems R, Skorecki K. 2006. The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. Am J Hum Genet 78:487-497.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J. 2001. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between north-western Africa and the iberian peninsula. Am J Hum Genet 68:1019-1029.

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408-1415.

Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. 2004. Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114:127-148.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Beraud Colomb E, Zaharova B, Lavinha J, Vona G, Aman R, Calì F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R. 2004. Phylogeographic

analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74:1014-1022.

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am J Hum Genet 70:1197–1214.

Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, Papola F, Ciavarella G, Cucci F, Di Stasi L, Gavrila L, Kerimova MG, Kovatchev D, Kozlov AI, Loutradis A, Mandarino V, Mammì C, Michalodimitrakis EN, Paoli G, Pappa KI, Pedicini G, Terrenato L, Tofanelli S, Malaspina P, Novelletto A. 2004. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. Hum Genet 115:357-371.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47-50.

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.

Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67:182-196.

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598-612

Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C. 2004. A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74:1183-1197.

Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, Henke J, Henke L, Hohoff C, Illing A, Jonkisz A, Kuzniar P, Lebioda A, Lessig R, Lewicki S, Maciejewska A, Monies DM, Pawlowski R, Poetsch M, Schmid D, Schmidt U, Schneider PM, Stradmann-Bellinghausen B, Szibor R, Wegener R, Wozniak M, Zoledziewska M, Roewer L, Dobosz T, Ploski R. 2005. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. Hum Genet 117:428-443.

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. Am J Hum Genet 70:192-206.

Malaspina P, Tsopanomichalou M, Duman T, Stefan M, Silvestri A, Rinaldi B, Garcia O, Giparaki M, Plata E, Kozlov AI, Barbujani G, Vernesi C, Papola F, Ciavarella G, Kovatchev D, Kerimova MG, Anagnou N, Gavrila L, Veneziano L, Akar N, Loutradis A, Michalodimitrakis EN, Terrenato L, Novelletto A. 2001. A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area. Ann Hum Genet 65:339-349.

Mitchell RJ, Reddy BM, Campo D, Infantino T, Kaps M, Crawford MH. 2006. Genetic diversity within a caste population of India as measured by Y-chromosome haplogroups and haplotypes: subcastes of the Golla of Andhra Pradesh. Am J Phys Anthropol. In press.

Peričić M, Barać Lauc L, Martinović Klarić I, Rootsi S, Janićijević B, Rudan I, Terzić R, Čolak I, Kvesić A, Popović D, Šijački A, Behluli I, Ðorđević D, Efremovska L, Bajec ÐD, Stefanović BD, Villems R, Rudan P. 2005. High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. Mol Biol Evol 22:1964-1975.

Roewer L, Croucher PJP, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M. 2005. Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116:279-291.

Sanchez JJ, Hallenberg C, Børsting C, Hernandez A, Morling N. 2005. High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. Eur J Hum Genet 13:856-866.

Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, Latini V, Varesi L, Memmi MM, Romano V, De Leo G, Gennarelli M, Jaruzelska J, Villems R, Parik J, Macaulay V, Torroni A. 2001. Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. Hum Immunol 62:871-884.

Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, Sellitto D, Arredi B, Destro-Bisol G, De Stefano G, Rickards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A. 1999. Combined use of

biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. Am J Hum Genet 65:829-846.

Semino O, Magri C, Benuzzi G, Lin AA , Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS. 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. Am J Hum Genet 74:1023-1034.

Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman MW, Oefner PJ. 2004. Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. Hum Mutat 24:248-260.

Shen P, Wang F, Underhill PA, Franco C, Yang W-H, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ. 2000. Population genetic implications from sequence variation in four Y chromosome genes. Proc Natl Acad Sci USA 97:7354-7359.

The Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339-348.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazón Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43-62.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. 2000. Y chromosome sequence variation and the history of human populations. Nat Genet 26:358-361.

Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF. 2004. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. Nat Genet 36:1122-1125.

Xue Y, Zerjal T, Bao W, Zhu S, Lim S-K, Shu Q, Xu J, Du R, Fu S, Li P, Yang H, Tyler-Smith C. 2005. Recent spread of a Y-chromosomal lineage in northern China and Mongolia. Am J Hum Genet 77:1112-1116.

Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 21:164-175.

Zei G, Lisa A, Fiorani O, Magri C, Quintana-Murci L, Semino O, Santachiara-Benerecetti AS. 2003. From surnames to the history of Y chromosomes: the Sardinian population as a paradigm. Eur J Hum Genet 11:802-807.

Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C. 2003. The genetic legacy of the Mongols. Am J Hum Genet 72:717-721.