Figure 2 Schematic phylogenetic tree of mtDNA haplogroups observed in the populations analyzed. The diagnostic mutations used to classify the whole data set are reported on the branches. Restriction enzyme sites are numbered from the first nucleotide of the recognition sequence. A plus sign (+) indicates the presence of a restriction site; a minus sign (−) indicates the absence of such a site. The restriction enzymes are designated by the following single-letter codes: a, *Alu*I; b, *Ava*II; c, *Dde*I; e, *Hae*III; f, *Hha*I; g, *Hin*fI; h, *Hpa*I; i, *Msp*I; j, *Mbo*I; k, *Rsa*I; l, *Taq*I; m, *Bam*HI; n, *Hae*II; o, *Hin*cII; p, *Bst*OI; q, *Nla*III; r, *Bfa*I; s, *Acc*I; t, *Mbo*II; u, *Mse*I; v, *Hph*I; z, *Mnl*I. Mutations in the HVS-I region are transitions unless the base change is specified explicitly. Boxes indicate novel information.

Kurds (KC) from Comas et al. (2000). From the Middle East/Arabian Peninsula, we included Iraqis (IQ), Syrians (SY), Yemenites (YM), Palestinians (PL), and Druze (DZ) from Richards et al. (2000); individuals from Dubai (DB) from A.T. (unpublished data); and Egyptians (EG) from Krings et al. (1999). From Pakistan/India, we included Pakistanis (PK) and Indians from Andhra Pradesh (AP), Gujarat (GK), Haryana (HY), Kashmir (KS), Maharashtra (MH), Punjab (PN), Rajasthan (RJ), Uttar Pradesh (UP), and Tamil Nadu (TN) from Kivisild et al. (1999a); and Indians (IN) from Mountain et al. (1995). From Central Asia, we included Kirghiz (KG), Uighur (UG), and Kazakh (KZ) samples from Comas et al. (1998). From western Eurasia, we included Basques (BS), Sicilians (SC), Bulgarians (BL), and Italians from Tuscany (TS) from Richards et al. (2000); Russians (RS) from Malyarchuk et al. (2002); Mansi (MN) from Derbeneva et al. (2002); and Sardinians (SD) from Di Rienzo and Wilson (1991). We also included Chinese (CH) from

Yao et al. (2002). The time to the most recent common ancestor of some clades and their SEs were calculated by means of the estimator ρ, the averaged distance to a specified founder haplotype, and were determined as described by Forster et al. (1996) and Saillard et al. (2000). Time estimates were also calculated, using the Network program. Principal-components (PC) analyses were performed using SPSS version 10.0.7 software, with basal mtDNA haplogroup frequencies as input vectors. Admixture proportions (mY) and their SEs were calculated, using information from all haplogroups, by means of the program Admix 2.0 (Dupanloup and Bertorelle 2001), on the basis of 1,000 bootstraps. The parental populations used for the analysis were Iranian populations and Gujarati for the Parsi population, and Pakistani populations (excluding the Makrani) and a geographically dispersed set of sub-Saharan African samples (Krings et al. 1999; Brakez et al. 2001; Brehm et al. 2002, Salas et al. 2002) for the Makrani population.