

Phylogenetic Star Contraction Applied to Asian and Papuan mtDNA Evolution

Peter Forster,* Antonio Torroni,† Colin Renfrew,* and Arne Röhl‡

*McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, England; †Dipartimento di Genetica e Microbiologia, Università de Pavia, Pavia, Italy; and ‡Mathematisches Seminar, Universität Hamburg, Hamburg, Germany

In the past decade, mitochondrial DNA (mtDNA) of 826 representative East Asians and Papuans has been typed by high-resolution (14-enzyme) restriction fragment length polymorphism (RFLP) analysis. Compared with mtDNA control region sequencing, RFLP typing of the complete human mitochondrial DNA generally yields a cleaner phylogeny, the nodes of which can be dated assuming a molecular clock. We present here a novel star contraction algorithm which rigorously identifies starlike nodes (clusters) diagnostic of prehistoric demographic expansions. Applied to the Asian and Papuan data, we date the out-of-Africa migration of the ancestral mtDNA types that founded all Eurasian (including Papuan) lineages at 54,000 years. While the proto-Papuan mtDNA continued expanding at this time along a southern route to Papua New Guinea, the proto-Eurasian mtDNA appears to have drifted genetically and does not show any comparable demographic expansion until 30,000 years ago. By this time, the East Asian, Indian, and European mtDNA pools seem to have separated from each other, as postulated by the weak Garden of Eden model. The east Asian expansion entered America about 25,000 years ago, but was then restricted on both sides of the Pacific to more southerly latitudes during the Last Glacial Maximum around 20,000 years ago, coinciding with a chronological gap in our expansion dates. Repopulation of northern Asian latitudes occurred after the Last Glacial Maximum, obscuring the ancestral Asian gene pool of Amerinds.

Introduction

Evolution of the Human mtDNA Molecule

As a nonrecombining locus (Ingman et al. 2000; Jorde and Bamshad 2000; Kivisild and Villems 2000; Kumar et al. 2000; Parsons and Irwin 2000), mitochondrial DNA can be analyzed in modern humans to yield a chronology of ancient genetic prehistory (Wilson et al. 1985). This feature has been successfully exploited to argue for a recent African origin of the human species (Vigilant et al. 1991; Penny et al. 1995). Soon after the appearance of anatomically modern humans about 130,000 years ago (Day and Stringer 1982; Bräuer 1989), an early expansion across Africa left its footsteps particularly in the mitochondria of the Bushmen and the west Pygmies (Watson et al. 1997). Subsequently, an east African reexpansion 60,000–80,000 years ago repopulated Africa and ultimately led to a migration out of Africa of at least one (Watson et al. 1997) or at least two major mtDNA types (Quintana-Murci et al. 1999), corresponding to the minimum number of emigrant women. Their descendants appear to have replaced all preexisting Eurasian *Homo erectus* or *Homo neanderthalensis* mtDNA types, given that no divergent mtDNA types have been found in any survey of modern humans (e.g., Torroni et al. 1994b; Richards et al. 1996, p. 196). According to the mtDNA sequences recovered from Neanderthal bones (Krings et al. 1997; Ovchinnikov et al. 2000), Neanderthal mtDNA diverged from the lineage leading to the modern human mtDNA lineage about 500,000 years ago, and Eurasian *H. erectus* necessarily diverged even earlier. Several human biomolecules (Ar-

mour et al. 1996; Tishkoff et al. 1996; Forster et al. 2000), but not all (Harding et al. 1997, 2000), are thought to have prehistories similar to that of mtDNA, underlining the need to study many loci before generalizing from molecules to a whole species.

With this overall picture of mitochondrial prehistory coming into focus, geneticists soon shifted their attention to the question of whether modern European, Asian, Papuan, and Australian mtDNA types derive from an uninterrupted demographic expansion of the out-of-Africa founders (strong Garden of Eden model), or whether an initial expansion was followed by the formation of regional gene pools, which, after a period of isolation and drift, expanded demographically and geographically to form the present mtDNA variation in different continents and regions (weak Garden of Eden model). One early methodology with which to explore these alternative scenarios was that of pairwise sequence difference distributions, or “mismatch distributions,” which were identified with global demographic expansions 80,000 to 40,000 years ago, thought to be in agreement with the weak Garden of Eden model (Harpending et al. 1993; Sherry et al. 1994). However, the mismatch distribution approach as used by Harpending et al. (1993) and Sherry et al. (1994) relied on implicit assumptions concerning the underlying mtDNA tree (Bandelt and Forster 1997). Nevertheless, independent phylogenetic studies on mtDNA restriction fragment length polymorphisms (RFLPs) in Europeans (Torroni et al. 1996), Asians (Ballinger et al. 1992a, 1992b; Torroni et al. 1993b, 1994c; Starikovskaya et al. 1998; Schurr et al. 1999), Papuans (Stoneking et al. 1990), Americans (Torroni et al. 1993a, 1994a), and Indians (Kivisild et al. 1999) implicitly confirmed the weak Garden of Eden model by finding distinct and phylogenetically deep mtDNA branches in each continent or region.

In this study, we present a chronology for the out-of-Africa migration and the onset of demographic ex-

Key words: phylogenetic network, Ice Age, refugium, Mongoloid, race.

Address for correspondence and reprints: Peter Forster, McDonald Institute for Archaeological Research, University of Cambridge, Cambridge CB2 3ER, United Kingdom. E-mail: pf223@cam.ac.uk.

Mol. Biol. Evol. 18(10):1864–1881. 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

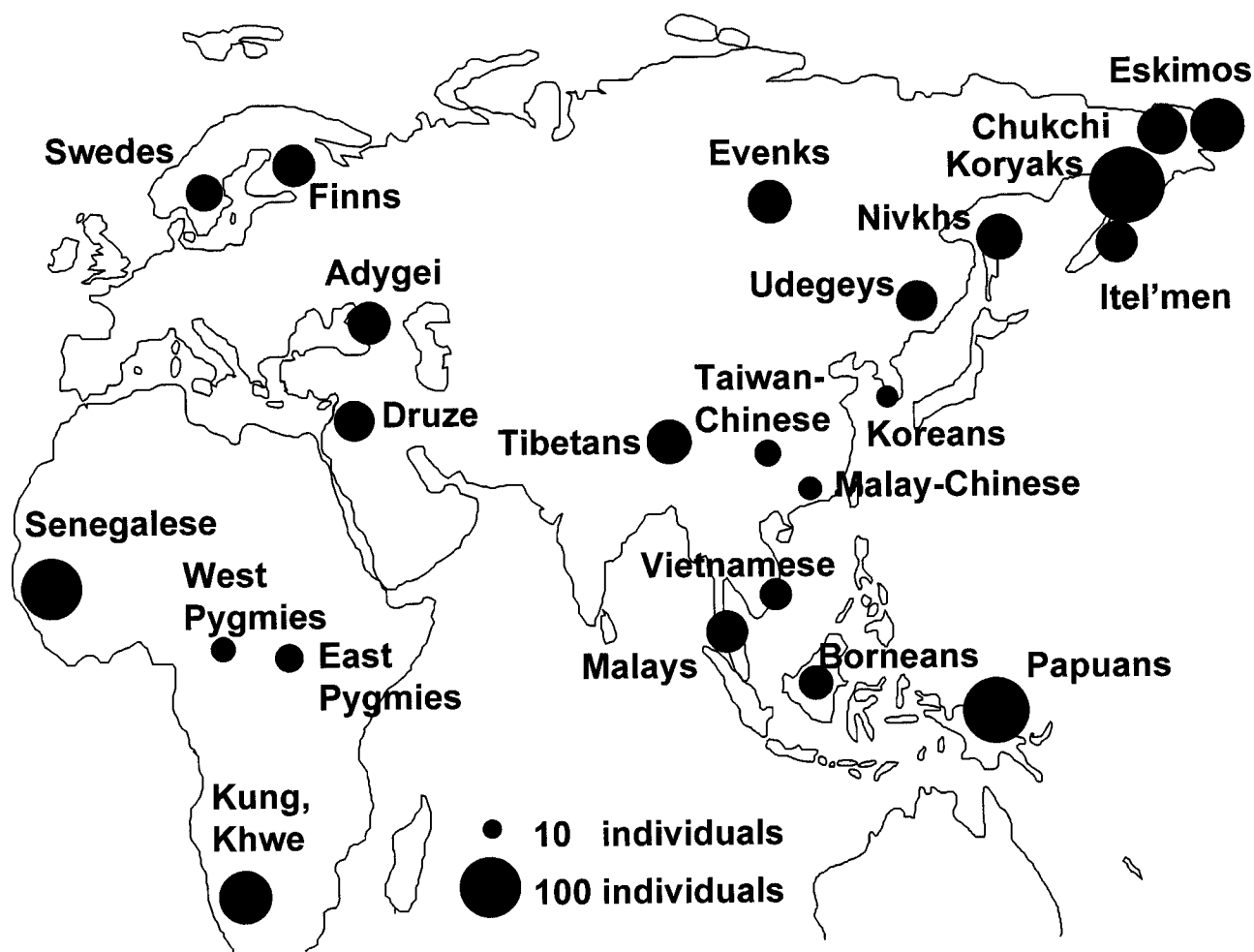


FIG. 1.—Samples typed by 14-enzyme restriction fragment length polymorphism analysis as of August 2000. Most of the 1,221 samples shown on the map were typed and published by the research groups of Douglas Wallace and Antonio Torroni, except for the Papuans, which were typed by Allan Wilson's group. The following samples are not displayed: the 381 Amerinds of Torroni et al. (1993a, 1994a), the worldwide sample by Cann, Stoneking, and Wilson (1987) as explained in *Materials and Methods*, and the preselected sample of 30 Indians published by Passarino et al. (1996).

pansions in Papua New Guinea and Asia. To achieve this aim, we first identified starlike mtDNA clusters diagnostic for demographic expansions by applying a new phylogenetic star contraction algorithm on high-resolution (14-enzyme) mtDNA RFLP data for 826 Asians and Papuans (fig. 1). The star contraction method identifies and distinguishes starlike phylogenetic clusters from nonstarlike branches according to a parameter specifying mutational time depth, one that is akin to cutting through a bush at a fixed height with a trimmer and then examining the diameters of the branches. An increased diameter, i.e., a greater molecule census, of any single cluster may arguably be the result of local circumstance; however, we observed that the clusters grouped together in geographically and temporally distinct sets, with each set therefore indicating a general demographic expansion. Genetic dating of these sets of clusters then allowed us to identify the relative and absolute times of different demographic expansions in Asia, which ultimately led to the peopling not only of the whole of Asia and Papua New Guinea, but also of America and Polynesia. To compare our RFLP-based results with pub-

lished mtDNA control region sequences, we took advantage of the hitherto unpublished correspondence table (appendix) linking the samples RFLP-typed by Cann, Stoneking, and Wilson (1987) and Stoneking et al. (1990) with the overlapping sample set sequenced for the mtDNA control region by Vigilant et al. (1991).

Alternative Data Contraction Methods

Two prerequisites for classifying ancestral mtDNA molecules into expansion classes are (1) a large mtDNA data set and (2) a rigorous procedure that classifies every node in the evolutionary tree, including nodes that are extinct and including expansion nodes within expansion clades. Previously, Oota et al. (1999) and Wang et al. (2000) had qualitatively classified and interpreted certain mtDNA control region sequences as radiation types, taking into account these factors. The phylogenetic star contraction algorithm we present here is designed to tackle both of these points for mtDNA RFLPs. The rather large Asian data set (826 individuals) is not amenable to conventional phylogenetic analysis and visualization methods, and an

analysis of subsets of the data (as previously published) would be of little help here, as expansion nodes will inevitably transgress arbitrary population and haplogroup subdivisions. The second point, i.e., characterization of clusters, equally requires our new star contraction algorithm rather than conventional data contraction methods. For example, the “frequency >1” option in the Network 3.0 program package does reduce mtDNA data as efficiently as star contraction (see *Results*), but at the cost of losing all singletons, without which time estimates would be skewed. More seriously, this option cannot classify extinct nodes as expansion nodes. The topiary pruning method (Wills 1995) is not an alternative due to its inherent weaknesses. In the first step, topiary pruning determines the consensus sequence of the data set and successively contracts the data in the direction of the consensus. A secondary aim of topiary pruning is to determine the root. However, inhomogeneous data sets consisting of widely different subsample sizes will considerably influence the consensus sequence and hence the rooting, such that the algorithm will produce a correct answer only in exceptional circumstances. Even with representative sampling, strong expansion events creating dominant sequence clusters (such as the major L2 and L3 clusters in Africa) will inevitably misguide the consensus rooting. Another serious problem in the algorithm is that obvious parallelisms are not identified as such, leading the contraction astray (Röhl 1999).

Molecules, Humans, and Populations

Two radically different approaches are currently popular in the endeavor to reconstruct human prehistory by means of molecular genetics (Pritchard and Feldman 1996; Risch, Kidd, and Tishkoff 1996; Stumpf and Goldstein 2001). There are those geneticists who set out from the DNA molecule as the basic unit of investigation (Forster et al. 1996; Torroni et al. 1998; Quintana-Murci et al. 1999; Richards et al. 2000), and there are those who set out from a (suitably defined) population as the basic unit (e.g., Relethford and Jorde 1999). Both approaches then attempt to reconstruct the prehistory of their respective units, delivering results which are not intended to be comparable (Harpending et al. 1998). For example, an increase in one type of molecule does not necessarily entail a net increase of humans in the “population.” While the molecular approach aspires to reconstruct the evolution of a genetic locus with the long-term aim of assembling many independent locus histories into a history of (human) evolution (Templeton 1998), the population approach attempts to take a short cut by implicitly or explicitly postulating a complete population model which is then tested on the basis of the data. To prevent confusion over terminology, we briefly summarize in this section some concepts of the molecular approach we adopt in this paper. In molecular mtDNA usage, an mtDNA type is said to have “geographically expanded” to a greater or lesser degree according to its observed greater or lesser geographic distribution and the distribution of its descendant types, i.e., the distribution of the clade (haplogroup). (If an mtDNA

clade is very widespread, e.g., the L2/L3 clade which is found in most of Africa and which is ancestral to more than 99% of non-African mtDNA types, then it goes without saying that the geographic expansion can only have been effected by concomitant “demographic expansion” of women carrying that mtDNA type.) While the geographic expansion of a molecule can be directly ascertained if no sweeping extinctions have occurred, a demographic expansion of a molecule is usually indirectly inferred from the more starlike or less starlike phylogenetic structure of its clade. The degree of “starlikeness” can be measured by the pairwise difference distribution within the clade (Watson et al. 1997) or by a “star index” (Slatkin 1996; Mateu et al. 1997; Torroni et al. 1998). Incidentally, in contrast to population terminology, the expression “constant-sized” has no application here: every observed molecule type has expanded from nonexistence to existence, whereas populations are usually defined as initially consisting of a number of individuals, which can increase, remain constant, or decrease until the time level of observation.

In this paper, we attempt a more direct approach to classifying molecules into demographically “more expanded” and “less expanded” types, namely, by taking a direct census of the number of molecules descended from a strongly expanded ancestral molecule. (Strongly expanded molecules, relative to other molecules, are first determined using our star density measure. This measure includes the feature that a descendent molecule which itself initiates a major expansion is classified separately as an expansion type.) Although taking a census of molecules to evaluate expansion may sound trivial, this approach is not equivalent to its theoretical counterpart of counting humans: an increase in the number of humans may be obscured by subsequent population declines in a quite unpredictable and unreconstructable manner. This is the strength of the molecular approach: population increase, decline, or mixing can systematically change neither the relative number of molecules descended from an ancestral molecule, nor their average mutational distance, nor therefore the time estimate of coalescence to that ancestral molecule or the starlike signature of expansion.

Materials and Methods

mtDNA Nomenclature

Throughout this paper, we refer to the nucleotide numbering system in the Cambridge reference sequence (CRS) of Anderson et al. (1981). The 11 errors recently found in the CRS by Andrews et al. (1999) did not seem to affect any of the restriction sites in the data we used below. When referring to branches in the mtDNA tree, we employ the mtDNA phylogenetic nomenclature updated by Macaulay et al. (1999) and T. Kivisild et al. (personal communication).

Data Revision

Australians and Asians of Cann, Stoneking, and Wilson (1987)

The Cann, Stoneking, and Wilson (1987) worldwide data consist of 46 Caucasians (Caucasoids), 34

Table 1
Enzyme Batteries Employed in the Seven Published Asian High-Resolution mtDNA RFLP Analyses

Enzyme	Recognition Site ^a	Cann (1987) ^b	Stoneking (1990) ^b	Ballinger (1992) ^b	Torrioni (1993/4) ^b	Starikov- skaya (1998) ^b	Schurr (1999) ^b	Macaulay (1999) ^b
<i>AluI</i>	AG.CT	a	a	a	a	a	a	a
<i>AvaII</i>	G.G(A/T)CC	b	b	b	b	b	b	b
<i>DdeI</i>	C.TNAG	c	c	c	c	c	c	c
<i>FnuDII</i>	CG.CG	d	d	—	—	—	—	—
<i>HaeIII</i>	GG.CC	e	e	e	e	e	e	e
<i>HhaI</i>	GC.GC	f	f	f	f	f	f	f
<i>HinfI</i>	G.ANTC	g	g	g	g	g	g	g
<i>HpaI</i>	GTT.AAC	h	—	h	h	h	h	h
<i>HpaII</i>	C.CGG	i	i	i	i	i	i	—
<i>MboI</i>GATC	j	j	j	j	j	j	j
<i>RsaI</i>	GTAC	k	k	k	k	k	k	k
<i>TaqI</i>	T.CGA	l	l	l	l	l	l	l
<i>BamHI</i>	G.GATCC	—	—	m	m	m	m	m
<i>HaeII</i>	(A/G)GCGC.(T/C)	—	—	n	n	n	n	n
<i>HincII</i>	GT(T/C).(A/G)AC	—	h	o	o	o	o	o
<i>PstI</i>	CTGCA.G	—	—	p	—	—	—	—
<i>PvuII</i>	CAG.CTG	—	—	q	—	—	—	—
<i>XbaI</i>	T.CTAGA	—	—	r	—	—	—	—
<i>XhoI</i>	C.TCGAG	—	—	s	—	—	—	—
<i>MspI</i>	C.CGG	—	—	—	—	—	—	i
<i>NlaIII</i>	CATG.	—	—	—	—	—	—	q
<i>BfaI</i>	C.TAG	—	—	—	—	—	—	r
<i>AccI</i>	GT.(A/C)(G/T)AC	—	—	—	—	—	—	s
<i>BstOI</i>	CC.(A/T)GG	—	—	—	—	—	—	t
<i>MseI</i>	T.TAA	—	—	—	—	—	—	u
9-bp deletion	nps 8272–8289	9 bp	9 bp	9 bp	9 bp	—	—	—

^a Dots indicate cleavage positions.

^b See *Materials and Methods*.

Asians, 26 Papua New Guineans, 21 Australians, 18 African Americans, and 2 African-born Africans. The Papua New Guinean individuals are a subset of the Stoneking et al. (1990) Papuans (see below). We eliminated four sites, as they could not be experimentally confirmed (see table 1 for the letter codes used to specify the enzymes): “8j,” “1484e,” “7750c,” and “13031g” (M. Stoneking, personal communication). Furthermore, “1403a” was corrected to 10397a, as pointed out by Ballinger et al. (1992a). We added site +1851 in Cann types 8 and 9 in accordance with the corresponding Vigilant sequences (appendix). The publication by Cann, Stoneking, and Wilson (1987) does not specify length variability of the 9-bp duplication at nucleotide positions (nps) 8272 through 8289, but this information is available: individuals 44, 62, 71, 73, 74 (Asians), and 72 (African American) have a 9-bp deletion (Wrishnik et al. 1987). Even with these corrections, the Cann data set was the only RFLP set unable to produce a low-dimensional phylogenetic network (not shown), indicating undetected data problems (Bandelt et al. 1995). For this reason, we discarded the Cann data from further analyses.

Papuans of Stoneking et al. (1990)

All 119 Papuans of Stoneking et al. (1990) were taken for the present study. The “1403a” site was corrected to 10397a. All but two of the Cann, Stoneking, and Wilson (1987) samples, including all Papuans, had been retyped with *RsaI* (enzyme k) for the Stoneking et al. (1990) study, explaining the corrections in two Pap-

uans (Papuan 24, alias Cann 12, correctly has +16310k; Papuan 106, alias Cann 134, correctly has +16310k). We corrected “16178l” to 16143l according to the sequences of Vigilant et al. (1991). Papuan 116 is published as having –16310k, contradicting the corresponding Vigilant (1990) sequence SH17 (appendix), which has 16311T.

Southeast Asians of Ballinger et al. (1992a, 1992b)

All 153 southeast Asians (14 Malaysian Chinese, 14 Malays, 32 Malay Aborigines “Orang Asli,” 32 Sabah Borneo Aborigines, 20 Taiwanese Han, 28 Vietnamese, and 13 South Koreans) were incorporated into the present study. We transferred site –4711i from mtDNA type 68 to mtDNA type 76, as explained in the corrigendum by Ballinger et al. (1992b). The –4685a variant was transferred from type 69 to type 77 in our file (A. Torrioni, personal communication). The site 6618e in Ballinger et al.’s (1992a) appendix B is inconsistent with clade G in the parsimony tree in their figure 2; original documentation, however, is not available. Sites “13284a” and “13284e” are artifacts (T. Schurr, personal communication) and were deleted here. Site “+4735k” in KN99 was changed to 4732k to agree with Tib142 (Torrioni et al. 1994c) and Ev20 (Torrioni et al. 1993b). Site “15431e” in KN102 was renamed 15437e to agree with RFLP type SIB40 of Schurr et al. (1999) and Starikovskaya et al. (1998): KN102 and SIB40 (found in one Chukchi, four Siberian Eskimos, and two Koryaks) are mtDNA clade D and, moreover,

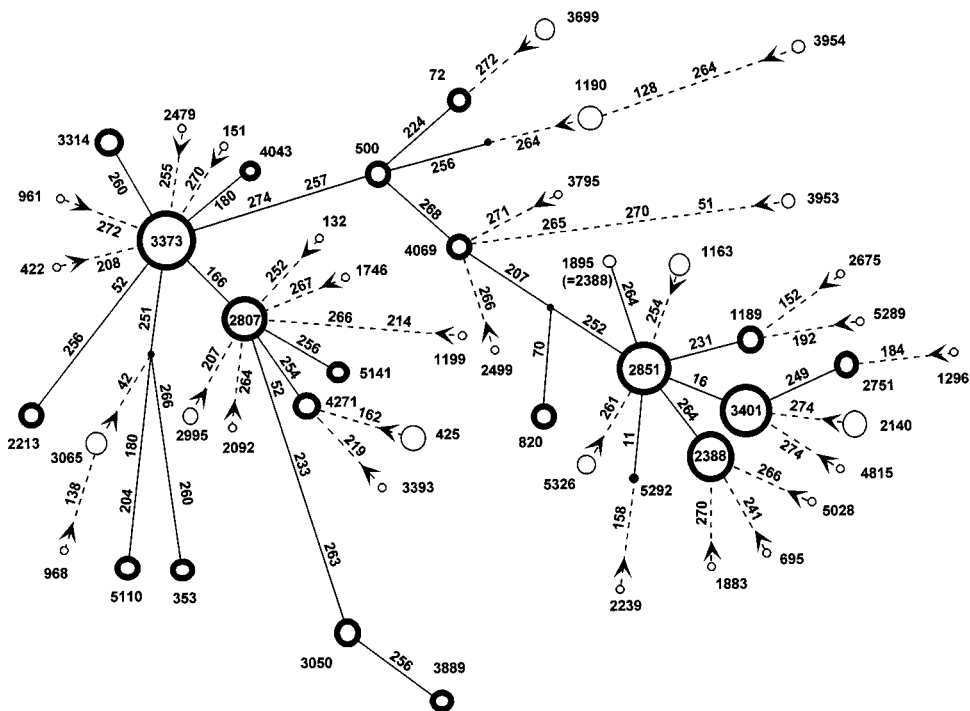


FIG. 2.—True tree of 200 randomly chosen molecules of the simulated data set. The numbers on the nodes refer to molecule designations. The numbers on the links refer to mutated nucleotide positions. The circle sizes are proportional to the numbers of molecules they represent. Black nodes represent reconstructed median vectors. The thick circles correspond to sequence types which are retained after the first round of the star contraction algorithm. Arrows point to sequence types to which peripheral sequence types are assigned by the star contraction algorithm. Sequence types 500 and 3373 are the two surviving original types with which the simulation was initiated.

share the absence of site 101801. The ambiguous sequence region according to Anderson et al. (1981) is as follows: 15430-C GCC CTC GGC TAC-15442. The corresponding amino acid sequence is ALGL. Site 15431e implies a change from A to G, and site 15437e implies a change L to H. The following sites were corrected mainly by consulting laboratory notes: “1063e” is 1062e; “3569j” is 3659j; “7672g” is 7672j; “8270k” is probably 8269k; the purported single site change “8569c” in type 110 is a double site, +8569c/–8572e; “9329f” is 9327f; “10256j” is 10256r; “11557b” is 11577b; “15660c” is 15460c; and “+16096g” is –16096k. The following sites are incorrect or misplaced, but documentation is not available: “160f,” “3659o,” “6534e” (presumably an “abbreviation” of 16534e: both “6534e” and 16534e are recorded for VN47), “9386e,” “13180j,” “15595e,” and “16512l.” These incorrect or misplaced sites were retained, as they do not create phylogenetic conflicts.

Siberians of Torroni et al. (1993b)

All 153 Siberians (57 Nivkhs, 51 Evenks, and 45 Udegys) were used in the present study. The sign of site 101801 is inverted; i.e., only Siberian 15 in fact has a loss at this site.

Tibetans of Torroni et al. (1994c)

The complete set of 54 Tibetans was used in the present study. Tibetan 118 was corrected as having +10394c and +10397a (Bandelt, Forster, and Röhl

1999). We corrected the following positions: “5259b” is 5260b, “6331b” is 6332b, “14773c” is 14774c, and “16388e” is 16398e.

Chukchi and Siberian Eskimos of Starikovskaya et al. (1998)

The complete data set of 145 individuals (79 Siberian Eskimos and 66 Chukchi) was used in this study. We note that variation at nucleotide position (np) 16311 was not detected by 16310k.

Kamchatkans of Schurr et al. (1999)

The complete data set of 202 Kamchatkans (56 Aulitor Koryaks, 44 Karagin Koryaks, 55 Palan Koryaks, and 47 Ite'l'men) was used in this study. We note that variation at np 16311 was not detected by 16310k.

Data Harmonization

The five published RFLP studies use enzyme sets which overlap but are not quite congruent (see table 1), necessitating harmonization. In the harmonized data we used for the analysis, the additional enzymes p, q, r, and s employed only by Ballinger et al. (1992a) were deleted. These enzymes recognize longer recognition sites and therefore cut infrequently: in the diverse Ballinger et al. (1992a) data, p and s never cut, q cuts once but can be expressed as o, and r cuts in only three mtDNA types. Furthermore, due to overlapping recognition sequences, enzymes m and n used in the Ballinger et al.

(1992a) and Torroni et al. (1993b, 1994c) studies can always be expressed by j and f, respectively, in the studies used here. Enzyme h could usually be expressed as enzyme o in the four studies in which both were used (table 1). The only exceptions were in the Ballinger et al. (1992a) data, for which we retained the independent 12406h variant of Asian 19, in the Schurr et al. (1999) data, for which we retained 1004h in Koryak 66, and in the Cann, Stoneking, and Wilson (1987) data, for which we retained 3592h. We omitted enzyme *FnuDII*, employed only by Cann, Stoneking, and Wilson (1987) and Stoneking et al. (1990); this enzyme detects variation in only one of the 119 Papuan individuals. We expressed sites +12345k and 64i by the alternative sites +12528k and 16494i, respectively, to harmonize the Cann, Stoneking, and Wilson (1987)/Stoneking et al. (1990) data with the Ballinger et al. (1992a) data. We consistently scored variation at 16517e as gains (some publications scored it as losses). The harmonization hence loses a minimum of information with respect to the standard 14-enzyme system, permitting direct comparisons with the RFLP mutation rate estimates based on European and Amerind mtDNA (Torroni et al. 1998, and with the African RFLP data of Chen et al. (1995). A file (asiapng.tor) of the revised and harmonized Asian data (excluding the Asians and Australians of Cann, Stoneking, and Wilson [1987]) is available at <http://www.fluxus-engineering.com>.

Star Contraction Algorithm

The aim of the star contraction algorithm is to identify starlike clusters of sequences in a given sequence set and contract these clusters to single representative sequences. The resulting reduced sequence set can be entered into a phylogenetic algorithm to generate a tree or network. There are two potential applications for this method. First, large population data sets (several hundred to over a thousand) are rapidly becoming the norm in population genetic studies, and it is becoming increasingly difficult to display the corresponding phylogeny as a figure in a publication or even to visually analyze it on a computer screen. The star contraction algorithm in conjunction with a phylogenetic analysis can display the much smaller “skeleton” of the tree, with the clusters indicated as single nodes. The second application of the star contraction method is to rigorously define dense starlike clusters which are potentially diagnostic for demographic expansions. The time to coalescence of such phylogenetic clusters can then be dated via the molecular clock and compared with historic or prehistoric records of other disciplines.

The following algorithm specifies a certain time depth or mutational distance radius δ (range 0 to ∞) up to which clusters are to be identified, with the founding sequence of a cluster being reconstructed if it is absent in the original data (step 2). Any node ancestral to a cluster is excluded from contraction into the cluster (step 9). Potential clusters are evaluated on the basis of a star density measure (step 3). Sequences which are equidistant to two or more potential stars are preferentially as-

signed to existing sequences rather than to inferred nodes (steps 4–6). The following algorithm can be executed several times in succession (step 11) to progressively contract the network, but the conditions in step 3 ensure that successive contraction rounds converge on a skeleton phylogeny rather than on a single node.

- Step 1: Pooling of identical sequences.
- Step 2: Generation of all median vectors as hypothesized ancestral nodes from triplets (U, V, W) which fulfill the following conditions, expressed in Hamming distances:
 - a) $d(U, V) = 2$
 - b) $d(U, W) \leq \delta + 1$
 - c) $d(V, W) \leq \delta + 1$.
- Step 3: Calculation of the clique (“clade” in biological parlance) for every sequence of the data set and for every median vector generated in step 2. Stored in the memory are those cliques which fulfill the following conditions:
 - a) $\#C_1(U) \geq 2$
 - b) $\#C_1(U) + 0.5 \#C_2(U) + 0.25 \#C_3(U) \geq 3.5$,
 where $\#C_k(U)$ is the number of sequences (i.e., individuals) within the clique $C(U)$ at distance radius k . The second subcondition enforces an exponentially declining contribution by increasingly distant sequences to the star center, with the empirical threshold 3.5 found to be discriminatory down to the time depth of the human mtDNA coalescent (Röhl 1999). A further condition in this step is that the sequences contracted into a clique center A (in previous rounds of the algorithm) are contracted to their ancestral clique B only if less than half as many sequences are assigned to A compared with B. A subcondition for noncontraction of A is that at least one sequence must have been assigned to A.
- Step 4: Elimination of all multiply-assigned sequences out of those cliques in which the assigned sequence is at a greater distance radius than in the clique with the smallest distance radius.

Exception: sequences are not removed if the clique at the smallest distance radius has, in contrast to the other clique, a median vector as a clique center.
- Step 5: Elimination of all stored cliques without any assigned sequences or with only one assigned sequence and a median vector as a clique center.
- Step 6: Repetition of step 4 but without the exception.
- Step 7: Elimination of all multiply-assigned sequences.
- Step 8: Calculation of the minimum spanning network (Bandelt, Forster, and Röhl 1999) of the non-assigned sequences with the clique centers added.
- Step 9: Elimination of all sequences which are candidates for links lying on or branching off from the minimum spanning network.

- Step 10: Addition of median vectors to the data set if there are cliques left with a median vector in the center. In the following, they will be treated like originally sampled sequences. Pooling of the sequences assigned to one clique.
- Step 11: Repetition from step 2 onward until no further reduction occurs or until the algorithm is terminated externally.
- Step 12: Reduction of the sequence length to those characters that mutate at least once in the contracted data set.

Phylogenetic and Demographic Analysis

Prior to the star contraction analysis, we first deleted three RFLP sites from the data set: sites 16303k and 16310k, which were difficult to detect in some studies for technical reasons, and site 16517e due to its erratic, possibly directional, mutation mechanism (Chen et al. 1995; Forster et al. 1997). For the star contraction analysis, all the RFLP data sets described above except for the Cann, Stoneking, and Wilson (1987) data were entered in the star contraction option of Network 3.0. Three rounds of star contraction were run, with δ set at 3 mutations, corresponding to the coalescence time of the two African mtDNA founders for modern Eurasians and Papuans (see *Results*). The star contraction output file was then entered into the reduced median network (RM) option (Bandelt et al. 1995) using the default settings except for the weighting of three sites as follows: sites 10394c/10397a were weighted half to counter known recurrent double-site losses through a single point mutation (Bandelt, Forster, and Röhl 1999), and site 1715c was weighted half due to its known hyper-variability (Macaulay et al. 1999). Finally, the RM output file was entered into the median-joining (MJ) network option (Bandelt, Forster, and Röhl 1999), using default parameters and the same weighting system, to eliminate nonparsimonious links (Forster et al. 2000). In this final network, any node containing more than 1% of the total number of individuals (i.e., at least 9 given the sample size of 826) was defined as a demographic expansion cluster. This threshold was chosen for the practical reason that statistical analysis of samples of size <9 does not appear sensible. In other words, for the detection of smaller demographic expansions, much larger sample sizes would be necessary. (Note that our procedure only distinguishes dense from less dense phylogenetic nodes and stars, indicating greater or lesser demographic expansion. We do not seek to draw an artificial line between “stationary” and “expanding” lineages.) The resulting expansion clusters were then analyzed for geographic distribution and for expansion age using the RFLP mutation rate.

RFLP Mutation Rate and Genetic Dating

The mutation rate of the 14-enzyme RFLP typing system for the whole mtDNA molecule has previously been calibrated (Torrioni et al. 1998) using Amerindian samples which had been both sequenced for the first hypervariable region (np16090–np16365) of the mtDNA

control region and RFLP-typed for sites throughout the molecule. A comparison of the control region tree with the RFLP tree yielded a ratio of 1.21 control region mutations for 1 RFLP mutation. Macaulay et al. (1999) performed a similar comparison in a European sample but obtained a ratio of only 0.82 control region mutations for 1 RFLP mutation. The apparently large discrepancy was only partly mitigated when the authors confined their comparison to shallow clusters in their trees (to avoid overlooking saturated recurrent mutations). A major reason for the discrepancy was that 10 of the RFLP mutations had in fact been detected by control region sequencing, namely, seven 16310k mutations, two 16303k mutations, and one 16208k mutation (V. Macaulay, personal communication), and one RFLP mutation (12308g) had been detected by a mismatched primer. When considering only the shallow clusters in their phylogeny (H, pre-HV, J, T, K, U1, U5, R1, and X) this still leaves seven RFLP mutations which would not have been detected by conventional RFLP typing in most of the RFLP publications used here. After subtracting these mutations, we obtained a control region/RFLP mutation ratio for their tree of 0.9500. For the purpose of this paper, we adopted the average ratio between 0.95 and 1.21, namely 1.08, which translates into 1 RFLP mutation every 21,800 years using the control region calibration of Forster et al. (1996) and Saillard et al. (2000). This point estimate is similar to the calibration of Horai et al. (1995) using primate mtDNA. As a time measure for dating nodes in the phylogeny, we use the demographically unbiased parameter ρ , which is the average mutational distance to the node of interest (Morral et al. 1994; Forster et al. 1996). For estimating the standard error σ of ρ , we employed the method of Saillard et al. (2000). The values for ρ and σ are converted into years by multiplication with the mutation rate. The standard error σ does not include uncertainty in the mutation rate. However, any future improved calibration for the mutation rate can directly be multiplied with the ρ and σ values presented throughout this paper to obtain improved absolute time estimates. Thus, for example, if the out-of-Africa migration date were doubled from 55,000 years to 110,000 years (e.g., to accommodate the Skhul/Qafzeh remains as our ancestors), then the mtDNA date for the migration into the Americas would correspondingly increase from 25,000 to 50,000 years.

Results

Star Contraction Applied to Simulated mtDNA Control Region Evolution

Conventional methods for verifying data models used in the phylogenetic analysis of molecular sequences are largely based on coalescence theory (cf. Wakeley and Hey 1997), in which the point of departure is a sample of current molecules whose ancestry is traced back in time to the most recent common ancestral molecule. The more realistic approach of simulating molecular evolution forward in time enables the geneticist to better study founder effects and bottlenecks with the

concomitant loss of sequences, as well as the effect on the phylogeny of multiple hits at the same nucleotide positions. For the following forward simulation, we assumed a founding set of 500 mtDNA molecules (four sequence types with a frequency of 125 each) whose number undergoes various expansions, contractions, and phases of stagnation in a period of 60,000 years, with an average generation time of 25 years. The simulation assumed that a woman could have maximally 3 daughters who would survive to reproductive age. In total, the number of molecules was permitted to rise to 5,000. In addition, we chose the mutation rate of Forster et al. (1996), which posits one transition per 20,180 years between nps 16090 and 16365 in the mtDNA control region. To begin with, we used the infinite-sites model, which requires that mutational events never hit the same nucleotide position twice; i.e., the number of nucleotide positions is unlimited. After the simulation, every mutation was randomly assigned to one of 275 nucleotide positions, with the positions being grouped into three mutation rate classes. The sizes of the three groups and their mutation rates were defined with reference to Aris-Brosou and Excoffier's (1996) gamma distribution with $\alpha = 0.4$ for the mutation probability of the positions. The borders for the maximum probabilities within the three groups were chosen as follows: conservative positions = probability 0 to 0.0015; average positions = probability >0.0015 to 0.01; hypervariable positions = probability >0.01 . Average values for each of the three groups were determined from these probabilities. Since the gamma distribution of Aris-Brosou and Excoffier (1996) refers not to 275 nucleotide positions but to 300, we multiplied the ranges by 11/12 and rounded to 5 positions. Transition probabilities were as follows: for nps 1–150, (conservative), 1/1,500; for nps 151–250 (average), 3/1,000; for nps 251–275 (hypervariable) 3/125.

At this point, we compressed the infinite-sites model into the finite-sites model by assigning several independent mutation events to a single nucleotide position (parallel mutations and reversions); nevertheless, the historical order of mutations was stored in the computer, allowing later comparisons of the real molecular tree with the reconstructed phylogenies.

The simulation spanned a period of 60,000 years (2,400 generations) with a nearly stagnant census punctuated by the following phases of increase and decline: expansion to 1,500 molecules—start at generation 20, end at generation 130; expansion to 2,500 molecules—start at generation 400, end at generation 450; contraction to 1,000 molecules—start at generation 700, end at generation 750; expansion to 5,000 molecules—start at generation 850, end at generation 950; contraction to 3,500 molecules—start at generation 1200, end at generation 1220; contraction to 3,000 molecules—start at generation 1600, end at generation 1605; expansion to 5,000 molecules—start at generation 2000, end at generation 2030.

The expansion and contraction phases were simulated according to a sinelike function in the interval $[-\pi/2, \pi/2]$. Within the specified restraints, the simulation allowed a maximum random change of 2% of the overall number of molecules from one generation to the

next. In other words, minor expansions and contractions were permitted even during the stagnation phases. For example, in the simulation described here, a minor expansion occurred in the last 400 generations, yielding a molecule census of 5,371 at the end of the simulation. In the course of the simulation, two of the original four founding sequence types went extinct, the first one after only 5,000 years (200 generations), and the second after about 20,000 years (800 generations). Finally, from the resulting data set, 200 sequences were randomly sampled and submitted to the star contraction algorithm.

Figure 2 displays the true tree and within it the star-contracted tree of the random sample. It should be pointed out that the expectation of exactly reconstructing the true tree (fig. 2) is not justified. For one thing, parallel mutations may make certain sequence types look identical, as in the case of sequence types 1895 and 2388 and sequence types 2140 and 4815. Moreover, a reversal has occurred at nucleotide position 264 on the link between type 1190 and type 3954 which no parsimony method would identify. In the first round, the star contraction algorithm reduced this data set from 50 sequence types with 43 variable positions to 23 sequence types with 23 variable positions, hence a reduction by 75.4% (number of sequence types multiplied by number of variable positions). In this first round, no type was incorrectly assigned, and two median vectors were reconstructed. In the second star contraction round, the data set was reduced by another 11 sequence types and 8 variable positions, and the third round caused another reduction by one sequence type and one position (not shown).

To summarize, the SC algorithm in this simulation was found to assign, without any errors, numerous sequence types to ancestral nodes, in spite of the high level of homoplasy. This not only contracts the data set by over 70% after a single round of star contraction, it also accurately identifies and contracts numerous parallel and reverse mutations, thus simplifying subsequent phylogenetic analyses, e.g., by applying a tree-building or a network algorithm to the star-contracted data set.

Star Contraction Applied to Real mtDNA Coding-Region RFLP Evolution

Threefold application of the star contraction algorithm to the 826 east Asian and Papuan RFLP sequences (encompassing 247 types) contracted the number of types to 115 after the first round, 85 after the second round, and 83 after the third round. The reduction efficiency was thus similar to the frequency >1 option of Network 2.0 (654 sequences occurred more than once, yielding 75 different types, which yielded a structurally similar RM network not shown here). The reduced data set was weighted as described in the *Materials and Methods* and submitted to a reduced median network analysis which yielded a three-dimensional phylogenetic network containing both parsimonious and nonparsimonious links (not shown). To eliminate nonparsimonious links, the RM output was submitted to an MJ network analysis, and the resulting simpler network is shown in figure 3. Inspection of the Asian/Papuan net-

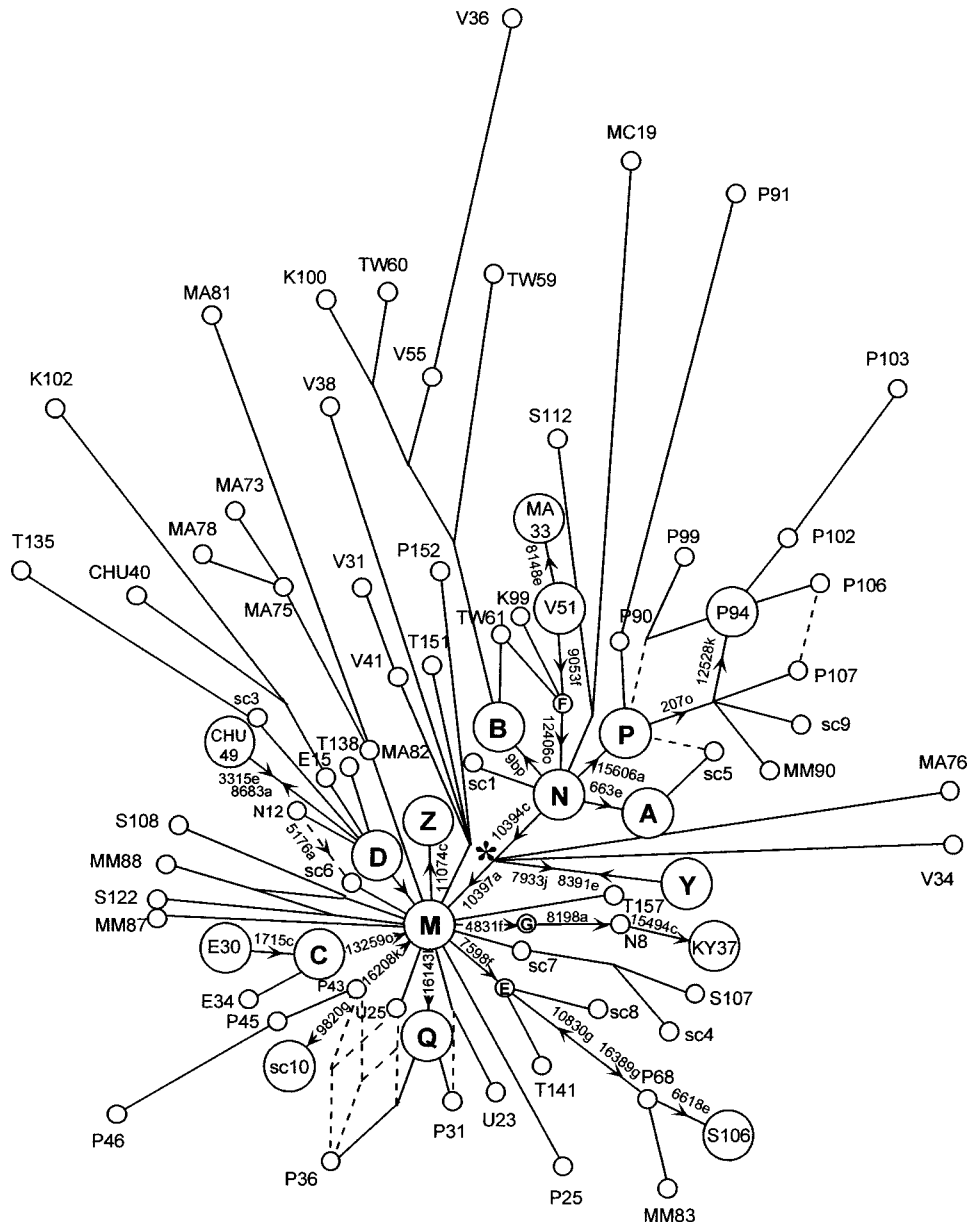


FIG. 3.—Star-contracted phylogenetic network of Asian mtDNA restriction fragment length polymorphism types. The larger circles indicate clusters, defined as comprising $>1\%$ (≥ 9 individuals) of the total sample. The smaller circles represent single, multiple, or star-contracted sequence types comprising $<1\%$ of the sample. For space reasons, only those mutations distinguishing clusters are entered along the links, with arrows pointing to restriction-site gains. Dotted links indicate reticulations which were resolved manually by consulting synapomorphies in available mtDNA control-region sequences. Most node names are labeled as in the original publications (V—Vietnamese; T—Tibetan; P—Papuan; MC—Malay Chinese; TW—Taiwanese Han; K—Korean; KY—Koryak; CHU—Chukchi; MM—Malay Malaysian; MA—Malaysian aboriginal; S—Sabah Bornean aboriginal; U—Udegey; E—Evenk; N—Nivkh). Star contraction clusters with reconstructed ancestral nodes are labeled “sc.” Some nodes (indicated in bold and without numbers) are renamed according to standard mtDNA clade nomenclature A, B, C, etc. (Macauley et al. 1999; Quintana-Murci et al. 1999), except for P and Q, which are new labels proposed here. Note that any type derived from, e.g., a C node is itself within clade C. The asterisk denotes the African root.

work reveals that most Papuans fall into branches which are distinct from those for mainland Asians, except for mtDNA clades B, E, and F, which were found in coastal Papua New Guinea, and one Malay Malaysian (MM90) who was potentially maternally related to one of the highland Papuan clusters. To investigate the Papuan sample, we therefore constructed a phylogenetic network (fig. 4) of the Papuans including the Malay. The Papuan network was constructed by combining the RM

with the MJ algorithms as for the network of figure 3. We dispensed with the star contraction algorithm for the Papua New Guinea network because the expansion clusters were already defined in the overall network and we here wished to focus on individual correspondences between control-region and RFLP types (see *Discussion*). Intraspecific robustness tests such as bootstrapping are not available for shallow intraspecific phylogenies (Bandelt et al. 1995). However, confidence can be gained

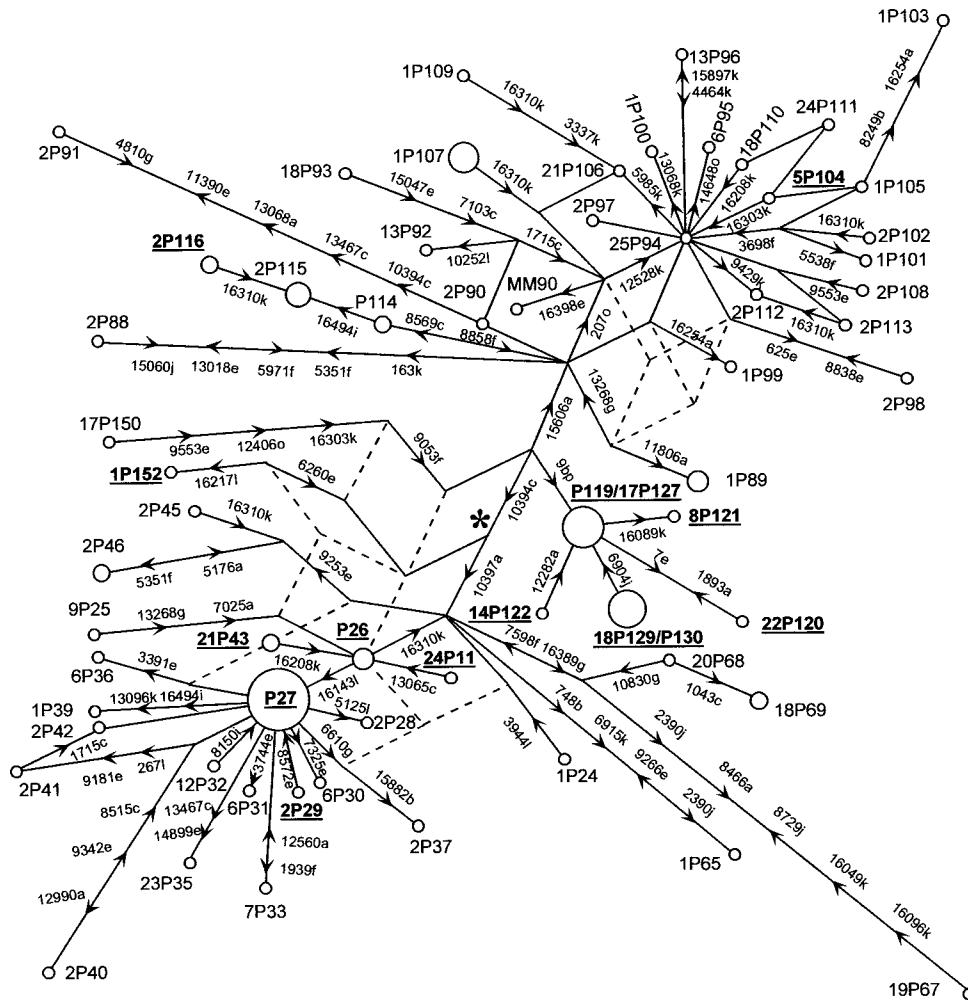


FIG. 4.—Phylogenetic network of Papuan mtDNA restriction fragment length polymorphism (RFLP) types. One Malay Malaysian aboriginal (MM90) is included. Unlike in figure 3, circle areas correspond to the numbers of individuals, and variation at 16303k and 16310k is included. Symbols and abbreviations are as in figure 3, except that the numbers preceding the Papuan labels indicate geographic origin: 1–3 indicate highland Papuans, and higher numbers indicate coastal Papuans as in Stoneking et al. (1990). Clade Q corresponds to clade II in Sykes et al. (1995). RFLP types which have been sequenced for the mtDNA control region (appendix) are underlined. The asterisk denotes the African root.

from the fact that most links in the Asian part of the phylogeny correspond to the published phylogenies of the individual data sets (for a reanalysis of the Tibetan data, see Bandelt, Forster, and Röhl [1999]) if variation at 16517e is disregarded. A separate tree of the Papuans has not previously been published, but published trees combining the Papuans with others are available (Stoneking et al. 1990; Ballinger et al. 1992a). In those analyses, most of the Papuans are grouped into the same two highland clusters and into the coastal B cluster as in our analysis, and the deep mutations within these clusters are the same. Only the tips are slightly different, but this is mainly due to 16310k and 16517e, which we weighted half and zero, respectively, unlike in the published analyses.

Next, we defined those nodes in the Asian-Papuan phylogeny (fig. 3) which contained more than 1% of the sample (i.e., a minimum of 9 individuals given a sample size of 826) as demographic expansion clusters. Typically, such a node consists of an existing sequence type with derived sequence types contracted into it by the star con-

traction algorithm; however, some nodes are ancestral nodes postulated by the star contraction algorithm, in which case they are designated by “sc” if descendant types have been contracted into them. The expansion nodes are listed in table 2, along with their age estimates. We decided to differentiate the diversity values for two of the expansion nodes in the network (A and B) into two values each because their diversities were dominated by a single population in both cases; thus, for A, we calculated the values for Beringians, i.e., Eskimos, Chukchi, and Kamchatkans (entered as A2), separately from the values for the other Asians (A1); this division was justified because Beringian A2 was mutated at np 16111 and was thus more closely related to American A2 than to Asian A1. For B, we calculated the values for coastal Papua New Guineans (B/png) separately from B for the other Asians (B/asia). This division was also justified by a consideration of the corresponding control-region sequences (appendix), as all of the Papuan B types had the Polynesian sequence motif, known to have expanded recently (Melton et al. 1995; Sykes et al. 1995; Richards,

Table 2
Time Estimates for mtDNA Expansion Clusters

Cluster ^a	Distribution	<i>n</i>	Rho	Sigma	Rho (years)	Sigma (years)
P	Papua New Guinea	13	3.0000	1.0686	65,400	23,300
M	Southern and central Asia	59	1.5932	0.2570	34,700	5,600
F/v51	Southern and central Asia	16	1.5625	0.4881	34,100	10,600
P/p94	Papua New Guinea	19	1.5263	0.3722	33,300	8,100
N	Asia	33	1.3636	0.4525	29,700	9,900
B/asia	Southern and central Asia	18	1.3333	0.3239	29,100	7,100
A1	Southern and central Asia	12	1.2500	0.3891	27,300	8,500
D	Northern and central Asia	26	0.7692	0.5189	16,800	11,300
Q	Papua New Guinea	37	0.7027	0.1622	15,300	3,500
D/chu49	Northern Asia	19	0.6842	0.6338	14,900	13,800
B/png	coastal Papua New Guinea	21	0.5714	0.3927	12,500	8,600
E/s106	Southeastern Asia	9	0.5556	0.3685	12,100	8,000
A2	Northern Asia	117	0.5128	0.2961	11,200	6,500
C	Northern and central Asia	110	0.4955	0.2303	10,800	5,000
F/ma33	Southeastern Asia	12	0.4167	0.2205	9,100	4,800
Z	Northern Asia	9	0.2222	0.1571	4,800	3,400
Y	Northern Asia	59	0.2203	0.1137	4,800	2,500
G/ky37	Northern Asia	102	0.2157	0.1009	4,700	2,200
M/sc10	Northern Asia	9	0.1250	0.1111	2,700	2,400
C/e30	Northern Asia	16	0	0	0	0
Total		716				

^a Defined as in figure 3; not to be confused with entire clades of the same name.

Oppenheimer, and Sykes 1998). A few cluster descendants were not contracted by the algorithm because they were involved in a network reticulation; we therefore performed some modest postprocessing of the clusters by resolving reticulations on the basis of control-region information where available (these resolved reticulations are indicated as dotted lines in fig. 3), or in the case of the Papuan clusters involved in reticulations, by referring to the separate Papuan network (fig. 4). The age estimates in table 2 are sorted by decreasing age and range from 65,400 ± 23,300 years for one of the Papuan clusters to ages of less than 10,000 years for several north Asian clusters. Out of the total number of 826 individuals, 716 (87%) are assigned to one of the 20 expansion nodes in table 2. The current geographic distributions of the expansion clusters are shown in figures 5 and 6.

All 826 Asian and Papuan sequences (as, in fact, most non-African sequences) belong to the African sub-cluster L3, defined by the lack of site 3592h (Watson et al. 1997). Consequently, the root of the Asian-Papuan network is the node separating the 10394c/10397a links between M and N and is indicated by asterisks in figures 3 and 4. The root node has two major derivatives—the M and N clusters—and, at first glance, three minor derivatives—MA76, V34, and mtDNA clade Y. However, these minor branches are probably derived from N via reversions at 10394c, which is known to be a hyper-variable site (Macaulay et al. 1999). This is clearest in the case of clade Y, whose control region motif includes np 16223C, as do those of clades B, F, and P, which represent a major subset of N sequences. Another case of reversion at 10394c is apparent in the Papuan network (fig. 4). Here, according to its control region sequence, Papuan 152 is clade K, which is a derivative of N (Macaulay et al. 1999). It is incidentally surprising to find a single K sequence, typical of Europe, in the Papuan Highlands. In short, there are at most two likely

out-of-Africa founder types identifiable at 14-enzyme resolution, namely, the root nodes of M and N. Incidentally, control-region sequencing cannot further resolve these two nodes (Watson et al. 1997; Quintana-Murci et al. 1999). We calculated the diversity ρ for these two potential founders, and thus the minimum age for the out-of-Africa migration, at 2.4861 ± 0.5232 (corresponding to 54,200 ± 11,400 years) for mtDNA clade M and 2.4512 ± 0.5364 (corresponding to 53,400 ± 11,700 years) for mtDNA clade N. These values are in excellent agreement with the age of the L3 expansion in Africa, dated at 60,000–80,000 years (Watson et al. 1997; Quintana-Murci et al. 1999), which is necessarily expected to predate the out-of-Africa migration. Note that the founding ages of 54,000 years for M and N should not be confused with their expansion ages of around 30,000 years, which we discuss below.

Discussion

Nearly all non-African mtDNA types are descended from at most two sequences at the resolution of both mtDNA control-region sequencing and 14-enzyme RFLP typing, indicating that possibly only a small number of women carrying closely related mtDNA types migrated out of Africa according to the following considerations. The founder ages of the corresponding root types of clades M and N are very similar (54,000 ± 11,500 years), raising the possibility that these two mtDNA types are derived from a single African migration. Furthermore, the fact that the M and N founder types differ by only two RFLP mutations from each other, while the average difference in the ancestral African L1 group is about 10 RFLP mutations (Chen et al. 1995), indicates that they may have belonged to a genetically closely knit group. The minimum age of about 54,000 years for the out-of-Africa migration is in agreement with the dentochronol-

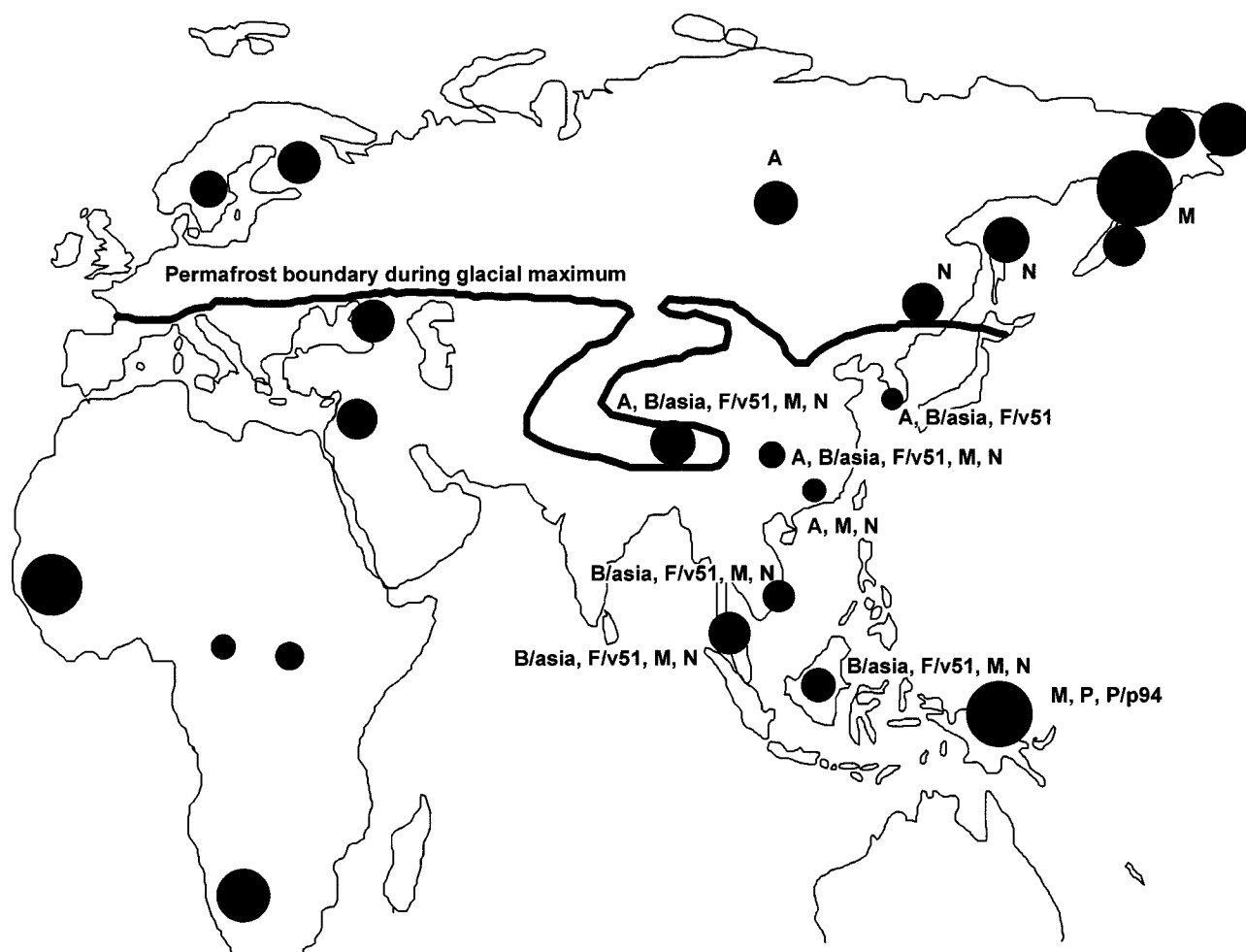


FIG. 5.—Current distribution of mtDNA expansion clusters older than 20,000 years. The permafrost boundary at the glacial maximum about 20,000 years ago is drawn according to Frenzel, Pécsi, and Velichko (1992). Circle areas are proportional to sample sizes; cluster labels refer to the local presence or absence of clusters without specifying local frequency. Note that the labels A, B/asia, etc. are used here to denote clusters as defined in figure 3; these clusters are within clades A, B, etc. defined by Macaulay et al. (1999). A single coastal Papuan (P150) has a F/v51 type which is omitted here, as it presumably represents a recent Austronesian contribution.

ogical date of $60,000 \pm 6,100$ years (Turner 1986) and with the expansion age of 60,000–80,000 years of the ancestral L2/L3 node (Watson et al. 1997), which repopulated much of Africa, with the exception of the Bushman and West Pygmy enclaves.

The oldest expansion in Eurasia occurred $65,000 \pm 23,000$ years ago (table 2) and is witnessed by mitochondrial descendants preserved in Papua New Guinea; the age estimate can be narrowed somewhat by considering that the Papuan node is derived from a Eurasian founder, so it should not be older than $54,000 \pm 12,000$ years. This is still about 20,000 years older than any mainland Asian cluster, although both the Papuans and the Asians are derived from the same two Eurasian founders. On the basis of this time difference, we tentatively propose the following scenario to account for the obvious phenotypic differences between Papuans and Asians despite their sharing a common mitochondrial ancestry: The M and N founders derive from a single African migration but split at an early stage (possibly before reaching Europe, which lacks M) into proto-Papuan and proto-Eurasian. The proto-Papuan M and N

immediately expanded demographically and geographically along a southern route until reaching Papua New Guinea, thus allowing Papuans to retain their overall genetic similarity to Africans (Stoneking et al. 1997). Meanwhile, proto-Eurasians spent 20 or more millennia genetically drifting to their present distinct European, Indian, and east Asian M and N types, as well as phenotypes (compare the common Papuan/Eurasian melanocortin receptor variants in table 1 of Harding et al. [2000]), long before expanding. The Papuan network in figure 4 shows that it may still be possible to trace the proposed southern route taken by proto-Papuan mtDNA. As had already been noted by Ballinger et al. (1992a), one Malay Malaysian (MM90) has two diagnostic sites—15606a and 2070o—in common with the Papuan P clade. Comparison of Papuan RFLP clade P with the corresponding control-region sequences (SH17 and WE17 in the appendix) demonstrates that the following control-region motif (relative to CRS) is ancestral to the Papuan cluster: 16223.C (as in CRS), 16357.C, 73.G, 212.C (corresponding to 2070o), and 263.G. Screening for these positions and 15606a in relevant populations

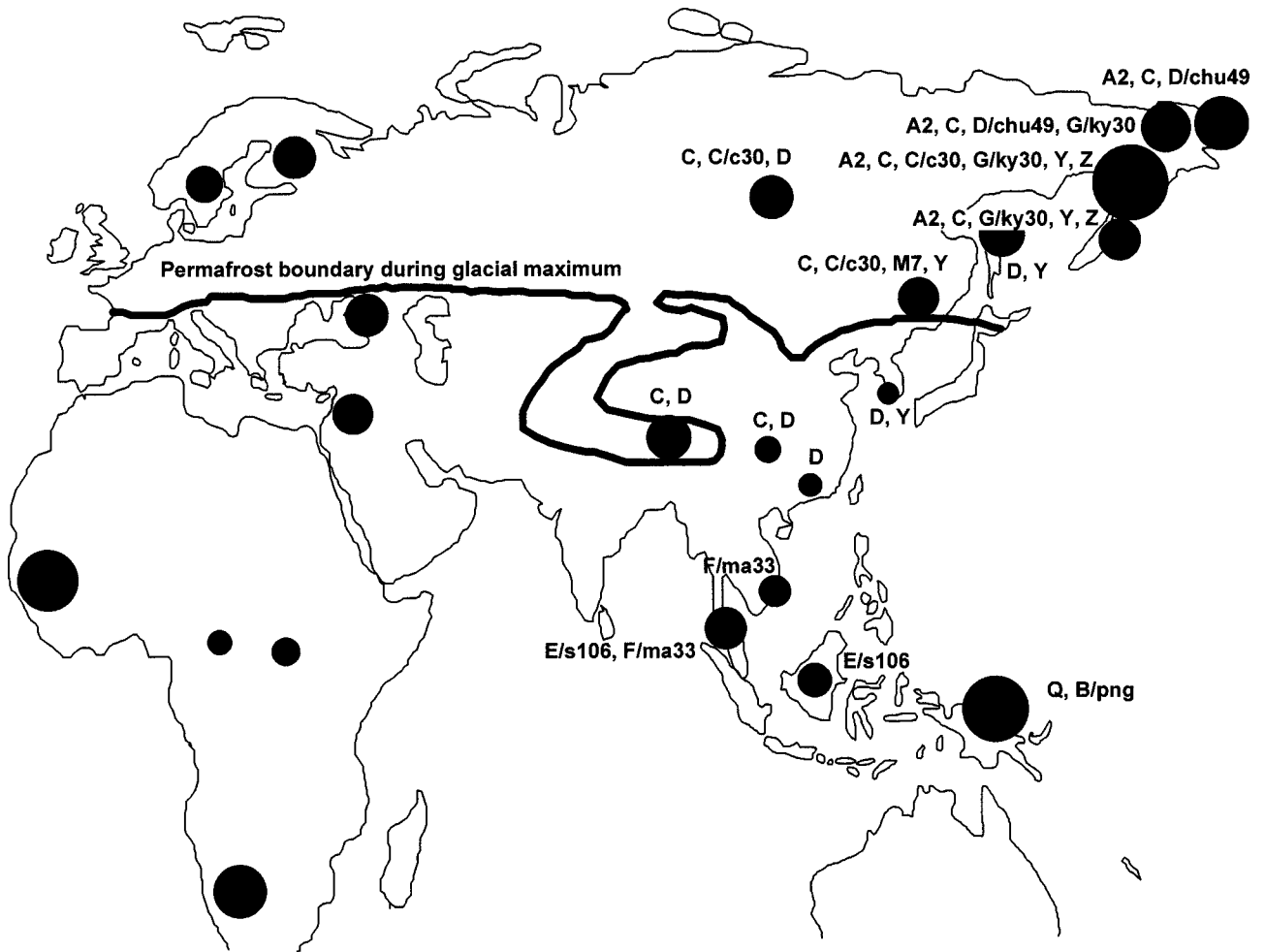


FIG. 6.—Current distribution of mtDNA expansion clusters younger than 20,000 years. The permafrost boundary at the glacial maximum about 20,000 years ago is drawn according to Frenzel, Pécsi, and Velichko (1992). Circle areas are proportional to sample sizes; cluster labels refer to the local presence or absence of clusters without specifying local frequency. Note that the labels A2, B/png, etc. are used here to denote clusters as defined in figure 3; these clusters are within clades A, B, etc. defined by Macaulay et al. (1999).

(Indians, Andaman islanders) may help to confirm the proposed southern route into Papua New Guinea. If Malaysian MM90 turns out to be representative of southeast Asia, then 12528k (fig. 4) would have mutated in or near Papua New Guinea, yielding a minimum age of $33,000 \pm 8,000$ years (table 2) for the settlement of Papua New Guinea, and a maximum age of $51,000 \pm 17,000$ years (i.e., the age of the node ancestral to MM90). Another interesting point to resolve would be the genetic relationship of Australians to Papuans and Eurasians. Unfortunately, the Australian RFLP data had to be discarded for this study (see *Materials and Methods*), but other studies on mtDNA control-region sequences (Redd and Stoneking 1999) and Y STRs (Forster et al. 1998) have shown that Papuans and Australians are not closely related as far as these loci are concerned. The oldest Australian human remains are found at Lake Mungo and dated to $62,000 \pm 6,000$ years (Thorne et al. 1999). Ancient DNA extracted from these remains (Adcock et al. 2001) may be genuine (although experimental reproduction and details are lacking), as it appears to be related to an ancient mtDNA type found to be inserted in

nuclear DNA of modern humans (Zischler et al. 1995). The absence of this mtDNA lineage in any mitochondria in over 17,000 published modern human mtDNA sequences (Röhl et al. 2001), including those of aboriginal Australians, would mean that the Lake Mungo mtDNA lineage was replaced by modern mtDNA (Adcock et al. 2001) at some time in the past 60,000 years.

At the next time level in table 2, six expansion nodes cluster closely at 27,000 to 35,000 years. If we take the mean value of 31,000 years as a lower limit for the arrival in east Asia of the African founders (a distance of about 8,000 km) and 54,000 years as the starting date, then the minimal eastward migration speed would have amounted to about 300 m/year. This rate appears plausible, as it is on the same order of magnitude as the minimal southward migration speed of Amerinds from Beringia to Chile of about 1 km per year, assuming an Alaskan entry date of 25,000 years and an arrival in Monte Verde by at least 14,000 calendar years ago (Forster et al. 1996). Similarly, the arrival of the east African L2/L3 expansion (60,000 years old) in west

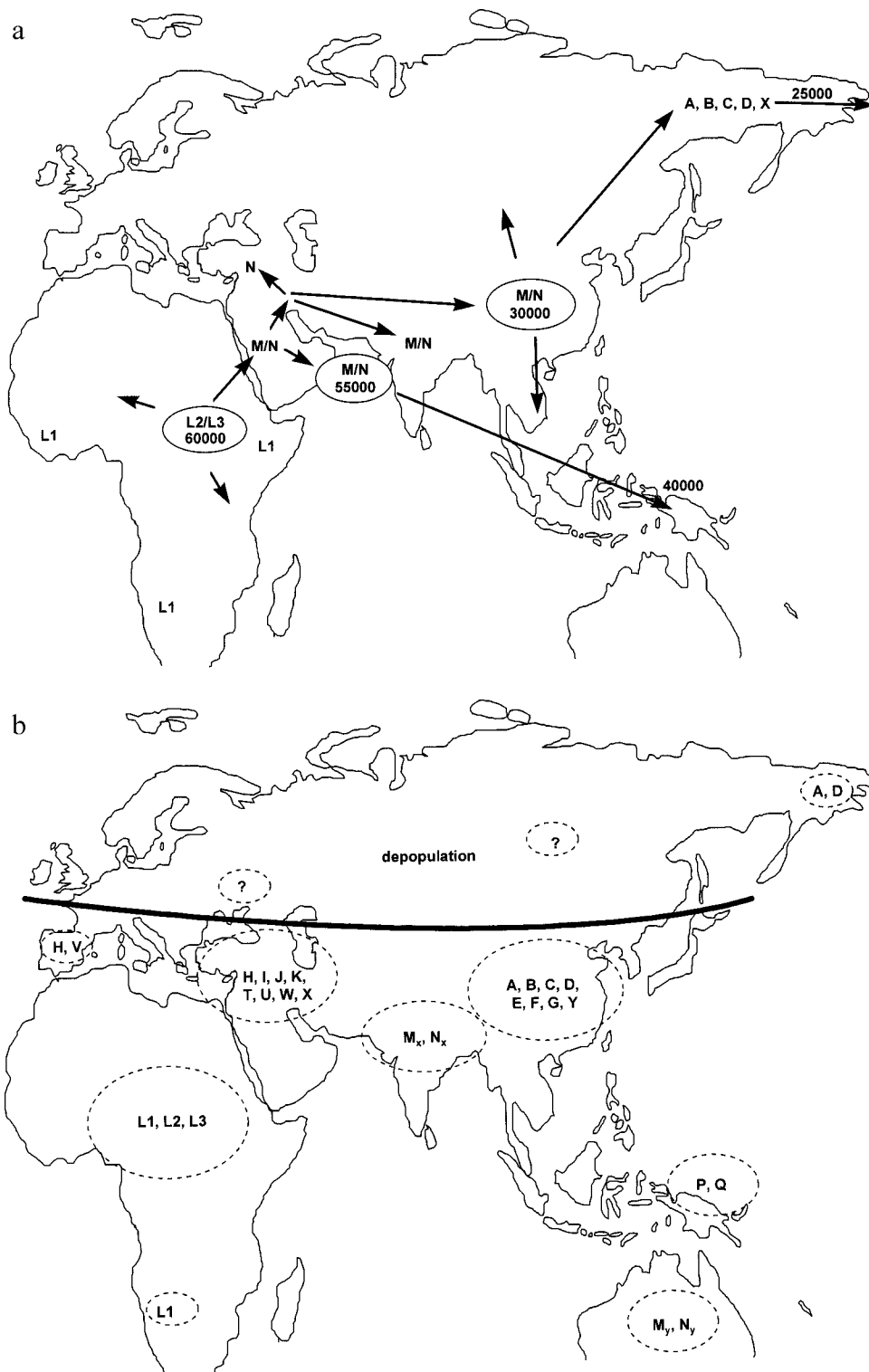


FIG. 7.—*a*, Out-of-Africa migration and establishment of regional founding mtDNA pools (weak Garden of Eden scenario) 60,000–20,000 years ago. The labels A, B, C, etc. denote mtDNA clades and are not listed exhaustively for each area. Dates within ellipses refer to demographic expansions; other dates refer to migration events; the dates may vary by more than 10,000 years within and between studies. *b*, Glacial refugia about 20,000 years ago as detected by mtDNA analysis. Question marks indicate lack of mitochondrial data for some representative areas. The postglacial re-expansion from the Iberian refugium was postulated by Torroni et al. (1998), and mtDNA analyses on the Polynesian expansion of clade B are reviewed by Richards, Oppenheimer, and Sykes (1998). The presence of further Near Eastern mtDNA clades in Europe is postulated by Richards et al. (2000). The ancient presence of N in India is corroborated by its Indian subclade of U (Kivisild et al. 1999).

Africa by 30,000 years ago (Watson et al. 1997) implies a westward migration speed of at least 200 m/year.

Inspection of the geographic distributions of the six oldest East Asian expansion clusters reveals that they are mainly located south of the permafrost boundary of the Last Glacial Maximum 20,000 years ago (fig. 5). A link with the Ice Age is strengthened by the gap in table 2 of 10,000 years during the Last Glacial Maximum before the next demographic expansion clusters occurred, all younger than 17,000 years and thus postglacial maximum. Most of these postglacial clusters are found today in northern Asia (fig. 6), excepting a few southeast Asian expansion clusters, notably, the Polynesian mtDNA clade B expansion to coastal Papua New Guinea (Stoneking et al. 1990). Taken together, this evidence strongly suggests that northern Asia was depopulated during the Last Glacial Maximum (fig. 7). The early expansions starting about 30,000 years ago in Asia did, however, reach America before they were swept back again, as is seen in the widespread presence of mtDNA clade B in America and central Asia, but not in northern Asia (Shields et al. 1993; Torroni et al. 1993a, 1993b). According to the time estimates in table 2 and the geographic distributions in figures 5 and 6, northern Asia was resettled partly from central Asia (from approximately

the latitude of Korea according to fig. 5) and partly from the Beringian glacial refuge from whence the Na Dene and Eskimo derive their mtDNA clade A2 types (Forster et al. 1996). The reexpansions from these two refugia may have contributed to the geographic patterns of the second principal component of autosomal variation in Asia (fig. 4.17.2 in Cavalli-Sforza, Menozzi, and Piazza 1994). Any attempt to search for the Amerind ancestors in modern Mongolians (Neel, Biggar, and Sukernik 1994) or elsewhere is thus going to overlook the actual ancestral Amerind population which crossed the Bering land bridge more than 20,000 years ago, presumably from northeast Asia, and then had most of its genetic traces in Asia obliterated by the ensuing glacial conditions.

Acknowledgments

We thank Mark Stoneking for providing information on RFLP data, and Lucy Forster and Matthew Hurles for assistance in preparing the manuscript. We are also grateful to Naruya Saitou and two anonymous reviewers for valuable comments. This project was supported by The British Council and by the Deutsche Akademische Austauschdienst.

APPENDIX

Correspondence of Samples RFLP-Typed for the Complete mtDNA Molecule by Cann, Stoneking, and Wilson (1987) (CAN)/Stoneking et al. (1990) (STO) with Samples Sequenced for the mtDNA Control Region by Vigilant (1990) (VIG)

ID	CAN	STO	VIG	ID	CAN	STO	VIG	ID	CAN	STO	VIG
SP1	73	nd	123	UC29	118	nd	nd	ENB2	nd	130	nd
SP2	80	nd	nd	UC30	17	nd	nd	ENB3	nd	11	50
SP3	53	nd	nd	UC31	25	nd	nd	ENB5	nd	130	129
SP4	2	nd	3	UC32	76	nd	nd	NP1	nd	69	nd
SP5	107	nd	nd	UC33	43	nd	nd	NP2	nd	119	132
SP6	80	nd	nd	UC34	50	nd	86	NP3	nd	129	133
SP7	116	nd	nd	UC35	77	nd	nd	NP4	nd	93	nd
SP8	104	nd	nd	UC36	113	nd	nd	NP5	nd	69	nd
SP9	115	nd	nd	UC37	8	nd	23	NP6	nd	119	nd
SL2	100	nd	nd	UC38	51	nd	nd	GP1	nd	119	nd
AB1	24	nd	nd	UC39	96	nd	98	GP2	nd	106	nd
AB2	114	nd	nd	DH1	36	nd	91	GP3	nd	27	nd
AB4	11	nd	74	DH2	7	nd	35	GP4	nd	43	nd
AB5	115	nd	nd	DH3	71	nd	nd	GP5	nd	68	nd
HD1	122	nd	nd	DH4	16	nd	nd	GP6	nd	130	131
HD2	97	nd	nd	DH5	39	nd	nd	GP7	nd	43	79
HD3	106	nd	nd	DH6	40	nd	63	GP8	nd	110	nd
HD4	15	nd	87	DH7	18	nd	nd	ENGA1	nd	27	nd
HD5	41	nd	nd	AUS1	23	nd	nd	WE1	nd	33	nd
101	61	nd	nd	AUS2	10	nd	49	WE2	nd	120	135
102	108	nd	nd	AUS3	35	nd	nd	WE3	nd	94	nd
103	114	nd	nd	AUS4	58	nd	nd	WE4	nd	32	nd
105	47	nd	nd	AUS5	69	nd	nd	WE5	nd	27	nd
106	94	nd	nd	AUS6	33	nd	nd	WE6	nd	35	nd
237	93	nd	nd	AUS7	52	nd	nd	WE7	nd	130	nd
KP1	62	nd	128	AUS8	64	nd	nd	WE8	nd	27	nd
KP2	84	nd	nd	AUS9	14	nd	nd	WE9	nd	31	nd
KP3	105	nd	nd	AUS10	22	nd	nd	WE10	nd	95	nd
KP4	81	nd	nd	AUS11	32	nd	nd	WE11	nd	119	135
NC1	4	nd	nd	AUS12	42	nd	nd	WE12	nd	92	nd
NC2	5	nd	33	AUS13	63	nd	nd	WE13	nd	27	nd
NC3	70	nd	100	AUS14	68	nd	nd	WE14	nd	26	nd
WB1	103	nd	nd	AUS15	126	nd	nd	WE15	nd	130	135
WB2	90	nd	nd	AUS16	124	nd	nd	WE16	nd	25	nd
WB3	91	nd	nd	AUS17	125	nd	nd	WE17	nd	104	110
WB4	74	nd	nd	AUS20	59	nd	nd	WE18	nd	119	nd

APPENDIX
Continued

ID	CAN	STO	VIG	ID	CAN	STO	VIG	ID	CAN	STO	VIG
WB5	21	nd	nd	AUS21	123	nd	nd	WE20	nd	30	nd
WB6	80	nd	nd	AUS22	59	nd	nd	WE21	nd	27	nd
WB7	6	nd	36	AUS23	31	nd	nd	WE22	nd	114	nd
WB8	72	nd	124	SA1	1	nd	13	WE23	nd	36	nd
WB9	99	nd	nd	EH1	95	152	97	WE24	nd	121	134
WB10	20	nd	nd	EH2	65	89	nd	WE25	nd	27	nd
WB11	109	nd	nd	EH3	26	39	nd	WE26	nd	122	135
WB12	119	nd	nd	EH4	13	40	nd	SH2	nd	37	nd
WB13	78	nd	nd	EH5	130	101	nd	SH3	nd	108	nd
WB14	48	nd	nd	EH6	131	105	nd	SH5	nd	27	nd
WB15	3	nd	27	EH7	65	89	nd	SH7	nd	115	nd
WB16	66	nd	nd	EH8	29	27	nd	SH8	nd	115	nd
WB17	60	nd	nd	EH9	133	109	nd	SH9	nd	27	nd
WB18	121	nd	nd	EH10	29	27	nd	SH10	nd	102	nd
WB19	79	nd	nd	EH11	128	100	nd	SH11	nd	45	nd
WB20	67	nd	nd	SH1	27	40	nd	SH12	nd	113	nd
HELA	45	nd	nd	EH12	49	65	nd	SH13	nd	116	nd
UC1	112	nd	nd	EH13	29	27	nd	SH14	nd	27	nd
UC2	37	nd	nd	EH14	127	99	nd	SH15	nd	26	80
UC3	55	nd	nd	EH15	134	107	nd	SH16	nd	27	nd
UC4	86	nd	nd	EH16	12	24	nd	SH17	nd	116	108
UC5	87	nd	nd	EH17	28	34	nd	SH18	nd	27	nd
UC6	19	nd	nd	EH18	134	107	nd	SH19	nd	46	nd
UC7	30	nd	nd	EH19	132	103	nd	SH20	nd	97	nd
UC8	57	nd	nd	EH20	134	107	nd	SH21	nd	46	nd
UC9	102	nd	nd	EH21	134	107	nd	SH22	nd	115	nd
UC10	56	nd	nd	MOR1	129	96	nd	SH23	nd	29	81
UC11	101	nd	nd	EH22	134	107	nd	SH24	nd	27	nd
UC12	85	nd	nd	EH23	65	89	nd	SH25	nd	27	nd
UC13	46	nd	nd	EH24	134	107	nd	SH26	nd	112	nd
UC14	82	nd	nd	Anderson	110	nd	nd	SH27	nd	27	nd
UC15	98	nd	nd	EH25	nd	114	nd	SH28	nd	115	nd
UC16	88	nd	nd	CP1	nd	130	nd	SH29	nd	88	nd
UC17	38	nd	nd	CP2	nd	119	nd	SH30	nd	41	nd
UC18	34	nd	nd	CP3	nd	67	nd	SH31	nd	38	nd
UC19	111	nd	nd	CP4	nd	119	130	SH32	nd	27	nd
UC20	92	nd	93	CP5	nd	27	nd	SH33	nd	27	82
UC21	89	nd	95	CP6	nd	127	125	SH34	nd	28	nd
UC22	75	nd	nd	CP7	nd	130	nd	SH35	nd	42	nd
UC23	54	nd	nd	CP8	nd	119	nd	SH36	nd	90	nd
UC24	9	nd	28	CP10	nd	119	nd	SH37	nd	91	nd
UC25	117	nd	113	CP11	nd	119	nd	SH38	nd	27	nd
UC26	44	nd	121	CP12	nd	150	nd	SH39	nd	98	nd
UC27	83	nd	nd	CP14	nd	130	nd				
UC28	120	nd	nd	ENB1	nd	111	nd				

NOTE.—The numbers in the CAN, STO, and VIG columns refer to the published phylogenetic trees. nd = not determined.

LITERATURE CITED

- ADCOCK, G. J., E. S. DENNIS, S. EASTEAL, G. A. HUTTLEY, L. S. JERMIIN, W. J. PEACOCK, and A. THORNE. 2001. Mitochondrial DNA sequences in ancient Australians: implications for modern human origins. *Proc. Natl. Acad. Sci. USA* **98**:537–542.
- ANDERSON, S., A. T. BANKIER, B. G. BARRELL et al. (14 co-authors). 1981. Sequence and organisation of the human mitochondrial genome. *Nature* **290**:457–465.
- ANDREWS, R. M., I. KUBACKA, P. F. CHINNERY, R. N. LIGHTOWLERS, D. M. TURNBULL, and N. HOWELL. 1999. Re-analysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**:147.
- ARIS-BROSOU, S., and L. EXCOFFIER. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol. Biol. Evol.* **13**:494–504.
- ARMOUR, J. A. L., T. ANTTINEN, C. A. MAY, E. E. VEGA, A. SAJANTILA, J. R. KIDD, K. K. KIDD, J. BERTRANPETIT, S. PÄÄBO, and A. J. JEFFREYS. 1996. Minisatellite diversity supports a recent African origin for modern humans. *Nat. Genet.* **13**:154–160.
- BALLINGER, S. W., T. G. SCHURR, A. TORRONI, Y. Y. GAN, J. A. HODGE, K. HASSAN, K.-H. CHEN, and D. C. WALLACE. 1992a. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* **130**:139–152.
- . 1992b. Corrigendum. *Genetics* **130**:957.
- BANDEL, H.-J., and P. FORSTER. 1997. The myth of bumpy hunter-gatherer mismatch distributions. *Am. J. Hum. Genet.* **61**:980–983.
- BANDEL, H.-J., P. FORSTER, and A. RÖHL. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**:37–48.
- BANDEL, H.-J., P. FORSTER, B. C. SYKES, and M. B. RICHARDS. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* **141**:743–753.
- BRÄUER, G. 1989. The evolution of modern humans: a comparison of the African and non-African evidence. Pp. 123–154 in

- P. MELLARS and C. STRINGER, eds. The human revolution: behavioural and biological perspectives in the origins of modern humans. Edinburgh University Press, Edinburgh, Scotland.
- CANN, R. L., M. STONEKING, and A. C. WILSON. 1987. Mitochondrial DNA and human evolution. *Nature* **325**:31–36.
- CAVALLI-SFORZA, L. L., P. MENOZZI, and A. PIAZZA. 1994. The history and geography of human genes. Princeton University Press, Princeton, N.J.
- CHEN, Y.-S., A. TORRONI, L. EXCOFFIER, A. S. SANTACHIARA-BENERECETTI, and D. C. WALLACE. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* **57**:133–149.
- DAY, M. H., and C. B. STRINGER. 1982. A reconsideration of the Omo Kibish remains and the erectus-sapiens transition. Pp. 814–846 in H. DE LUMLEY, ed. *L'Homo erectus et la place de l'homme de Tautavel parmi les hominidés fossiles*. Centre National de la Recherche Scientifique/Louis-Jean Scientific and Literary, Nice, France.
- FORSTER, P., R. HARDING, A. TORRONI, and H.-J. BANDELT. 1996. Origin and evolution of native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* **59**:935–945.
- . 1997. Reply to Bianchi and Bailliet. *Am. J. Hum. Genet.* **61**:246–247.
- FORSTER, P., M. KAYSER, E. MEYER, L. ROEWER, H. PFEIFFER, H. BENKMANN, and B. BRINKMANN. 1998. Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. *Mol. Biol. Evol.* **15**:1108–1114.
- FORSTER, P., A. RÖHL, P. LÜNNEMANN, C. BRINKMANN, T. ZERJAL, C. TYLER-SMITH, and B. BRINKMANN. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. *Am. J. Hum. Genet.* **67**:182–196.
- FRENZEL, B., M. PÉCSI, and A. VELICHKO. 1992. Atlas of Paleoclimates and paleoenvironments of the Northern Hemisphere. Late Pleistocene–Holocene. Hungarian Academy of Sciences, Gustav-Fischer-Verlag, Budapest/Stuttgart.
- HARDING, R. M., S. M. FULLERTON, R. C. GRIFFITHS, J. BOND, M. J. COX, J. A. SCHNEIDER, D. S. MOULIN, and J. B. CLEGG. 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* **60**:772–789.
- HARDING, R. M., E. HEALY, A. J. RAY et al. (11 co-authors). 2000. Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet.* **66**:1351–1361.
- HARPENDING, H. C., M. A. BATZER, M. GURVEN, L. B. JORDE, A. R. ROGERS, and S. T. SHERRY. 1998. Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA* **95**:1961–1967.
- HARPENDING, H. C., S. T. SHERRY, A. R. ROGERS, and M. STONEKING. 1993. The genetic structure of ancient human populations. *Curr. Anthropol.* **34**:483–496.
- HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE, and N. TAKAHATA. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci. USA* **92**:532–536.
- INGMAN, M., H. KAESSMANN, S. PÄÄBO, and U. GYLLENSTEN. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**:708–713.
- JORDE, L. B., and M. BAMSHAD. 2000. Questioning evidence for recombination in human mitochondrial DNA. www.sciencemag.org/cgi/content/full/288/5474/1931a.
- KIVISILD, T., M. J. BAMSHAD, K. KALDMA et al. (15 co-authors). 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* **9**:1331–1334.
- KIVISILD, T., and R. VILLEMS. 2000. Questioning evidence for recombination in human mitochondrial DNA. www.sciencemag.org/cgi/content/full/288/5474/1931a.
- KRINGS, M., A. STONE, R. W. SCHMITZ, H. KRAINITZKI, M. STONEKING, and S. PÄÄBO. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* **90**:19–30.
- KUMAR, S., P. HEDRICK, T. DOWLING, and M. STONEKING. 2000. Questioning evidence for recombination in human mitochondrial DNA. www.sciencemag.org/cgi/content/full/288/5474/1931a.
- MACAULAY, V., M. RICHARDS, E. HICKEY, E. VEGA, F. CRUCIANI, V. GUIDA, R. SCOZZARI, B. BONNÉ-TAMIR, B. SYKES, and A. TORRONI. 1999. The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**:232–249.
- MATEU, E., D. COMAS, F. CALAFELL, A. PÉREZ-LEZAUN, A. ABADE, and J. BERTRANPETIT. 1997. A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann. Hum. Genet.* **61**:507–518.
- MELTON, T., R. PETERSON, A. J. REDD, N. SAHA, A. S. M. SOFRO, J. MARTINSON, and M. STONEKING. 1995. Polynesian genetic affinities with southeast Asian populations as identified by mtDNA analysis. *Am. J. Hum. Genet.* **57**:403–414.
- MORRAL, N., J. BERTRANPETIT, X. ESTIVILL et al. (31 co-authors). The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nat. Genet.* **7**:169–175.
- NEEL, J. V., R. J. BIGGAR, and R. I. SUKERNIK. 1994. Virologic and genetic studies relate Amerind origins to the indigenous people of the Mongolia/Manchuria/southeastern Siberia region. *Proc. Natl. Acad. Sci. USA* **91**:10737–10741.
- OOTA, H., N. SAITOU, T. MATSUSHITA, and S. UEDA. 1999. Molecular genetic analysis of remains of a 2000-year-old human population in China—and its relevance for the origin of the modern Japanese population. *Am. J. Hum. Genet.* **64**:250–258.
- OVCHINNIKOV, I. V., A. GÖTHERSTRÖM, G. P. ROMANOVA, V. M. KHARITONOV, K. LIDÉN, and W. GOODWIN. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* **404**:490–493.
- PARSONS, T. J., and J. A. IRWIN. 2000. Questioning evidence for recombination in human mitochondrial DNA. www.sciencemag.org/cgi/content/full/288/5474/1931a.
- PASSARINO, G., O. SEMINO, L. F. BERNINI, and A. S. SANTACHIARA-BENERECETTI. 1996. Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am. J. Hum. Genet.* **59**:927–934.
- PENNY, D., M. STEEL, P. J. WADDELL, and M. D. HENDY. 1995. Improved analysis of human mtDNA sequences support a recent African origin for *Homo sapiens*. *Mol. Biol. Evol.* **12**:863–882.
- PRITCHARD, J. K., and M. W. FELDMAN. 1996. Genetic data and the African origin of humans. *Science* **274**:1548.
- QUINTANA-MURCI, L., O. SEMINO, H.-J. BANDELT, G. PASSARINO, K. MCELREAVEY, and A. S. SANTACHIARA-BENERECETTI. 1999. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat. Genet.* **23**:437–441.
- REDD, A. J., and M. STONEKING. 1999. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. *Am. J. Hum. Genet.* **65**:808–828.
- RELETHFORD, J. H., and L. B. JORDE. 1999. Genetic evidence for larger African population size during recent human evolution. *Am. J. Phys. Anthropol.* **108**:251–260.
- RICHARDS, M. B., H. CÔRTE-REAL, P. FORSTER, V. MACAULAY, H. WILKINSON-HERBOTS, A. DEMAINE, S. PAPIHA, R. HEDGES, H.-J. BANDELT, and B. C. SYKES. 1996. Palaeolithic and

- neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **59**:185–203.
- RICHARDS, M., S. OPPENHEIMER, and B. SYKES. 1998. mtDNA suggests Polynesian origins in eastern Indonesia. *Am. J. Hum. Genet.* **63**:1234–1236.
- RICHARDS, M., V. MACAULAY, E. HICKEY et al. (37 co-authors). 2000. Tracing European founder lineages in the near Eastern mitochondrial gene pool. *Am. J. Hum. Genet.* **67**:1251–1276.
- RISCH, N., K. K. KIDD, and S. A. TISHKOFF. 1996. Genetic data and the African origin of humans. *Science* **274**:1548–1549.
- RÖHL, A. 1999. Phylogenetische Netzwerke. Ph.D. thesis, University of Hamburg, Hamburg, Germany.
- RÖHL, A., B. BRINKMANN, L. FORSTER, and P. FORSTER. 2001. An annotated mtDNA database. *Int. J. Legal Med.* **115**:29–39.
- SAILLARD, J., P. FORSTER, N. LYNNERUP, H.-J. BANDEL, and S. NØRBY. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am. J. Hum. Genet.* **67**:718–726.
- SCHURR, T. G., R. I. SUKERNIK, Y. B. STARIKOVSKAYA, and D. C. WALLACE. 1999. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea–Bering Sea region during the Neolithic. *Am. J. Phys. Anthropol.* **108**:1–39.
- SHERRY, S. T., A. R. ROGERS, H. HARPENDING, T. JENKINS, and M. STONEKING. 1994. Mismatch distributions of mtDNA reveal recent human population expansions. *Hum. Biol.* **66**:761–775.
- SHIELDS, G. F., A. M. SCHMIECHEN, B. L. FRAZIER, A. REDD, M. I. VOEVODA, J. K. REED, and R. H. WARD. 1993. mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am. J. Hum. Genet.* **53**:549–562.
- SLATKIN, M. 1996. Gene genealogies within mutant allelic classes. *Genetics* **143**:579–587.
- STARIKOVSKAYA, Y. B., R. I. SUKERNIK, T. G. SCHURR, A. M. KOGELNIK, and D. C. WALLACE. 1998. mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *Am. J. Hum. Genet.* **63**:1473–1491.
- STONEKING, M., J. J. FONTIUS, S. L. CLIFFORD, H. SOODYALL, S. S. ARCOT, N. SAHA, T. JENKINS, M. A. TAHIR, P. L. DEININGER, and M. A. BATZER. 1997. *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* **7**:1061–1071.
- STONEKING, M., L. B. JORDE, K. BHATIA, and A. C. WILSON. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* **124**:717–733.
- STUMPF, M. P. H., and D. B. GOLDSTEIN. 2001. Genealogical and evolutionary inference with the human Y chromosome. *Science* **291**:1738–1742.
- SYKES, B., A. LEIBOFF, J. LOW-BEER, S. TETZNER, and M. RICHARDS. 1995. The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet.* **57**:1463–1475.
- TEMPLETON, A. R. 1998. Nested cladistic analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* **7**:381–397.
- THORNE, A., R. GRÜN, G. MORTIMER, N. A. SPOONER, J. J. SIMPSON, M. McCULLOCH, L. TAYLOR, and D. CURNOE. 1999. Australia's oldest human remains: age of the Lake Mungo 3 skeleton. *J. Hum. Evol.* **36**:591–612.
- TISHKOFF, S. A., E. DIETZSCH, W. SPEED et al. (15 co-authors). 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**:1380–1387.
- TORRONI, A., H.-J. BANDEL, L. D'URBANO et al. (11 co-authors). 1998. mtDNA analysis reveals a major late Palaeolithic population expansion from southwestern to northeastern Europe. *Am. J. Hum. Genet.* **62**:1137–1152.
- TORRONI, A., Y.-S. CHEN, O. SEMINO, A. S. SANTACHIARA-BENERECETTI, C. R. SCOTT, M. T. LOTT, M. WINTER, and D. C. WALLACE. 1994a. mtDNA and Y-chromosome polymorphisms in four native American populations from southern Mexico. *Am. J. Hum. Genet.* **54**:303–348.
- TORRONI, A., K. HUOPONEN, P. FRANCALACCI, M. PETROZZI, L. MORELLI, R. SCOZZARI, D. OBINU, M.-L. SAVONTAUS, and D. C. WALLACE. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**:1835–1850.
- TORRONI, A., M. T. LOTT, M. F. CABELL, Y.-S. CHEN, L. LAVERGNE, and D. C. WALLACE. 1994b. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am. J. Hum. Genet.* **55**:760–776.
- TORRONI, A., J. A. MILLER, L. G. MOORE, S. ZAMUDIO, J. ZHUANG, T. DROMA, and D. C. WALLACE. 1994c. Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am. J. Phys. Anthropol.* **93**:189–199.
- TORRONI, A., T. G. SCHURR, M. F. CABELL, M. D. BROWN, J. V. NEEL, M. LARSEN, D. G. SMITH, C. M. VULLO, and D. C. WALLACE. 1993a. Asian affinities and continental radiation of the four founding native American mtDNAs. *Am. J. Hum. Genet.* **53**:563–590.
- TORRONI, A., R. I. SUKERNIK, T. G. SCHURR, Y. B. STARIKOVSKAYA, M. F. CABELL, M. H. CRAWFORD, A. S. G. COMUZIE, and D. C. WALLACE. 1993b. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* **53**:591–608.
- TURNER, C. G. 1986. Dentochronological separation estimates for Pacific rim populations. *Science* **232**:1140–1142.
- VIGILANT, L. 1990. Control region sequences from African populations and the evolution of human mitochondrial DNA. Ph.D. thesis, University of California, Berkeley.
- VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES, and A. C. WILSON. 1991. African populations and the evolution of human mitochondrial DNA. *Science* **253**:1503–1507.
- WAKELEY, J., and J. HEY. 1997. Estimating ancestral population parameters. *Genetics* **145**:847–855.
- WANG, L., H. OOTA, N. SAITOU, F. JIN, T. MATSUSHITA, and S. UEDA. 2000. Genetic structure of a 2,500-year-old human population in China and its spatiotemporal changes. *Mol. Biol. Evol.* **17**:1396–1400.
- WATSON, E., P. FORSTER, M. RICHARDS, and H.-J. BANDEL. 1997. Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.* **61**:691–704.
- WILLS, C. 1995. Topiary pruning and weighting reinforce an African origin for the human mitochondrial DNA tree. *Evolution* **50**:977–989.
- WILSON, A. C., R. L. CANN, S. M. CARR et al. (11 co-authors). 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* **26**:375–400.
- WRISHNIK, L. A., R. G. HIGUCHI, M. STONEKING, H. A. ERLICH, N. ARNHEIM, and A. C. WILSON. 1987. Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acids Res.* **15**:529–542.
- ZISCHLER, H., H. GEISERT, A. VON HAESLER, and S. PÄÄBO. 1995. A nuclear “fossil” of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**:489–492.

NARUYA SAITOU, reviewing editor

Accepted June 6, 2001