Human Molecular Genetics, 2000, Vol. 9, No. 3 421–430

A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome

Fabrício R. Santos[+], Arpita Pandya, Manfred Kayser[1,§], R. John Mitchell[2], Aiping Liu[3], Lalji Singh[4], Giovanni Destro-Bisol[5], Andrea Novelletto[6], Raheel Qamar[7], S. Qasim Mehdi[7], Raju Adhikari[¶], Peter de Knijff[8] and Chris Tyler-Smith[‡]

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK, [1]Institut für Rechtsmedizin, Humboldt-Universität zu Berlin, Berlin, Germany, [2]Unit of Human Genetics, School of Biochemistry and Genetics, La Trobe University, Bundoora, Victoria 3083, Australia, [3]Auckland Cancer Research Centre, Faculty of Medicine and Health Science, University of Auckland, Private Bag 92019, Auckland, New Zealand, [4]Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad, India, [5]Dipartimento di Biologia Animale e dell'Uomo, Università 'La Sapienza', Roma, Italy, [6]Department of Cell Biology, University of Calabria, Rende, Italy, [7]Biomedical and Genetic Engineering Laboratories, PO Box 2891, Islamabad, Pakistan, [8]Forensic Laboratory for DNA Research, MGC-Department of Human and Clinical Genetics, Leiden University Medical Center, PO Box 9503, 2300 RA Leiden, The Netherlands

Received 18 October 1999; Revised and Accepted 30 November 1999 DDBJ/EMBL/GenBank accession nos AF207859 and AF189307

We have identified a novel polymorphic L1 retroposon insertion, designated LY1, in the centromeric alphoid array of the human Y chromosome. The element belongs to the transpositionally active Ta subset and its presence is compatible with normal centromere function. It was found at highest frequency in China, where it accounts for 23% of the Han sample, and was present at low frequencies in the surrounding areas, but was not found at all outside Asia. Chromosomes carrying LY1 show considerable microsatellite diversity, suggesting an ancient origin for the lineage at ~10 000 years ago (with wide confidence limits), but only limited subsequent migration.

INTRODUCTION

The Y chromosome has a unique role in human population genetics with several properties that distinguish it from all other segments of the genome (1,2). The major portion, ~60 million base pairs (3), is transmitted exclusively from father to son. It is haploid, so there are no recombination events. It therefore makes up the largest DNA segment in the human genome where variation can only accumulate due to mutations. The number of Y chromosomes in the population is generally one quarter of the number of each autosome. This reduced effective population size and male behaviour, where marriage customs and migration can differ from female behaviour, are likely to influence the way in which Y diversity is distributed among human populations.

Some early surveys of Y variation (4) emphasized the paucity of DNA polymorphisms, but subsequently the use of more efficient methods for detecting variation and the inclusion of Y chromosomes from distinct geographical regions has yielded many variants. These include slowly evolving markers, mainly SNPs (5–12); in addition, several rapidly evolving Y microsatellite loci have been discovered and found to display high variability (5,13–16).

Since there are few genes on the Y and most polymorphisms lie in non-coding DNA, patterns of Y variation in distinct populations are usually considered to result from neutral processes such as mutation and random genetic drift. However, because of the absence of recombination, the occurrence of a mutation in coding sequences or other sequence structures could lead to a selective sweep affecting all polymorphisms on the Y by a hitchhiking effect (17). These processes, whether neutral or selective, have led to high levels of geographical clustering of Y variants and a correlation with language (9,10,12,18–20). This makes the Y a very useful tool for investigations of human population genetics, but leads to the requirement of a large number of polymorphic markers, including ones ascertained from each population and arising at different historical and pre-historical times.

Among slowly evolving markers, retrotransposon insertion polymorphisms are particularly useful because their ancestral state (the absence of the insertion) is known, recurrent insertion at the same position is unlikely, and there is no mechanism known for specific deletion (21). Most retroposons have resulted from very ancient insertion events and can be used in phylogenetic analysis between species (22). However, some retroposon insertions have happened more recently, after speciation, and can reveal intra-specific polymorphisms. Some

[+]Present address: Departamento de Biologia Geral, ICB/UFMG, Caixa Postal 486, 31.270–910 Belo Horizonte, MG, Brazil
[§]Present address: Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, Inselstrasse 22, D-04103 Leipzig, Germany
[¶]Present address: Nuffield Department of Clinical Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, UK
[‡]To whom correspondence should be addressed. Tel: +44 1865 275222; Fax: +44 1865 275259; Email: chris@bioch.ox.ac.uk

autosomal polymorphic retroposons have been found in the human genome (21). In addition, the only reported retroposon insertion on the human Y chromosome, the Y Alu polymorphism YAP (7), has been widely used in human evolutionary studies. It is found at high frequency in Africa and in parts of Asia and Europe and the geographical distribution of its variants has led to the novel suggestion that most African Y chromosomes have been replaced by Asian ones within the last ~30 000 years (23,24).

We now describe a polymorphic L1 retroposon element on the human Y chromosome. The insertion occurred in the alphoid DNA of the Y centromere, where it appears to be compatible with normal centromere function and is now found in a substantial percentage of Chinese males. The analysis of microsatellite variation in Y chromosomes with the L1 insertion indicates that, despite having a distribution essentially restricted to China, it is likely to be an old event originating in the Palaeolithic.

RESULTS

Identification of an L1 insertion within the Y chromosome alphoid array

Y alphoid DNA has a simple tandemly repeating structure with a high degree of homology between units (25). Consequently, restriction enzymes are expected to cut either within each unit, or not at all within the entire array. Cleavage patterns have been determined experimentally using pulsed-field gel electrophoresis (PFGE) (26,27) and, in most individuals, fulfil these expectations. However, in a small number of individuals [initially m38 (26)], an additional site for many enzymes has been found within the array. The mapping of these sites in one individual (m255) (Fig. 1a) revealed that 12 of 13 lie within a single cluster with a size below the limit of resolution of PFGE. We therefore set out to establish the molecular basis of this variation.

A cosmid library was constructed from m255 and screened with the pYα1 probe using conditions that detect typical and diverged Y alphoid DNA, but not alphoid DNA from other chromosomes. We expected that clones spanning the variant region would contain typical Y alphoid units linked to sequences with a different restriction pattern, perhaps consisting of non-alphoid DNA. One cosmid (53C), containing both typical Y alphoid DNA fragments and atypically sized fragments, was isolated and characterized further.

Hybridization of *Eco*RI digests of cosmid 53C with pYα1 revealed the presence of a band containing only non-alphoid DNA (data not shown). This non-alphoid fragment was subcloned and sequenced. The 334 bp fragment (Fig. 1b) was highly homologous to the L1 family of repeats, showing that, in this cosmid clone, alphoid DNA was linked to an L1 element. Experiments described below show that the clone was representative of genomic DNA and was not an *in vitro* cloning artefact. We therefore conclude that the alphoid array of m255 contains an L1 insertion, designated LY1.

Characterization of the L1 element

L1 elements differ in sequence, and a classification thought to reflect their evolutionary history has been proposed (28,29). The subcloned region of 334 bp had high homology to specific members of the Ta subset of L1 elements. The presence of an *Mbo*I site at the beginning of the sequence attached to the SuperCos multicloning site indicated that it was derived from the edge of the 53C clone. As the sequenced L1 was highly homologous (333/334 bp) to position 3097–3430 of the typical Ta subset retrotransposon L1.2 (GenBank accession no. M80343), cosmid 53C was predicted to carry ~3 kb of L1 sequence, including its 3′ end with the poly(A) tail probably attached to alphoid sequences. To test this prediction, we designed PCR primers based on the known Ta subset of L1 sequences (30) and the Y alphoid consensus (25) to allow amplification and subcloning of the likely 3′ insertion site of the L1 element. Excluding the primers, poly(A) tail and alphoid DNA, the PCR subclones produced 72 bp of sequence with complete homology to position 5952–6023 of the retrotransposon L1.2. Nucleotide variants characteristic of the small number of L1 sequences that currently undergo active transposition are found at the 3′ end. LY1 shows a perfect match to most members of the Ta subset, followed by a poly(A) tail of ~40 nucleotides. The 3′ flanking sequence is Y alphoid DNA, but the exact insertion site cannot be deduced as we lack the duplicated target site just before the 5′ end of LY1 which has not been cloned. However, the next recognizable alphoid sequence after the poly(A) tract is composed of GAAAA-GAAAGG, a polypurine tract forming a likely insertion site for retroposons as previously suggested (31,32).

The cosmid clone ends within the L1 sequence and contains its 3′ end with a poly(A) tail, showing that the element is at least 3 kb long, but does not reveal the full length of LY1. Attempts to amplify the 5′ end of the element from the genomic DNA of m255 using primers designed from the L1 and Y alphoid sequences were unsuccessful. However, the map of the array (Fig. 1a) includes sites for *Bgl*I, *Pvu*II, *Bst*EII and *Sac*I within the internal cluster, and the typical L1 Ta subset L1.2 sequence contains these sites only at positions 17, 656 + 798, 1627 and 1857, respectively, from the 5′ end. It is therefore likely that additional 5′ L1 sequences are present and that LY1 represents a full-length or nearly full-length member of the actively transposing Ta subset.

World-wide distribution of the LY1 chromosomes

We tested 633 males from around the world by PFGE for the presence of an internal cluster of three or more restriction enzyme sites. Of these, 15 (~2.5%) contained these sites (Table 1; but see below) and are therefore likely to carry LY1. A PCR assay was established to detect the presence of LY1 directly, using one primer from within the L1 sequence, and the other from within the alphoid DNA. No product was obtained from female DNA, or from males lacking the cluster of restriction sites, but a product of ~140 bp was seen when either the cosmid 53C or males carrying the cluster of restriction sites were amplified (Fig. 2). Another Y locus, 92R7 (Fig. 2, 55 bp), was included in the same PCR reaction as a control for successful amplification. The LY1 PCR product (Fig. 2) appeared as a fuzzy band probably due to *Taq* polymerase slippage across the poly(A) tail during PCR. Indeed, the two distinct PCR clones sequenced (53Cj1 and 53Cj2) displayed a 10 bp difference in the length of the poly(A) tail (data not shown). A few individuals also displayed visible variation in

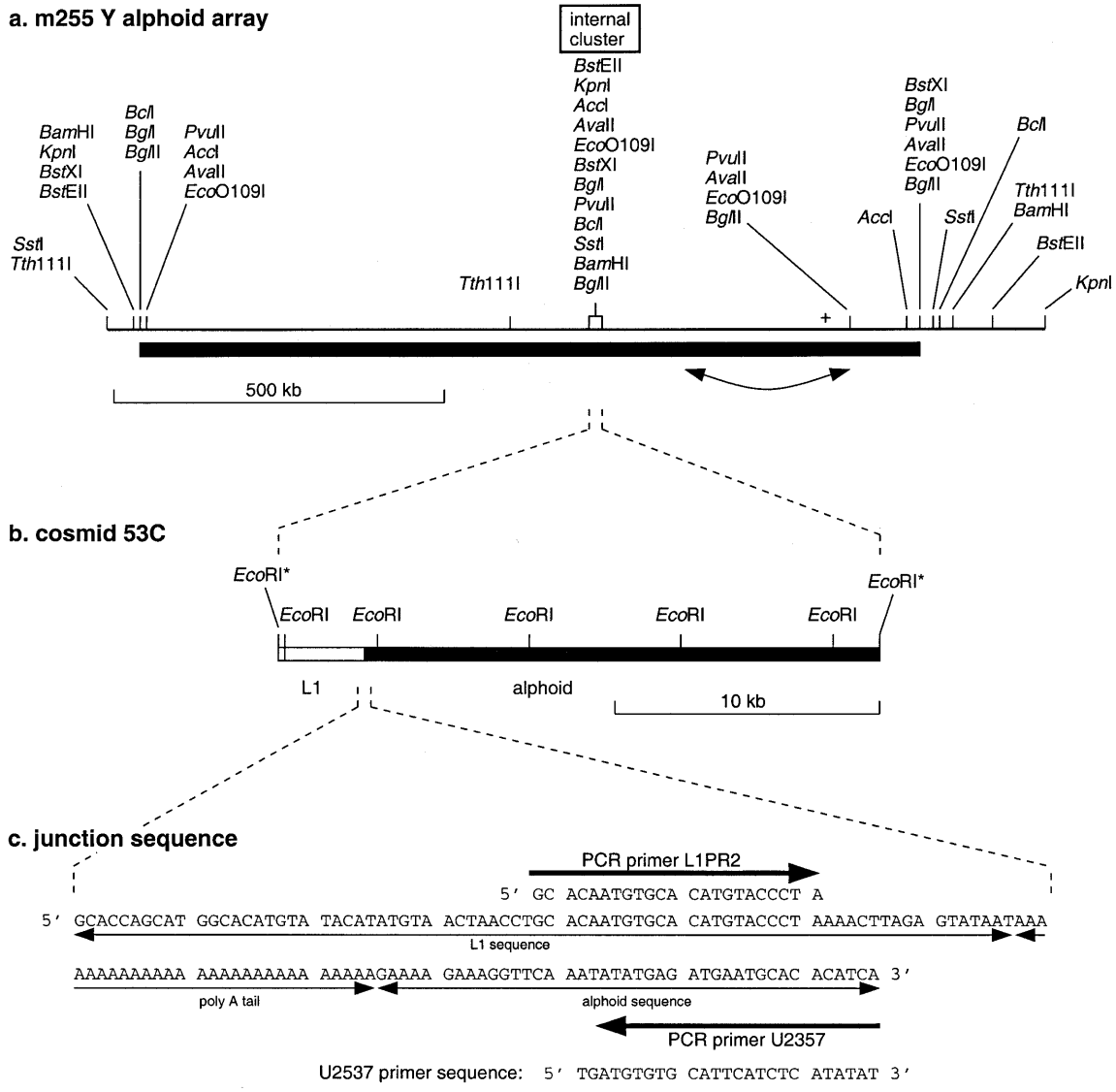

Figure 1. (**a**) Restriction site map of the m255 Y alphoid array. The black bar shows the approximate location of the alphoid array, the curved arrow points to the two possible positions for the small cluster of *Pvu*II, *Ava*II, *Eco*O109I and *Bgl*II sites, and the '+' indicates that an additional *Ava*II site was present but could not be precisely located. (**b**) Structure of cosmid 53C. L1 DNA is shown in white, alphoid DNA in black. *Eco*RI* sites were introduced by the cloning procedure. The orientation of the cosmid map with respect to the genomic map was deduced from the orientation of the alphoid DNA (27). (**c**) Sequence of the junction between the L1 and alphoid sequences. The primers used for PCR detection of the L1 insertion are shown.

size of the amplified LY1 band, but this has not been investigated further. Variation in the poly(A) tract from another retroposon insertion, YAP, has been observed previously (12). In a comparison of the PFGE and PCR assays, 15 males contained LY1 according to both assays and 497 lacked it. A single male showed the presence of LY1 according to the PCR assay but appeared to lack the internal cluster of restriction sites; on further analysis he was found to carry the cluster of sites precisely in the centre of the array so that the two fragments co-migrated on the pulsed-field gel. (The revised result has been incorporated into Table 1.) The PCR assay therefore provides a reliable and convenient way to score the presence or absence of LY1. Chromosomes carrying the L1 insertion are designated haplogroup 13.

In a total of 2311 males, 33 carried LY1 (1.4%) (Table 1). LY1 was absent from the African, European, Pacific and American individuals tested, but was present in several Asian populations (Fig. 3). It was found at highest frequency in China, where it represented 23% of the Han tested. It was also found at 11 and 26%, respectively, in smaller samples of two Chinese minority populations, the Tujia and Miao from Hunan. It was present at much lower frequencies (1–2%) in the surrounding geographical areas: Mongolia to the north, Japan to the east, Indonesia to the south and India to the west, producing a symmetrical geographical pattern centred on China (Fig. 3).

Diversity within the LY1 lineage

The geographical distribution seen in Figure 3 could arise if LY1 had inserted recently into a Chinese Y chromosome and the lineage had increased in frequency in the population, but had not spread very far. In order to investigate this scenario, we

Table 1. Geographical distribution of LY1-positive (+) and -negative (–) chromosomes

Continent	Population	PFGE		PCR		Total		*n*
		–	+	–	+	–	+	
Africa	San	7	0	7	0	7	0	7
	Kenya	14	0	15	0	15	0	15
	Bamileke (Cameroon)			40	0	40	0	40
	CAR Pygmy			22	0	22	0	22
	Ewondo (Cameroon)			8	0	8	0	8
	Ghana	1	0	6	0	6	0	6
	Algeria			33	0	33	0	33
	Egypt			12	0	12	0	12
	Other	7	0	12	0	12	0	12
Europe	Iceland	28	0	27	0	28	0	28
	Norway	10	0	53	0	53	0	53
	Saami			12	0	45	0	12
	Russia			28	0	28	0	28
	Mari			48	0	48	0	48
	Britain	21	0	26	0	26	0	26
	Basque			26	0	26	0	26
	Germany	1	0	13	0	13	0	13
	Italy	6	0	13	0	14	0	14
	Sardinia			17	0	17	0	17
	Hungary			37	0	37	0	37
	Other	35	0	26	0	40	0	40
Asia	Yakut	5	0	21	0	21	0	21
	Mongolia	60	1	60	1	60	1	61
	Buryat	4	0	105	1	105	1	106
	Pakistan	2	0	577	1	577	1	578
	India	316	1	214	1	316	1	317
	Sri Lanka	24	0	24	0	24	0	24
	Nepal	3	0	22	0	24	0	24
	China (Han)	38	14	58	17	58	17	75
	China (Miao)			17	6	17	6	23
	China (Tujia)			8	1	8	1	9
	Taiwan (Bunun)			10	0	10	0	10
	Taiwan (Paiwan)			14	0	14	0	14
	Taiwan (Atayal)			9	0	9	0	9
	Taiwan (Ami)			13	1	13	1	14
	Japan	3	0	143	2	143	2	145
	Indonesia (Alor)			20	0	20	0	20
	Indonesia (Palu)			17	0	17	0	17
	Indonesia (Bali)			9	0	9	0	9
	Indonesia (Manado)			22	0	22	0	22
	Indonesia (Medan)			12	0	12	0	12
	Indonesia (Padang)			17	0	17	0	17
	Indonesia (Pekanba)			11	1	11	1	12
	Indonesia (Borneo)			8	0	8	0	8
	Indonesia (other)			93	1[a]	93	1[a]	94
	Other	19	0	5	0	19	0	19
Pacific	PNG			43	0	43	0	43
	Tahiti			8	0	8	0	8
	Melanesia	3	0	3	0	3	0	3
	Australia	3	0	3	0	3	0	3
America	Inuit			25	0	25	0	25
	North			18	0	18	0	18
	Quechua			47	0	47	0	47
	Other	7	0	7	0	7	0	7
Totals		617	16	2132	33	2278	33	2311

[a]The one LY1-positive chromosome was the single sample from Toraja.

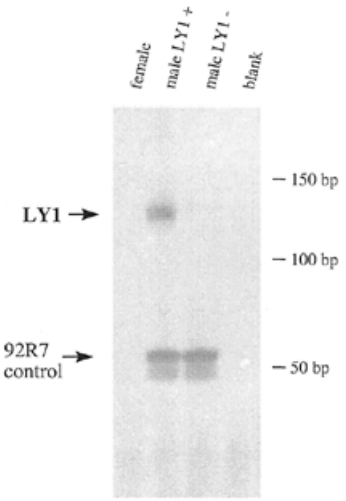

Figure 2. PCR detection of LY1. Samples were amplified with the 92R7 and LY1 primers, fractionated on a native polyacrylamide gel and silver stained. Samples are identified at the top of each track.

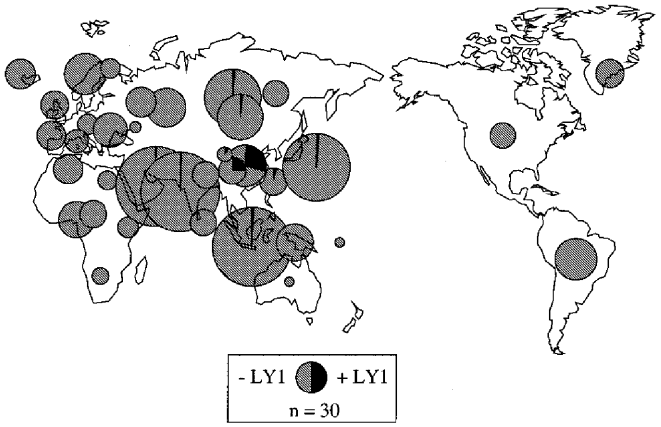

Figure 3. World-wide distribution of LY1-containing Y chromosomes. The circles represent the populations tested. Area is proportional to sample size (see $n = 30$ reference), except that the maximum size used is equivalent to 200 individuals. Some of the populations in Table 1 have been omitted or combined for clarity. Grey, samples lacking LY1; black, samples with LY1.

have analysed a set of seven Y microsatellites within the LY1 lineage. This provides a measure of the diversity within the lineage and allows us to compare the haplotypes of Chinese and non-Chinese LY1 chromosomes (Table 2). The resulting microsatellite haplotypes were used to construct a median-joining network (Fig. 4) (33). The network has several striking features. Most haplotypes are represented by only a single chromosome, and are usually separated by several steps, showing that they represent a diverse set of chromosomes. Chromosomes from each ethnic group within China or from each country are scattered within the network and do not cluster according to their origin.

Two dating methods were used to estimate the time to the most recent common ancestor (TMRCA) of these chromosomes incorporating the mutation rate of 0.0021 per locus per generation measured for this set of Y microsatellites (34). With a method relying on the variance being proportional to time (35), and an effective population size for males of ~4500 (36), our observed average variance per locus was 0.74, which is equivalent to 367 generations. Using a generation time of 25 years, we obtain ~9200 years; with 95% confidence limits for the variance, the range is between ~7100 and ~11 700, whereas with 95% confidence limits for the mutation rate [0.0006–0.0049 per locus per generation (34) as well], the range is ~3300 to ~59 000 years. The second method used the average number of mutation steps per locus from a likely root (37) as applied previously (10). For LY1 chromosomes, it was difficult to establish the root (Fig. 4), so we used an unobserved haplotype carrying the average repeat number at each locus. We obtained ~0.58 mutations on average from the likely root, which gives 275 generations, equivalent to ~6900 years if the generation time is 25 years. Using the 95% confidence intervals for the mutation rate, we obtained ~2900 and ~24 000 years. Because of the uncertainties in these estimates, we refer to the age as '~10 000 years'.

DISCUSSION

L1 elements may comprise as much as 17% of the human genome (38). Although most of the inserted copies are truncated and very old (39) some are known to be active in retrotransposition, but only a few polymorphic L1 retroposons have been described previously (30,40). Their use as polymorphic markers has not been extensively exploited. In addition, the centromeric position and Y-chromosomal location provide unique features of interest for LY1.

A centromeric L1 element

A considerable body of evidence supports the hypothesis that alphoid DNA is the normal human centromeric DNA sequence. However, entirely different DNA sequences can form neocentromeres (41), demonstrating that centromere function is not determined directly by the DNA sequence. It has been suggested that the last segment of DNA to replicate forms the centromere, and thus that the function of the satellite DNA is to delay replication (42). Retroposon insertions into a satellite array can potentially disrupt the normal pattern of replication. Retroposon locations within alphoid arrays have therefore been investigated extensively and all previously known elements have been found within diverged alphoid subsets or close to the ends of arrays (32,40). Our results now show that an L1 insertion within a major homogeneous alphoid array is compatible with normal centromere function. Perhaps only a short region of uninterrupted satellite is needed. However, the most likely evolutionary fate of such an element is loss during the normal satellite expansion/contraction processes, and their transience may explain why they have not been detected previously.

Dating the LY1 lineage

We assume that: (i) the insertion occurred only once, in agreement with all the available typing of the haplogroup 13 chromosomes with additional biallelic markers; and (ii) the ancestral state was the absence of the L1 element, so that the initial LY1-containing chromosome carried a single allele at each microsatellite locus. The microsatellites would subsequently have accumulated backward and forward slippage

Table 2. Microsatellite haplotypes of LY1-positive chromosomes

Sample	Origin	*DYS19*	*DYS389(D)*	*DYS389(A+B)*	*DYS390*	*DYS391*	*DYS392*	*DYS393*
m228	Mongolia	17	10	16	23	10	14	12
D25	Buryat	17	10	15	25	10	13	12
m39	China, Han	17	10	15	25	10	13	12
m255	China, Han	15	10	16	25	10	13	12
m259	China, Han	17	9	16	25	10	13	12
m299	China, Han	16	9	15	25	10	15	12
m301	China, Han	16	11	17	24	10	13	12
m310	China, Han	16	10	15	21	10	14	12
m312	China, Han	16	11	17	24	10	13	12
m313	China, Han	15	8	17	22	10	14	12
m315	China, Han	15	9	16	24	11	13	12
m323	China, Han	17	9	16	24	10	13	14
m324	China, Han	18	11	16	25	10	13	12
m327	China, Han	15	9	15	25	10	13	12
m329	China, Han	15	9	17	22	10	14	12
m330	China, Han	16	11	17	23	10	13	12
m332	China, Han	16	9	19	25	11	12	14
Ap33	China, Han	15	9	17	25	10	13	11
Ap43	China, Han	16	9	16	25	10	13	12
M19	China, Miao	16	10	15	24	11	13	14
M33	China, Miao	17	9	16	24	10	13	12
T1	China, Tujia	17	9	17	24	10	13	12
Ami3	Taiwan	15	10	18	23	10	13	14
J135	Japan	17	9	16	27	10	15	12
J141	Japan	16	10	15	25	11	13	12
Tor23	Indonesia	17	9	16	23	10	13	12
m547	India	17	9	15	24	10	13	12

mutations. One potential complication is that the L1 and surrounding alphoid DNA may be deleted during normal satellite DNA evolution. If a change in alphoid array size occurs in 1 in 20 meioses, and changes generally involve only one or two units (43), the one unit of ~150 carrying the L1 may be deleted in ~1 in 3000 meioses, an insignificant frequency.

We used two distinct dating methods for the TMRCA of the chromosomes carrying the LY1 insertion; the insertion itself will have occurred earlier than this. The two methods have given consistent results for dating three different Y lineages: haplogroup 16, defined by the Tat C allele (10), haplogroup 24, defined by the M4 G allele (44), and haplogroup 22, defined by the *SRY-2627* T allele (45). The most controversial parameter used is the mutation rate, which was calculated from meioses in deep-rooting pedigrees (34), using the same set of microsatellite loci. There may be problems related to the heterogeneity of rates among loci and alleles (46), and, for mtDNA, the mutation rates measured in pedigrees give more recent estimates than those measured indirectly through phylogenetic analysis (47). Thus the calculations may give consistent results but underestimate the true TMRCA.

Although our absolute dating for haplogroup 13 may be dubious, the insertion lineage is probably older than the other Y lineages analysed in this way (10,44,45). The comparison between distinct Y lineages, their microsatellite networks, dating and geographical distribution can still be informative for elucidating the history of haplogroup 13 chromosomes in China and the surrounding areas.

Chinese population history

LY1 was found in 23% of our Chinese Han sample. If the sample is representative of the entire Han population, this lineage will be found in ~6% of the world's Y chromosomes. Its microsatellite diversity suggests that the lineage arose ~10 000 years ago, in the Palaeolithic period. This high diversity, and correspondingly ancient age, linked to a limited geographical range, contrast with the patterns found in other Y lineages. Haplogroup 22 (45) has a restricted distribution within Europe and a low microsatellite diversity. Haplogroup 16 chromosomes (10) have a wide distribution spanning most of northern Asia and Europe, but still a lower microsatellite diversity than LY1 chromosomes.

This may reflect the larger population in China and/or a lower geographical mobility. The origin of the lineage is likely to predate the Neolithic in China, and diversification in relatively small populations during the Palaeolithic may account for the lack of a star-like network. LY1 now provides a useful marker for further investigating genetic diversity and population relationships within China, and the migrations of Chinese elsewhere. Indeed, the presence of a few LY1 chromosomes found in bordering countries, whose microsatellite haplotypes

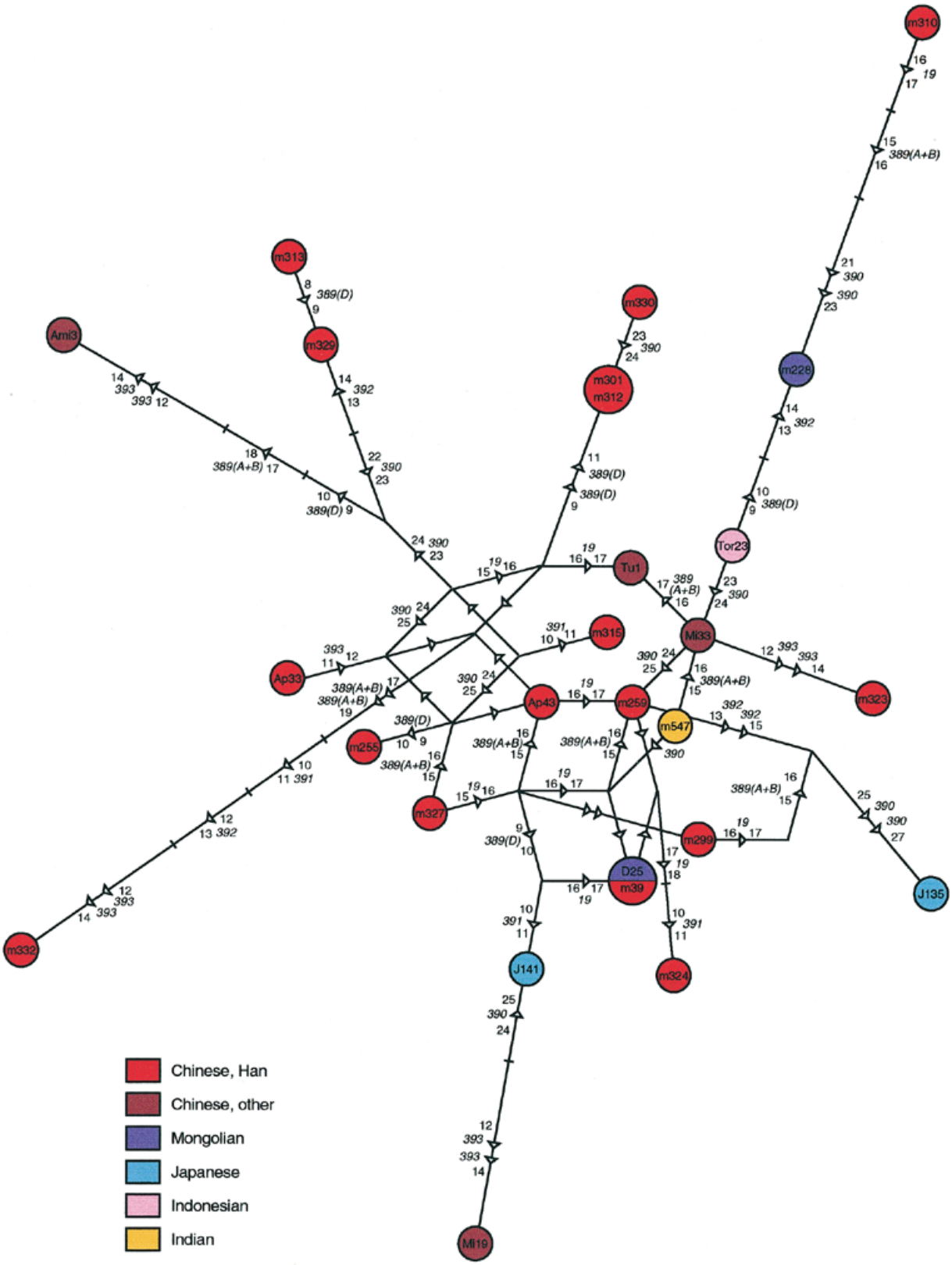

Figure 4. Median-joining network of microsatellite haplotypes of chromosomes carrying LY1. The circles represent the microsatellite haplotypes, with an area proportional to the number of individuals (one or two). The lines represent the microsatellite mutational steps; loci and repeat numbers are indicated.

were found scattered in the network (Fig. 4) indicate migrations from China.

MATERIALS AND METHODS

DNA samples

DNA samples are listed in Table 1. Most were derived from previously described collections of the authors (5,10,26,43,48–50). Additional samples were collected by the authors in Pakistan (R.Q., S.Q.M. and collaborators), China (A.L.), Nepal (R.A.) and Greenland (P.d.K.), or were generously provided by John Clegg (Taiwan, Indonesia, Tahiti), Rebecca Oakey (Africa), Eduardo M. Tarazona-Santos and Davide Pettener (Quechua), and Mark Stoneking and Alan Redd (PNG and Indonesia).

Mapping of the m255 alphoid array

Most procedures were as described previously (27). Single digests were carried out with the enzymes shown in Figure 1a and *Sfi*I, *Hpa*I and *Apa*I; the last three enzymes produced only a single alphoid fragment, showing that only a single alphoid array was present. Double digests with *Asp*I (an isoschizomer of *Tth*111I) were used to place some sites, but, since DNA stocks were limited, other sites were placed using the known maps from other arrays, or remained ambiguous as indicated.

Cosmid library construction and screening

Samples (2.5 µg) of genomic DNA from m255 were partially digested with dilutions of *Mbo*I for 30 min at 37°C. Suitably digested DNA was dephosphorylated and ligated to *Bam*HI-cleaved SuperCos arms, packaged with GigaPack II (Stratagene, La Jolla, CA) and incubated with *E.coli* XL1 Blue MR host cells (Stratagene). Infected cells were spread by suction onto a nylon membrane (20 × 20 cm) and grown at 37°C on an LB plate with ampicillin until colonies were visible, after which a replica was made and allowed to grow 4 h longer at 37°C. The replica filter was then denatured and hybridized with 50 ng of ^{32}P-labelled pYαI (5) at moderate stringency (75°C) overnight, washed three times with 0.1× SSC, 1% SDS at 65°C for 10 min and autoradiographed.

Approximately 250 000 colonies (~2.5 genome equivalents) were screened and 145 Y alphoid positive colonies were isolated for further analysis. They were digested with *Eco*RI, which cuts once per unit, or double digested with *Eco*RI plus *Kpn*I, *Sac*I, *Pvu*II or *Ava*II, which do not cut within the Y alphoid unit, but do cut within the variant region, and separated by agarose gel electrophoresis. Cosmids likely to contain non-alphoid DNA linked to typical alphoid repeats were transferred to a membrane and probed again with ^{32}P-labelled pYαI to allow the detection of bands containing alphoid and non-alphoid DNA.

Subcloning and sequencing

A single cosmid clone (53C) was selected for further analysis. A non-alphoid *Eco*RI fragment was purified by GeneClean (Bio 101, Vista, CA) and ligated into pTZ18R to generate subclones 53Ca5 and 53Ca7. The subclones were either sequenced manually using the T7 Sequenase kit (USB, Cleveland, OH), or the CycleSequencing kit (Amersham Pharmacia Biotech, Little Chalfont, UK). Reactions from the latter were run on an ALF DNA sequencer (Amersham Pharmacia Biotech). The sequence was deposited in GenBank under accession no. AF189307.

Subsequently, a specific PCR subcloning strategy was designed to allow the cloning of the 3′ junction between the L1 element and the alphoid DNA of the cosmid 53C. The L1–alphoid junction was amplified using the primer L1α1, 5′-aca cat tag tgg gtg cag cgc acc-3′, designed from the consensus 3′ end sequence of L1 elements of the Ta subset (30), and a general alphoid primer YαR, 5′-ttt gga gcc ctt tga ggc cta ttg-3′, based on the consensus sequence of Y alphoid 171 bp subunits (25). Amplified bands were inserted into pUC18 using the SureClone kit (Amersham Pharmacia Biotech). Two PCR subclones called 53Cj1 and 53Cj2 were sequenced as described above. The sequence was deposited in GenBank under accession no. AF207859.

PCR detection of the L1 insertion

For typing of males for the presence or absence of LY1, we designed more specific primers which allow amplification of the L1–alphoid junction from genomic DNA. As a positive control for amplification, we amplified another Y-chromosomal locus named 92R7 (45) in the same tube. The primers specific for the L1–alphoid junction were: L1PR2, 5′-gca caa tgt gca cat gta ccc ta-3′, and U2357, 5′-tga tgt gtg cat tca tct cat ata t-3′, and the primers for amplification of the 55 bp 92R7 locus were 92R7a, 5′-tgc atg aac aca aaa gac gta-3′, and 92R7b, 5′-gca ttg tta aat atg acc agc-3′. Each 12.5 µl PCR tube contained 200 µM dNTPs, 1.5 mM $MgCl_2$, 1 µM primers L1PR2 and U2357, 0.1 µM control primers for the 92R7 locus, 5–50 ng of DNA template and 1 IU/tube of Ampli-Taq Gold DNA polymerase and 1× buffer (both from Perkin Elmer, Warrington, UK). The cycling conditions were: a first denaturation step at 94°C for 11 min, followed by 35 cycles with steps of 60°C for 30 s, 72°C for 45 s and 94°C for 20 s. After amplification, the samples were separated in an 18 cm native polyacrylamide gel in 1× TBE buffer for 1 h at 150 V and silver stained as described previously (51), or run on a 2% NuSieve, 1% agarose gel and stained with ethidium bromide.

Microsatellite analysis

Twenty-seven males bearing the L1 insertion were further analysed for the seven Y chromosome microsatellite loci *DYS19*, *DYS389(D)*, *DYS389(A+B)*, *DYS390*, *DYS391*, *DYS392* and *DYS393* by automated allele typing on an ALF*express* DNA sequencer (Amersham Pharmacia Biotech) or ABI 377 (Applied Biosystems, Foster City, CA) as previously described (15). Haplotypes consisting of the allelic state (repeat number) at each microsatellite locus were used for phylogeny reconstruction with the median network program Network1.1 (33). For dating analysis we used either the average repeat number differences between haplotypes and a likely root (37) as applied by Zerjal *et al.* (10), or the average variance for all microsatellite loci proportional to time (35).

ACKNOWLEDGEMENTS

We thank Ed Southern for stimulating discussions and suggestions about the experiments, Hans Bandelt for helpful discus-

sions, all DNA donors and John Clegg, Eduardo M. Tarazona-Santos, Davide Pettener, Rebecca Oakey, Mark Stoneking and Alan Redd for providing samples, and Gerard Roizes and Mark Jobling for comments on the text. F.R.S. was supported by the Leverhulme Trust, A.P. by the BBSRC, M.K. by the Deutsche Forschungsgemeinschaft, R.A. by a Wellcome Trust vacation grant and C.T.-S. by the C.R.C.

REFERENCES

1. Jobling, M.A. and Tyler-Smith, C. (1995) Fathers and sons: the Y chromosome and human evolution. *Trends Genet.*, **11**, 449–456.
2. Mitchell, R.J. and Hammer, M.F. (1996) Human evolution and the Y chromosome. *Curr. Opin. Genet. Dev.*, **6**, 737–742.
3. Harris, P., Boyd, E., Young, B.D. and Ferguson-Smith, M.A. (1986) Determination of the DNA content of human chromosomes by flow cytometry. *Cytogenet. Cell Genet.*, **41**, 14–21.
4. Malaspina, P., Persichetti, F., Novelletto, A., Iodice, C., Terrenato, L., Wolfe, J., Ferraro, M. and Prantera, G. (1990) The human Y chromosome shows a low level of DNA polymorphism. *Ann. Hum. Genet.*, **54**, 297–305.
5. Mathias, N., Bayes, M. and Tyler-Smith, C. (1994) Highly informative compound haplotypes for the human Y chromosome. *Hum. Mol. Genet.*, **3**, 115–123.
6. Seielstad, M.T., Hebert, J.M., Lin, A.A., Underhill, P.A., Ibrahim, M., Vollrath, D. and Cavalli-Sforza, L.L. (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet.*, **3**, 2159–2161.
7. Hammer, M.F. (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. *Mol. Biol. Evol.*, **11**, 749–761.
8. Santos, F.R., Pena, S.D. and Tyler-Smith, C. (1995) PCR haplotypes for the human Y chromosome based on alphoid satellite DNA variants and heteroduplex analysis. *Gene*, **165**, 191–198.
9. Underhill, P.A., Jin, L., Zemans, R., Oefner, P.J. and Cavalli-Sforza, L.L. (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl Acad. Sci. USA*, **93**, 196–200.
10. Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F.R., Schiefenhovel, W., Fretwell, N., Jobling, M.A., Harihara, S. *et al.* (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.*, **60**, 1174–1183.
11. Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L. and Oefner, P.J. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.*, **7**, 996–1005.
12. Hammer, M.F., Spurdle, A.B., Karafet, T., Bonner, M.R., Wood, E.T., Novelletto, A., Malaspina, P., Mitchell, R.J., Horai, S., Jenkins, T. and Zegura, S.L. (1997) The geographic distribution of human Y chromosome variation. *Genetics*, **145**, 787–805.
13. Roewer, L., Arnemann, J., Spurr, N.K., Grzeschik, K.H. and Epplen, J.T. (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum. Genet.*, **89**, 389–394.
14. Santos, F.R., Pena, S.D. and Epplen, J.T. (1993) Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. *Hum. Genet.*, **90**, 655–656.
15. Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M. *et al.* (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal Med.*, **110**, 125–133.
16. Scozzari, R., Cruciani, F., Malaspina, P., Santolamazza, P., Ciminelli, B.M., Torroni, A., Modiano, D., Wallace, D.C., Kidd, K.K., Olckers, A. *et al.* (1997) Differential structuring of human populations for homologous X and Y microsatellite loci. *Am. J. Hum. Genet.*, **61**, 719–733.
17. Jobling, M.A., Williams, G.A., Schiebel, G.A., Pandya, G.A., McElreavey, G.A., Salas, G.A., Rappold, G.A., Affara, N.A. and Tyler-Smith, C. (1998) A selective difference between human Y-chromosomal DNA haplotypes. *Curr. Biol.*, **8**, 1391–1394.
18. Ruiz Linares, A., Nayar, K., Goldstein, D.B., Hebert, J.M., Seielstad, M.T., Underhill, P.A., Lin, A.A., Feldman, M.W. and Cavalli Sforza, L.L. (1996) Geographic clustering of human Y-chromosome haplotypes. *Ann. Hum. Genet.*, **60**, 401–418.
19. Poloni, E.S., Semino, O., Passarino, G., Santachiara-Benerecetti, A.S., Dupanloup, I., Langaney, A. and Excoffier, L. (1997) Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am. J. Hum. Genet.*, **61**, 1015–1035.
20. Santos, F.R., Pandya, A., Tyler-Smith, C., Pena, S.D., Schanfield, M., Leonard, W.R., Osipova, L., Crawford, M.H. and Mitchell, R.J. (1999) The central Siberian origin for native American Y chromosomes. *Am. J. Hum. Genet.*, **64**, 619–628.
21. Stoneking, M., Fontius, J.J., Clifford, S.L., Soodyall, H., Arcot, S.S., Saha, N., Jenkins, T., Tahir, M.A., Deininger, P.L. and Batzer, M.A. (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.*, **7**, 1061–1071.
22. Nikaido, M., Rooney, A.P. and Okada, N. (1999) Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl Acad. Sci. USA*, **96**, 10261–10266.
23. Altheide, T.K. and Hammer, M.F. (1997) Evidence for a possible Asian origin of YAP+ Y chromosomes. *Am. J. Hum. Genet.*, **61**, 462–466.
24. Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R. and Zegura, S.L. (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.*, **15**, 427–441.
25. Tyler-Smith, C. and Brown, W.R. (1987) Structure of the major block of alphoid satellite DNA on the human Y chromosome. *J. Mol. Biol.*, **195**, 457–470.
26. Oakey, R. and Tyler-Smith, C. (1990) Y chromosome DNA haplotyping suggests that most European and Asian men are descended from one of two males. *Genomics*, **7**, 325–330.
27. Cooper, K.F., Fisher, R.B. and Tyler-Smith, C. (1993) Structure of the sequences adjacent to the centromeric alphoid satellite DNA array on the human Y chromosome. *J. Mol. Biol.*, **230**, 787–799.
28. Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, C.T., O'Hara, B., Rossiter, J.P., Cooley, T., Heath, P., Smith, K.D. and Margolet, L. (1987) Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics*, **1**, 113–125.
29. Skowronski, J., Fanning, T.G. and Singer, M.F. (1988) Unit-length LINE-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.*, **8**, 1385–1397.
30. Sassaman, D.M., Dombroski, B.A., Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J., Gabriel, A., Swergold, G.D. and Kazazian Jr, H.H. (1997) Many human L1 elements are capable of retrotransposition. *Nature Genet.*, **16**, 37–43.
31. Boeke, J.D. (1997) LINEs and Alus—the polyA connection. *Nature Genet.*, **16**, 6–7.
32. Laurent, A.M., Puechberty, J. and Roizes, G. (1999) Hypothesis: for the worst and for the best, L1Hs retrotransposons actively participate in the evolution of the human centromeric alphoid sequences. *Chromosome Res.*, **7**, 305–317.
33. Bandelt, H.J., Forster, P. and Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, **16**, 37–48.
34. Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E. and de Knijff, P. (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum. Mol. Genet.*, **6**, 799–803.
35. Goldstein, D.B., Zhivotovsky, L.A., Nayar, K., Linares, A.R., Cavalli-Sforza, L.L. and Feldman, M.W. (1996) Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes. *Mol. Biol. Evol.*, **13**, 1213–1218.
36. Hammer, M.F. (1995) A recent common ancestry for human Y chromosomes. *Nature*, **378**, 376–378.
37. Bertranpetit, J. and Calafell, F. (1996) Genetic and geographical variability in cystic fibrosis: evolutionary considerations. In Chadwick, D. and Cardew, G. (eds), *Ciba Foundation Symposium, 197*. Wiley, Chichester, New York, pp. 97–118.
38. Smit, A.F. (1996) The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.*, **6**, 743–748.
39. Smit, A.F., Toth, G., Riggs, A.D. and Jurka, J. (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.*, **246**, 401–417.
40. Prades, C., Laurent, A.M., Puechberty, J., Yurov, Y. and Roizes, G. (1996) SINE and LINE within human centromeres. *J. Mol. Evol.*, **42**, 37–43.
41. du Sart, D., Cancilla, M.R., Earle, E., Mao, J.I., Saffery, R., Tainton, K.M., Kalitsis, P., Martyn, J., Barry, A.E. and Choo, K.H. (1997) A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. *Nature Genet.*, **16**, 144–153.
42. Csink, A.K. and Henikoff, S. (1998) Something from nothing: the evolution and utility of satellite repeats. *Trends Genet.*, **14**, 200–204.

43. Mathias, N. (1993) *Y chromosome DNA polymorphisms and human evolution*. DPhil thesis, Oxford University, Oxford.
44. Hurles, M.E., Irven, C., Nicholson, J., Taylor, P.G., Santos, F.R., Loughlin, J., Jobling, M.A. and Sykes, B.C. (1998) European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. *Am. J. Hum. Genet.*, **63**, 1793–1806.
45. Hurles, M.E., Veitia, R., Arroyo, E., Armenteros, M., Bertranpetit, J., Pérez-Lezaun, A., Bosch, E., Shlumukova, M., Cambon-Thomsen, A., McElreavey, K. *et al.* (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am. J. Hum. Genet.*, **65**, 1437–1448.
46. Carvalho-Silva, D.R., Santos, F.R., Hutz, M.H., Salzano, F.M. and Pena, S.D. (1999) Divergent human Y-chromosome microsatellite evolution rates. *J. Mol. Evol.*, **49**, 204–214.
47. Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R., Berry, D.L., Holland, K.A., Weedn, V.W., Gill, P. and Holland, M.M. (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genet.*, **15**, 363–368.
48. Spedini, G., Destro-Bisol, G., Mondovi, S., Kaptue, L., Taglioli, L. and Paoli, G. (1999) The peopling of sub-Saharan Africa: the case study of Cameroon. *Am. J. Phys. Anthropol.*, **110**, 143–162.
49. Ciminelli, B.M., Pompei, F., Malaspina, P., Hammer, M., Persichetti, F., Pignatti, P.F., Palena, A., Anagnou, N., Guanti, G., Jodice, C. *et al*. (1995) Recurrent simple tandem repeat mutations during human Y-chromosome radiation in Caucasian subpopulations. *J. Mol. Evol.*, **41**, 966–973.
50. Malaspina, P., Cruciani, F., Ciminelli, B.M., Terrenato, L., Santolamazza, P., Alonso, A., Banyko, J., Brdicka, R., Garcia, O., Gaudiano, C. *et al*. (1998) Network analyses of Y-chromosomal types in Europe, northern Africa, and western Asia reveal specific patterns of geographic distribution. *Am. J. Hum. Genet.*, **63**, 847–860.
51. Santos, F.R., Bianchi, N.O. and Pena, S.D. (1996) Worldwide distribution of human Y-chromosome haplotypes. *Genome Res.*, **6**, 601–611.