Genetic characterization of the body attributed to the evangelist Luke

Cristiano Vernesi*, Giulietta Di Benedetto*, David Caramelli†, Erica Secchieri*, Lucia Simoni‡, Emile Katti*, Patrizia Malaspina§, Andrea Novelletto¶, Vito Terribile Wiel Marin‖, and Guido Barbujani*,**

*Department of Biology, University of Ferrara, Via Borsari 46, 44100 Ferrara, Italy; †Institute of Anthropology, University of Florence, 50125 Florence, Italy; ‡Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, 1227 Geneva, Switzerland; §Department of Biology, University of Rome 2, Tor Vergata, 00133 Rome, Italy; ¶Department of Biology, University of Calabria, 87030 Arcavacata di Rende, Italy; and ‖Institute of Pathological Anatomy, University of Padua, 35121 Padua, Italy

Historical sources indicate that the evangelist Luke was born in Syria, died in Greece, and then his body was transferred to Constantinople, and from there to Padua, Italy. To understand whether there is any biological evidence supporting a Syrian origin of the Padua body traditionally attributed to Luke, or a replacement in Greece or Turkey, the mtDNA was extracted from two teeth and its control region was cloned and typed. The sequence determined in multiple clones is an uncommon variant of a set of alleles that are common in the Mediterranean region. We also collected and typed modern samples from Syria and Greece. By comparison with these population samples, and with samples from Anatolia that were already available in the literature, we could reject the hypothesis that the body belonged to a Greek, rather than a Syrian, individual. However, the probability of an origin in the area of modern Turkey was only insignificantly lower than the probability of a Syrian origin. The genetic evidence is therefore compatible with the possibility that the body comes from Syria, but also with its replacement in Constantinople.

According to historical sources, the evangelist Luke was born in Antioch, in the Roman province of Syria, and died in Thebes (Greece) at age 84, around *anno Domini* (A.D.) 150 (1). His body was initially buried in Thebes, but then it was transferred to Constantinople in the second year of the reign of emperor Constantius (A.D. 338) and eventually to Padua, Italy, at an unspecified time before 1177, possibly under the reign of Julian the Apostate (361–363), or during the iconoclast controversy (726–846; Fig. 1; ref. 2). The availability of samples from the Padua body (PB) for DNA typing provides an opportunity to test the degree to which molecular biology could confirm history by assigning the remains of a single individual to a geographic region of origin.

The marble sarcophagus containing the body traditionally attributed to Luke was opened September 17, 1998, by a committee headed by V.T.W.M. It contained a leaden coffin, whose size fits into the tomb considered to be Luke's, in Thebes. Inside, along with various objects (including two plates with the dates 1463 and 1562, when the coffin was last opened), was the skeleton of a male individual. Evident signs of osteoporosis and skeletal deformations showed that that individual had died at age 70 or more. The pelvis had fused with the coffin. Complementary marks of larvae of the saprophage *Diptera*, on the pelvis and on the lead, suggest that the body decomposed in that coffin.

Materials and Methods

Ancient DNA. To minimize the risk of contamination, we followed the guidelines defined in ref. 3. The entire process of DNA extraction and PCR amplification was carried out in a dedicated room, where no modern, amplified, or cloned DNA was ever taken. Strict cleaning criteria were followed, including frequent treatment with bleach and UV light, and there were negative controls throughout the entire procedure. The root surfaces of both teeth was removed, the whole root was washed in 10% sodium hypoclorite for 10 seconds, and left 1 hour under UV light. Once dried, the cleaned root was powdered in a grinding mill, and DNA was extracted by using a silica-based protocol (modified from ref. 4). Because ancient DNA molecules tend to be damaged, after DNA extraction a 360-bp segment of the hypervariable region I of the mitochondrial genome was typed by amplifying 5 short overlapping fragments, hereafter designated as amplicons 1–5. A total of 39 sequences were obtained (31 initially, plus 8 of amplicon 3 in a second round of experiments). Amplifications were carried out by using primer pairs described in ref. 4 [(except for primer H16331, which we designed as 5′-TTGACTGTAATGTGCTATGT-3′) amplicon 1, L16055-H16139; amplicon 2, L16131-H16218; amplicon 3, L16131-H16379; amplicon 4, L16209-H16331; and amplicon 5, L16287-H16410]. Amplicon lengths can be obtained as the difference between nucleotide positions (for more details about the methods, see refs. 5 and 6). If primer dimers or nonspecific bands were not visible after gel electrophoresis, 4–6 μl of the amplification volume were directly ligated in a pCR2.1 vector (Original TA Cloning, Invitrogen) for 16 h at 16°C. *Escherichia coli* JM109 were transformed, and the clones were sequenced (6).

Of the nine criteria listed by Cooper and Poinar (3), we could comply with seven, namely: (*i*) a physically isolated work area, (*ii*) control amplifications, (*iii*) appropriate molecular behavior, (*iv*) reproducibility, (*v*) cloning, (*vii*) biochemical preservation, and (*ix*) study of associated faunal and floral remains, which is not of interest here. Because of the extreme paucity of the sample, however, replication of the sequencing in independent extracts (criteria *vi* and *viii*) proved impossible.

Modern Samples. To identify the most likely geographic origin of the PB, we collected samples of modern Greeks and Syrians. We typed 48 Greeks, 30 of them from Attica and 18 from Crete, and 49 Arabic-speaking northern Syrians. While no evidence suggests a discontinuity between the ancient and contemporary Greek gene pools, it is not obvious which modern sample is best suited to represent the population of which Luke was originally part. After Syria ceded Antioch to Turkey in 1939, immigration has been extensive there. Also, southeastern Turkey is home to a large community of Kurds, an ethnic group historically and linguistically differentiated from both Turks and Syrians, which is not officially recognized by the Turkish authorities. It was therefore impossible to obtain a sample from that area, with a reasonable certainty that it did not include members of the Kurdish community. We thus judged that the best possible comparison would be with people of northern Syria, who were sampled around Aleppo, in a radius of less
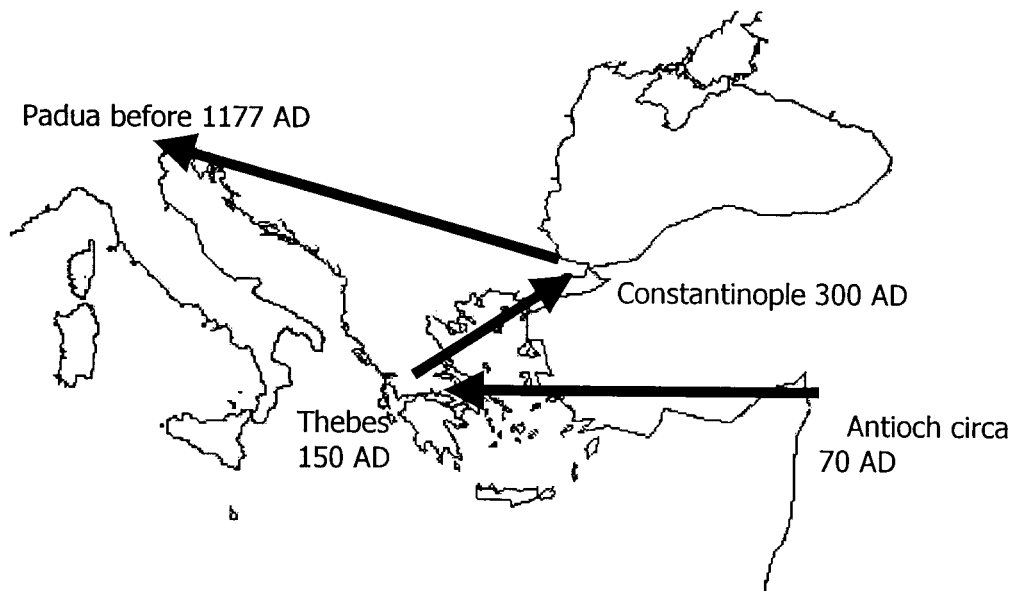
Fig. 1. Possible itinerary of the evangelist Luke and of his body. Dates are based on ref. 2.

than 100 km from Antioch (Fig. 5, which is published as supporting information on the PNAS web site, www.pnas.org.

In the time period of interest, Anatolia and Constantinople were not inhabited by Turkic speakers. However, the introduction of Turkic at the bout of this millennium was probably accompanied by limited genetic change (7). As a consequence, it seemed safe to assume that the modern population closest to the ancient Anatolian population is the one dwelling in the same area, for which 96 sequences were already available in the literature (8, 9).

Data Analysis. Differences between populations were represented in two dimensions by means of multidimensional scaling, a nonparametric method of multivariate ordination (10). Different measures of genetic distance gave highly concordant results, and here we report the results based on Tamura's index (11).

Mismatch distributions were estimated by using the DNASP3 (12) and ARLEQUIN 2.000 (13) packages. Odds ratios comparing alternative hypotheses on the origin of the PB were calculated, and a randomization procedure was devised to test whether they significantly differ from the value, 1, expected when the alternative hypotheses are equally likely. For that purpose, 9,000 pseudosamples were generated by randomly extracting (3,000 times) with replacement 49 sequences from the Syrian sample, 48 sequences from the Greek sample, and 96 sequences from the Turkish sample. Mismatch distributions were calculated from those pseudosamples, and pseudoodds ratios were then estimated for 3,000 comparisons of Syria with Greece and 3,000 comparisons of Syria with Turkey. From the distributions of these pseudovalues, we quantified empirical confidence intervals about the observed odds ratios.

Results

We obtained a whole canine tooth, a tooth root (Fig. 2), and 0.071 g of the bone powder generated during the removal of the femur fragment that was used, in other laboratories, for radiocarbon dating. The bone powder was hydrolyzed and analyzed by RP-HPLC (14) to estimate the degree of amino acid racemization, an indirect measure of the state of preservation of the sample's macromolecules. The aspartic acid D/L ratio was 0.06, suggesting a good probability (15) that amplifiable endogenous DNA could be retrieved from the teeth.

Only the whole tooth yielded sufficient amounts of DNA for the successive steps of the analysis; other biological material proved impossible to obtain. The 5 amplicons were cloned, and from 5 to 13 clones for each amplicon (39 in total) were sequenced.

As is common with ancient DNA (4, 5), nonreproducible substitutions were observed. However, almost all of the substitutions occurred only in single clones, and thus they were regarded as the results of occasional misincorporation of nucleotides during PCR; the endogenous sequence could be inferred with a good level of confidence by comparing clones. Two ambiguities remained: (*i*) In position 16319, a G–A substitution was observed in some sequences of amplicon 2 (sequenced between 16287 and 16410), but not in any of the sequences of amplicons 1 (between 16131 and 16379) and 3 (between 16209 and 16331); (*ii*) In position 16291, a C–T substitution was observed in all 5 clones of amplicon 1 and in 9 of the 13 clones of amplicon 3, but not in any clones of amplicon 2. No laboratory member carries in her/his mtDNA either substitution. Aside from the sequences obtained from amplicon 2, none of the 18 relevant replicates (amplicons 1 + 3) shows the 16319 substitution, and 14 show the 16291 substitution. The most parsimonious conclusion is that the results of amplification 2 seem affected by the presence of a contaminating competitor DNA, and that the PB sequence differs from the Cambridge reference sequence (16) for two transitions, at sites 16235 and 16291.

The sequence thus determined belongs to the a broad cluster of evolutionarily related alleles, comprising haplogroups V and H, the latter being the most common mitochondrial haplogroup in Europe and in the Levant (17). To characterize the PB further, we then reamplified its mtDNA four times with primers L6909-H 7115, and performed an *Alu*I restriction cut at position 7025 (18). The presence of the restriction site in all replicates shows that the PB mtDNA falls in what has been termed "pre HV" cluster (19). However, because very few European samples have been typed at the restriction fragment length polymorphism (RFLP) level, in successive steps of the numerical analysis we could consider only the sequence information. In a database of 2,819 Europeans (20), we found two identical sequences, in a Spaniard and in a Basque. Eleven other individuals show only the substitution at site 16291, whereas no known European mtDNAs contain the substitution at site 16235 alone. Also, the 16235-16291-16319

Fig. 3. Synthetic representation of the relationships among 12 samples and the PB sequence. Sy, Syria; Gr, Greece (present study). The other 10 samples (with their sizes in parentheses) are Al, Albanians (42); Ba, Basques (106); Dr, Druzes from Israel (45); Is, Southern Italians (37); It, Italians from Tuscany (49); Ne, Arabs from the Near East (42); Sa, Sardinians (73); Si, Sicilians (63); Sp, Spaniards (74); Tu, Turks (96).
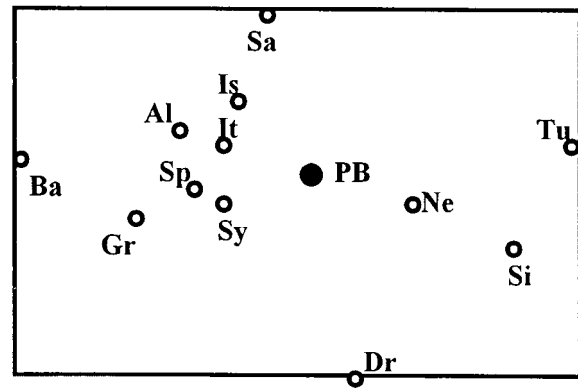
Fig. 2. The two teeth analyzed. Two laboratories (University of Arizona, Tucson, and Oxford University) independently estimated that the PB belonged to a person who died between A.D. 72 and A.D. 416 (95% confidence interval; G. Molin and V.T.W.M., personal communication).

motif is not present in the European database (20). The PB sequence is thus an uncommon variant of a set of alleles that is widespread in Europe and along the Mediterranean coasts.

One Syrian sequence shows an insertion between positions 16252 and 16253. All other sequences differ from each other, and from the PB sequence, for variable numbers of substitutions, all of them transitions. Representing the data as a network (21), or as a neighbor-joining tree (22), did not clarify their relationships (data not given). Both methods showed only the PB sequence close to an unstructured cluster, in which Greeks and Syrians were equally represented.

Then the PB, Syrian, and Greek sequences were compared with 10 population samples of the Mediterranean region (20). In a two-dimensional representation (10) of their genetic relationships (Fig. 3), the PB sequence falls close to the samples from central Italy, Syria, and the Near East. The Spanish and Greek samples are further apart, and the Basque sample is even more distant. The Arabic-speaking populations dwelling in the Levant, Druzes and Near Eastern Arabs, appear differentiated from Syria; this does not suggest that the composition of the Syrian sample of this study has been dramatically affected by the demographic changes associated with the diffusion of Arabic speakers in the seventh and eighth centuries A.D. (23). Although Eastern populations tend to fall on the right of the graph, and Western populations on the left, there is little geographical structuring, as shown by previous studies of mtDNA (20). Based

on this analysis, it would be difficult to attribute the PB sequence to a specific area with any degree of confidence.

We then resorted to an odds-ratio test, a statistical method that has been largely used in epidemiology to assign individuals to different risk classes, on the basis of a risk factor (24). The mean difference between pairs of sequences (MPSD) was 5.09 between Syria and Greece, 5.00 between PB and Syria, and 4.35 between PB and Greece (Fig. 4). However, the Syrian sample was internally more variable than the Greek one (MPSD = 5.76 and 4.47, respectively), so that the PB occupies a more eccentric position in the distribution of the Greek, than of the Syrian, sequences. Then
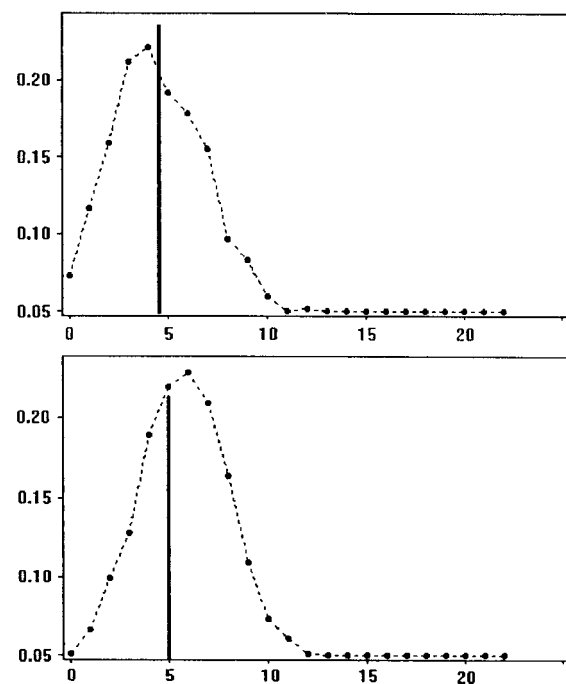
Fig. 4. Distributions of pairwise sequence differences (broken lines) within the Greek (*Top*) and the Syrian (*Bottom*) samples. *x* axis, pairwise sequence difference; *y* axis, frequency of that difference. The thick vertical lines indicate d_{PB}, the average difference between the sequences of either population sample and the PB sequence.

Table 1. Distribution of the pairwise sequence differences in the Syrian, Greek, and Turkish samples with respect to d_{PB}, the average *PSD* between the PB and each sample

Sample	$PSD \geq d_{PB}$	$PSD < d_{PB}$	Total comparisons
Syria ($n = 49$)	842	334	1,176
Greece ($n = 48$)	528	600	1,128
Turkey ($n = 96$)	2,878	1,682	4,560

n = sample size.

what is the relative probability that the PB sequence belongs to the Syrian or to the Greek gene pool, given the MPSDs observed between the PB and the two samples? This probability can be estimated as an odds ratio, ω (24), and is equivalent to the relative risk of having certain symptoms (here, coming from the Syrian or from the Greek gene pool) given a certain risk factor (here, an MPSD equal to, or greater than, that observed between the PB sequence and either population sample, d_{PB}). We thus estimated the probability to belong to either sample, for a sequence of unknown origin, which differs from the others at least as much as the PB sequence does. The calculations were then repeated for the comparison between Syria and Turkey.

The relative risk that an unknown sequence with $PSD \geq d_{PB}$ be of Syrian and not Greek origin is $\omega_{SG} = (842 \times 600)/(528 \times 334) = 2.87$ (Table 1). The empirical 95% confidence interval, estimated through 3,000 randomizations, is 1.06–6.17. Accordingly, the observed ω_{SG} is significantly higher than the value, 1, expected if the alternative origins are equally likely, as also shown by the fact that 98.2% of the random odds ratios are >1. The hypothesis of a Greek (and not Syrian) origin can therefore be rejected with an empirical probability of Type I error = 1.8%. For the comparison between Syria and Turkey, on the other hand, $\omega_{ST} = 1.41$, with 95% confidence interval 0.66–2.88 and 80.5% of the random odds ratios >1. The hypothesis of a Turkish (and not Syrian) origin, although less likely than its alternative, is therefore compatible with the data, with an empirical probability = 19.5%.

Discussion

The largest fraction of human diversity is found within populations. Less than 10% of the total DNA variance of our species occurs between populations of the same continent (25–27), whereas individual differences between members of the same community account for more than 80% of the total mitochondrial diversity (28). Therefore, it comes as no surprise that mitochondrial alleles equal or related to the PB sequence have been observed over much of Europe. However, the population samples that appear more closely related to the PB sequence in Fig. 3 are those from the Near East and Syria.

A better genetic characterization of the PB seems extremely difficult. For personal identification purposes, forensic scientists use DNA profiles based on several genes (29), but mtDNA seems at present the only marker that can be typed from ancient human remains without a high risk of errors caused by amplification of exogenous DNA (30).

Given that limitation, the evidence available, which is unlikely to be expanded further, indicates that the PB remains have a nearly 3-fold higher probability to come from a Syrian than from a Greek individual. Any error we may have made in the choice of the Syrian reference sample has probably obscured the relationship between it and the PB sequence, and therefore has acted against attribution of the PB sequence to the Syrian gene pool. Therefore, genetic data indicate that replacement of the body of the evangelist Luke with the body of a Greek individual is unlikely. On the other hand, the PB sequence also has a higher probability to be of Syrian, rather than Turkish, origin but that difference has an almost 20% chance to be simply the result of chance. Anatolia and Syria are geographically close and, along the coasts, they are not separated by major physical barriers. Therefore, under a simple model of isolation by distance, they are not expected to differ much genetically. In addition, estimates of low Central Asian admixture in Anatolia (7) suggest that recent demographic changes may have caused only limited divergence between the Syrian and Turkish gene pools. Both factors could explain why assigning the PB to either population proved complicated. At any rate, given the mitochondrial sequence determined, a Syrian origin of the PB is the most likely, but replacement of the body in Constantinople, which is still compatible with the upper confidence limit of his radiocarbon-estimated age (V.T.W.M. and G. Molin, personal communication; see legend to Fig. 2), cannot be ruled out.

1. Munck, J. (1967) *The Acts of the Apostles* (Doubleday, Garden City, NY).
2. Libanori, L. (1999) *Civiltà Cattolica* **150,** 158–166.
3. Cooper, A. & Poinar, H. R. (2000) *Science* **289,** 1139–1140.
4. Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M. & Pääbo, S. (1997) *Cell* **90,** 19–30.
5. Di Benedetto, G., Nasidze, I. S., Stenico, M., Nigro, L., Krings, M., Lanzinger, M., Vigilant, L., Stoneking, M., Pääbo, S. & Barbujani, G. (2000) *Eur. J. Hum. Genet.* **8,** 669–677.
6. Stone, A. C. & Stoneking, M. (1998) *Am. J. Hum. Genet.* **62,** 1153–1170.
7. Di Benedetto, G., Ergüven, A., Stenico, M., Castrì, L., Bertorelle, G., Togan, I. & Barbujani, G. (2001) *Am. J. Phys. Anthropol.* **115,** 144–156.
8. Calafell, F., Underhill, P., Tolun, A., Anglicheva, D. and Kalaydjieva, L. (1996) *Ann. Hum. Genet.* **60,** 35–49.
9. Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A. & Bertranpetit, J. (1996) *Mol. Biol. Evol.* **13,** 1067–1077.
10. Kruskal, J. B. (1964) *Psychometrika* **29,** 28–42.
11. Tamura, K. (1992) *Mol. Biol. Evol.* **9,** 678–687.
12. Rozas, J. & Rozas, R. (1999) *Bioinformatics (Oxford)* **15,** 174–175.
13. Schneider, S., Roessli, D. & Excoffier L. (2000) ARLEQUIN, Software for Population Genetics Analysis (Genetics and Biometry Lab., Univ. of Geneva, Switzerland), Version 2.000.
14. Poinar, H. N., Hoss, M., Bada, J. L. & Pääbo, S. (1996) *Science* **272,** 864–866.
15. Poinar, H. N. & Stankiewicz, B A. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 8426–8431.
16. Anderson, S., Bankier, T., Barrel, B. G., De Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, S., *et al.* (1981) *Nature (London)* **290,** 457–465.
17. Richards, M. B., Macaulay, V. A., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., *et al.* (2000) *Am. J. Hum. Genet.* **67,** 1251–1276.
18. Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L. & Wallace, D. C. (1996) *Genetics* **144,** 1835–1850.
19. Macaulay, V. A., Richards, M. B., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-Tamir, B., Sykes, B. & Torroni, A. (1999) *Am. J. Hum. Genet.* **64,** 232–249.
20. Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J. & Barbujani, G. (2000) *Am. J. Hum. Genet.* **66,** 262–278.
21. Bandelt, H. J., Forster, P., Sykes, B. C. & Richards, M. B. (1995) *Genetics* **141,** 743–753.
22. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
23. Renfrew, C. (1991) *Camb. Arch. J.* **1,** 3–23.
24. Sokal, R. R. & Rohlf, F. J. (1995) *Biometry* (Freeman, San Francisco), 3rd Ed., pp. 760–762.
25. Barbujani, G., Magagni, A., Minch, E. & Cavalli-Sforza, L. L. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 4516–4519.
26. Comas, D., Reynolds, R. & Sajantila, A. (1999) *Eur. J. Hum. Genet.* **7,** 459–468.
27. Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T. & Batzer, M. A. (2000) *Am. J. Hum. Genet.* **66,** 979–988.
28. Seielstad, M., Minch, E. & Cavalli-Sforza, L. L. (1998) *Nat. Genet.* **20,** 278–280.
29. Balding, D. (1999) in *Microsatellites: Evolution and Applications*, eds. Goldstein, D. B. & Schlötterer, C. (Oxford Univ. Press, Oxford), pp. 198–210.
30. Handt, O., Krings, M., Ward, R. H. & Pääbo, S. (1996) *Am. J. Hum. Genet.* **59,** 368–376.