

Balancing Selection at the Prion Protein Gene Consistent with Prehistoric Kurulike Epidemics

Simon Mead,¹ Michael P. H. Stumpf,² Jerome Whitfield,^{1,3} Jonathan A. Beck,¹ Mark Poulter,¹ Tracy Campbell,¹ James Uphill,¹ David Goldstein,² Michael Alpers,^{1,3,4} Elizabeth M. C. Fisher,¹ John Collinge^{1*}

¹Medical Research Council, Prion Unit, and Department of Neurodegenerative Disease, Institute of Neurology, University College, Queen Square, London WC1N 3BG, UK. ²Department of Biology (Galton Laboratory), University College London, Gower Street, London WC1E 6BT, UK. ³Institute of Medical Research, Goroka, EHP, Papua New Guinea. ⁴Curtin University of Technology, Perth, WA, Australia.

*To whom correspondence should be addressed. E-mail: j.collinge@prion.ucl.ac.uk

Kuru is an acquired prion disease largely restricted to the Fore linguistic group of the Papua New Guinea Highlands which was transmitted during endocannibalistic feasts. Heterozygosity for a common polymorphism in the human prion protein gene (*PRNP*) confers relative resistance to prion diseases. Elderly survivors of the kuru epidemic, who had multiple exposures at mortuary feasts, are, in marked contrast to younger unexposed Fore, predominantly *PRNP* 129 heterozygotes. Kuru imposed strong balancing selection on the Fore, essentially eliminating *PRNP* 129 homozygotes. Worldwide *PRNP* haplotype diversity and coding allele frequencies suggest that strong balancing selection at this locus occurred during the evolution of modern humans.

Prion diseases are invariably fatal transmissible neurodegenerative conditions, which include Creutzfeldt-Jakob disease (CJD) and kuru in humans and bovine spongiform encephalopathy (BSE) and scrapie in animals. According to the protein-only hypothesis, the central molecular event in prion replication is the post-translational recruitment of a normal neuronal glycoprotein (PrP^C) into a self-propagating conformational isomer that accumulates as aggregated material (PrP^{Sc}) (1). Kuru came to the attention of Western medicine in the 1950s as the affected area of the Eastern Highlands of Papua New Guinea came under Australian administrative control. The Fore and neighboring linguistic groups occupied a remote highland area that had had no direct contact with the outside world prior to this. It was the practice in these communities for kinship groups to consume deceased relatives at mortuary feasts. From the evidence of local oral history, this practice was not ancient amongst the Fore and is thought to have started around the end of the 19th century. The first remembered case of kuru was around 1920 and the disease rapidly increased in incidence. Adult women and children of both sexes were primarily affected, reflecting their selective exposure—adult males participated little at feasts. At its peak, kuru killed around 1% of the population annually and some villages were almost devoid of young adult women. Kuru was the first human prion disease shown to be transmissible, by inoculation of chimpanzees with autopsy-derived brain tissue (2). It is hypothesized that kuru originated from consumption of an individual with sporadic CJD (3), a disease with a remarkably uniform worldwide incidence of around 1 per million and a lifetime risk of around 1 in 50,000. The ban on

cannibalism imposed by the Australian authorities in the mid-1950s led to a decline in kuru incidence, and although rare cases still occur these are all in older individuals and reflect the long incubation periods possible in human prion disease. Kuru has not been recorded in any individual born after the late 1950s (4).

A coding polymorphism at codon 129 of *PRNP* is a strong susceptibility factor for human prion diseases. Methionine homozygotes comprise 37% of the UK population whereas valine homozygotes comprise 12% (5). Homozygosity at *PRNP* codon 129 predisposes to iatrogenic and sporadic CJD (5). In iatrogenic CJD caused by exposure to contaminated pituitary hormones, heterozygotes have a longer mean incubation period (6). All cases of variant CJD to date have been methionine homozygotes (7), and homozygotes of either allele have an earlier age of onset for kuru (8).

Heterozygosity at a different *PRNP* polymorphism, E219K, is also associated with resistance to sporadic CJD in Japan (9). Heterozygosity is thought to confer resistance to prion disease by inhibiting homologous protein-protein interactions (5). Blood was obtained from Fore women aged fifty years or above. All had a history of multiple exposure to mortuary feasts. Twenty-three from thirty women over the age of fifty were heterozygotes at codon 129, a finding that is significant (Fisher's Exact Test, $P = 0.01$) compared with the genotypes of the unexposed Fore population, which are in Hardy-Weinberg Equilibrium ($n = 140$). Two large samples of elderly Europeans have also displayed Hardy-Weinberg Equilibrium, suggesting this age effect is local to the Fore (10, 11). The age of onset of kuru in homozygotes of either allele at *PRNP* codon 129 has been estimated around nineteen, but over thirty years for heterozygotes (8). Thus the marked survival advantage for codon 129 heterozygotes provides a powerful basis for selection pressure in the Fore. This finding led us to examine for other evidence of balancing selection at *PRNP* in PNG and worldwide. The primary sequence of the prion protein is highly conserved between primates (12). Methionine is the ancestral amino acid at codon 129 (12). Sequencing and genotyping of over 2000 chromosomes in populations selected to represent worldwide genetic diversity, was used to determine which *PRNP* nucleotide changes were intermediate polymorphisms (defined as a frequency >0.05 in at least one population studied). M129V was polymorphic in all world populations with a reducing cline towards East Asia, the lowest frequency being found in Japan (0.01). The E219K polymorphism was found in Japan and other populations in the Indian

subcontinent and East Asia. One of these two established prion disease resistance polymorphisms was found in every population studied. Unusually, considering the reducing cline of 129V in East Asia, the Fore population have the highest polymorphic frequency of 129V in the world (0.55), the only population for which 129V is the most frequent allele. In total, we found only six intermediate frequency polymorphisms in our screen; five coding changes, M129V, E219K, N171S, G142S, 1-octapeptide repeat deletion; but only one intermediate frequency silent change, A117A (Table 1).

The McDonald and Kreitman (MK) test compares the number of silent and coding nucleotide changes between two species with polymorphisms within a species in a 2 x 2 contingency table (13). *PRNP* does not appear to be a rapidly evolving gene as there are only four non-coding and two coding changes between chimpanzee and human. When a comparison is made between humans and more divergent species, such as other Great Apes, all New and all Old-World Monkeys, the MK test suggests a significant departure from neutral evolution with an excess of human coding polymorphism (The ratio of non-coding to coding polymorphisms in humans is 1:5; the ratios in other species and the associated *P* values for the comparison with the human ratio are: Gibbon 13:2, *P* = 0.0055; Rhesus monkey 23:9, *P* = 0.0185; Spider monkey 27:7, *P* = 0.006; *P* values were calculated using Fisher's exact test).

A detailed linkage disequilibrium SNP map of the *PRNP* locus was generated in the European population (14). The region is block-like in structure and the *PRNP* gene is found within a single haplotype block extending over 30kb. Haplotype diversity was characterized in four world populations (European, African, Japanese and Fore). The most parsimonious explanation for the haplotype diversity world-wide requires only a single methionine-valine mutation. In terms of haplotype genealogy, the effect of balancing selection is to produce a gene tree with two long deep branches. The European, African and Fore-speaking PNG population are characterized by highly distinct divergent clades associated with different alleles at codon 129 (Fig. 1), suggesting that codon 129 is an ancient polymorphism. Within the *PRNP* haplotype block, 4.3kb was sequenced in a chimpanzee to compare nucleotide divergences between human and chimpanzee. 53 divergent nucleotides were found (1.2%). For the same 4.3kb in human, there were 5 divergent nucleotides between methionine and valine clades (0.12%), which was similar to the methionine-valine clade divergence across 13kb of the *PRNP* intron (0.10%).

Old balancing selection is characterized by a deep split in the genealogy which leaves a particular signature on the nucleotide diversity of tightly linked polymorphisms: an excess of highly polymorphic sites and a deficit of low frequency polymorphisms when compared to neutrally evolving sites (15). Tajima's *D* statistic allows us to assess this quantitatively. An excess of old polymorphisms will result in large positive values of the statistic and is indicative of population structure and/or balancing selection. The *D* values calculated in 5 different population samples and the corresponding *P* values (Table 2) show that apart from Japan, all populations have large positive *D* values that are significantly different from the neutral null-hypothesis. The *P* values were calculated through simulations assuming a neutrally evolving constant size and unstructured population. The real demography of the human population is of course very different from these assumptions. Empirical distributions

of Tajima's *D* statistic are now available from a study of 313 genes (16). The empirical distribution was determined using a mixed panel and differs markedly from the theoretical values with most *D* values being negative (Mean: -0.973; 95% CI: [-2.00, 0.955]; 99% CI: [-2.12, 1.68]; Minimum: -2.32; Maximum: 1.64). All populations are significant at the empirical 99% level. We note that the population sample used by Stephens (16) is geographically more inclusive than the populations considered here; as population structure will generally lead to more positive values of Tajima's *D* statistic we thus expect the comparison with the empirical values to be conservative. Thus on the level of background population expansion (corresponding to large negative values of *D*) that seems to have shaped much of human genetic variation, *PRNP* appears exceptional and our results are in agreement with what would be expected under balancing selection. If we plot the distribution of pairwise differences between M and V lineages and within M and V lineages we observe that in European, African and PNG populations the major contributions to the overall pairwise difference comes from the between lineage contribution. The distribution within the M and V lineages considered on their own is centered at low values, comparisons between M and V lineages are significantly different from zero and are centered around moderate to high values (fig. S1). Thus the M and V lineages are highly divergent within the *PRNP* genealogy, consistent with their being under old balancing selection, as indicated by Tajima's *D*. Age estimation of the most recent common ancestor of the M and V lineages suggests a genealogical depth of approximately 500,000 years (SOM text, note 1). Microsatellite and SNP diversity was examined at sites outside the *PRNP* haplotype block, ~30kb upstream and downstream of codon 129, in the Fore and European populations. In the European population, there are significant differences between allele frequencies at two microsatellites and two SNPs on M and V chromosomes. The findings in the Fore are more dramatic with extensive linkage disequilibrium (LD) between codon 129 alleles and the same microsatellites; allele diversity is also reduced (Fig. 2, A and B). More generally, average gene diversity as measured by the heterozygosity is also reduced significantly compared to other populations from around the world (and including PNG coastal populations). All of this points to a major bottleneck in the recent history of the Fore. However, simulations suggest that a simple demographic event is highly unlikely to be able to lead to the observed distribution of M/V allele distribution at codon 129 (Fig. 3, A and B; SOM text, note 2). Thus kuru in the Fore imposed an exceptionally strong balancing selection on the prion locus, and global patterns of diversity in the same gene indicate historical balancing selection. What might have provided the selection pressure at *PRNP* around the world? The well-established examples of balancing selection are the MHC-HLA locus (17), the β -hemoglobin locus and more recently the G6PD locus (18), related to resistance to common infectious disease. While a role for the prion protein in the defense against common conventional infectious agents cannot be ruled out, prion protein gene-ablated transgenic animals are not obviously immune impaired (19). Alternatively, we propose that acquired prion disease may have provided the necessary selective pressure worldwide, as it apparently has in PNG. The source of an acquired prion disease that might provide widespread and significant selection pressure is open to speculation. A prehistoric endemic animal prion disease that was able to cross the transmission barrier to carnivorous

humans is a possibility. However, there is now strong evidence for widespread cannibalistic practices in many prehistoric populations, for example telltale scratches and burn-marks on Neanderthal bones (20) and biochemical analysis of fossilized human feces (21). For kuru, it has been supposed that a rare individual with sporadic CJD was consumed and subsequently prion disease was amplified by the systematic and habitual practice of consumption of the dead at mortuary feasts, leading to recycling of prions in the population. A similar argument has been put forward for the origins of the BSE epidemic in the UK. With the recognition that prion incubation times can exceed a normal lifetime (22) and with increased awareness of sub-clinical prion disease (23) one may not need to postulate the consumption of a rare sporadic CJD patient; more prevalent pre- or sub-clinically infected individuals may be responsible. Although the prion gene could have been subject to other unknown forms of selection, available evidence appears consistent with the explanation that repeated episodes of endocannibalism-related prion disease epidemics in ancient human populations made coding heterozygosity at *PRNP* a significant selective advantage leading to the signature of balancing selection observed today.

References and Notes

1. S. B. Prusiner, *Science* **216**, 136 (1982).
2. D. C. Gajdusek, C. J. Gibbs, M. Alpers, *Nature* **209**, 794 (1966).
3. M. Alpers and L. Rail, *Proc. Aust. Assoc. Neurol.* **8**, 7 (1971).
4. J. Whitfield, M. Alpers, J. Collinge, unpublished observations.
5. M. S. Palmer, A. J. Dryden, J. T. Hughes, J. Collinge, *Nature* **352**, 340 (1991).
6. J. Huillard d'Aignaux *et al.*, *Neurology* **53**, 1197 (1999).
7. J. Collinge, J. Beck, T. Campbell, K. Estibeiro, R. G. Will, *Lancet* **348**, 56 (1996).
8. L. Cervenakova *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13239 (1999).
9. S. Shibuya, J. Higuchi, R. W. Shin, J. Tateishi, T. Kitamoto, *Ann. Neurol.* **43**, 826 (1998).
10. C. Berr *et al.*, *Neurology* **51**, 734 (1998).
11. O. Combarros *et al.*, *Neurology* **55**, 593 (2000).
12. H. M. Schatzl, M. Da Costa, L. Taylor, F. E. Cohen, S. B. Prusiner, *J. Mol. Biol.* **265**, 257 (1997).
13. J. H. McDonald and M. Kreitman, *Nature* **351**, 652 (1991).
14. S. Mead *et al.*, *Am. J. Hum. Genet.* **69**, 1225 (2001).
15. H. Harpending, A. Rogers, in *Annu. Rev. Genomics Hum. Genet.* (2000).
16. J. C. Stephens *et al.*, *Science* **293**, 489 (2001).
17. A. L. Hughes, *Adaptive Evolution of Genes and Genomes* (Oxford Univ. Press, New York, ed. 1, 1999), pp. 1–270.
18. S. A. Tishkoff *et al.*, *Science* **293**, 455 (2001).
19. H. Bueler *et al.*, *Nature* **356**, 577 (1992).
20. A. Defleur, T. White, P. Valensi, L. Slimak, E. Cregut-Bonnoure, *Science* **286**, 128 (1999).
21. R. A. Marlar, B. L. Leonard, B. R. Billman, P. M. Lambert, J. E. Marlar, *Nature* **407**, 74 (2000).
22. A. G. Dickinson, H. Fraser, G. W. Outram, *Nature* **256**, 732 (1975).
23. A. F. Hill *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10248 (2000).
24. This work was funded by the UK Medical Research Council and the Wellcome Trust. We thank the Fore communities and Professor John Reeder and his staff at the PNG Institute for Medical Research for their generous support of this work.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1083320/DC1

SOM text

Fig. S1

11 February 2003; accepted 21 March 2003

Published online 10 April 2003; 10.1126/science.1083320

Include this information when citing this paper.

Table 1. Frequency of human *PRNP* ORF polymorphisms in various world populations. *n* = number of individuals genotyped for M129V, A117A and the octapeptide repeat polymorphism by PCR and restriction digestion. UP = Uttar Pradesh. CEPH = Centre d'Etude du Polymorphisme Humain. Dash indicates no data. For sequencing details, see SOM text, note 3.

Table 2. Calculated values for Tajima's statistic in four world populations. These are shown together with their *P* values under the assumption of a standard neutral, constant size evolutionary model.

Fig. 1. Examples of predicted haplotype genealogy from four populations: (a) European, (b) Japanese, (c) Fore, and (d) African. Haplotypes are colored in orange (129M) blue (129V) or pink (129M, 219K) (14). Note the linear structure, loss of diversity in the Fore, and the location of 129V and 219K at extremes of all genealogies. Results are similar using maximum likelihood, parsimony or distance methods. For method details, see SOM text, note 4.

Fig. 2. Allele frequencies at microsatellites tightly linked to *PRNP*. Allele frequency at microsatellites in four populations: (A) microsatellite 53, 24kb downstream of codon 129, and (B) microsatellite 108, 30kb upstream of codon 129. Note the marked difference between PNG 129M and 129V when compared to CEPH 129M and 19V although this is less striking for 53 than for 108. Note also the reduced diversity of alleles in PNG compared with other populations. See SOM text, note 5.

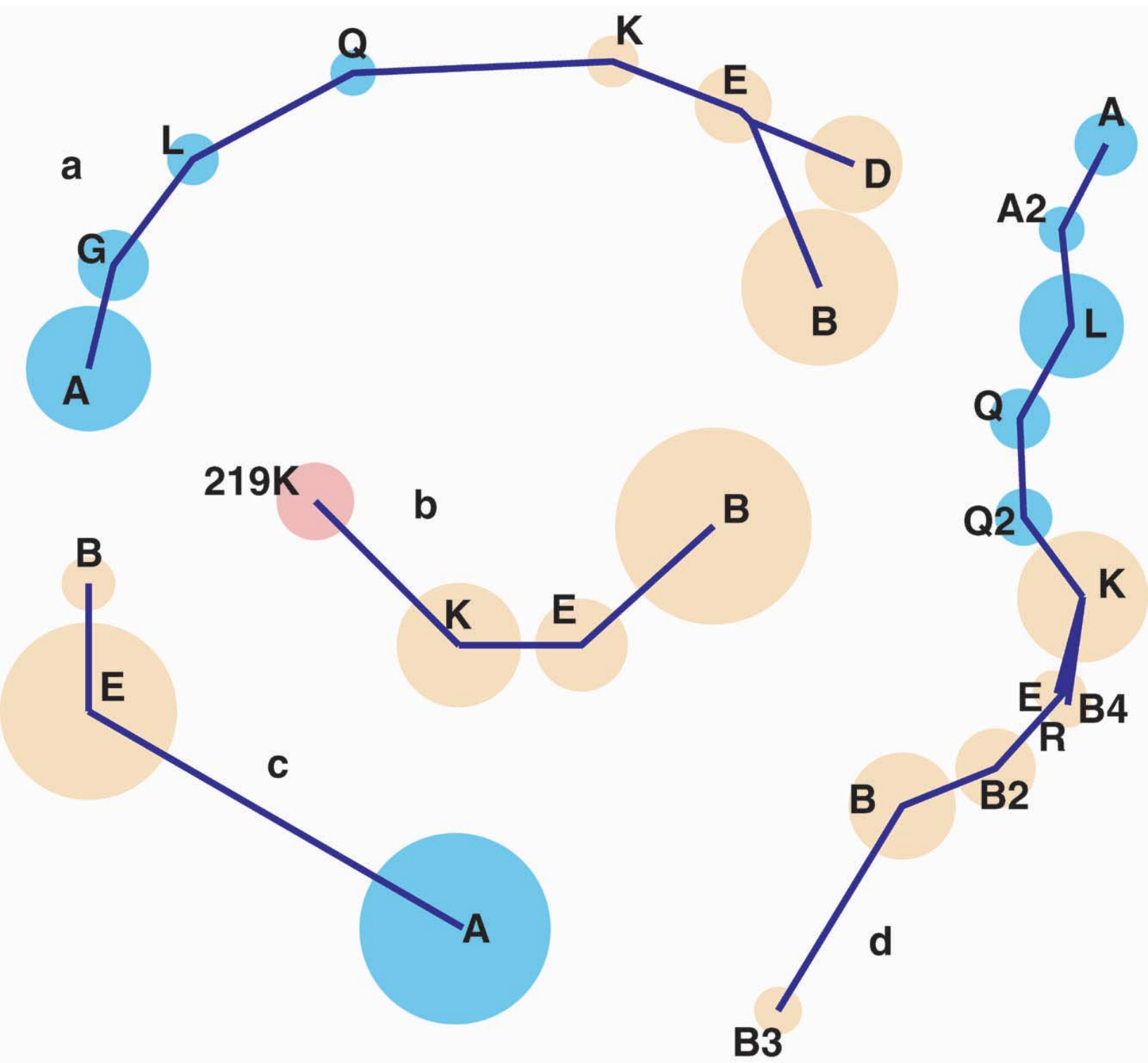
Fig. 3. Probability distribution resulting from a simulation of the effect of a population bottleneck on the frequency of the M129V allele. The model assumes a very severe recent bottleneck that reduced the population size to 1/300th of the initial size. The starting 129V frequency used was the CEPH population frequency. (A) bottleneck for one generation, (B) bottleneck for 10 generations. Using the African or Japanese 129V frequency makes the probability of attaining the present day PNG (Fore) 129V frequency, shown by the dashed vertical line, even smaller. See SOM text, note 2.

Table 1. Frequency of human *PRNP* ORF polymorphisms in various world populations. *n* = number of individuals genotyped for M129V, A117A and the octapeptide repeat deletion polymorphism by PCR and restriction digestion. UP = Uttar Pradesh. CEPH = Centre d'Etude du Polymorphisme Humain. Dash indicates no data. For sequencing details, see SOM text, note 3.

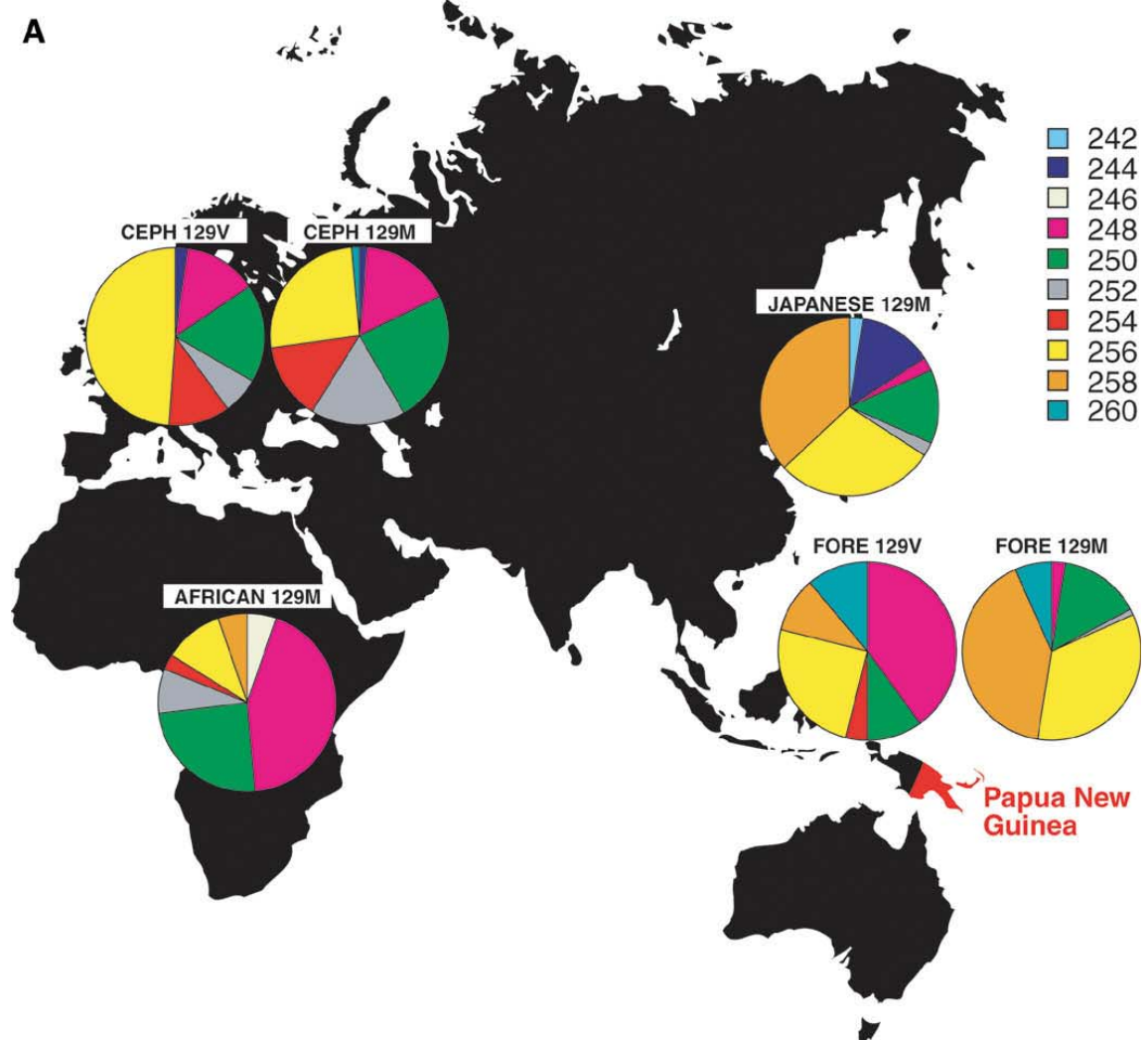
Population	<i>n</i>	M129V (%)	E219K (%)	N171S (%)	G142S (%)	1-OPRD (%)	A117A (%)
<i>African</i>							
Yemeni Hadramout	15	27	0	0	6	0	3
Yemeni Sena	22	32	0	0	0	2	5
Cameroon	39	35	0	8	0	-	9
Jamaican	100	32	0	5	3	30	0
<i>South Asian</i>							
Sri Lankans	35	23	13	0	0	0	1
Non-UP Indians	88	21	5	0	0	2	2
UP Indians	64	28	6	0	0	0	3
<i>East Asian</i>							
Japanese	36	1	7	0	0	-	0
Taiwan (5 populations)	70	3	3	0	0	0	0
<i>Pacific</i>							
PNG (Madang)	83	30	1	0	0	0	0
PNG (Fore)	48	55	0	0	0	0	0
Fiji (Taveuni)	10	5	-	-	-	5	0
Fiji (others)	18	17	-	-	-	3	0
Vanuatu (Port Olry)	33	5	13	0	0	8	0
Vanuatu (Maewo)	32	17	8	0	0	13	0
Tonga	22	14	7	0	0	0	0
<i>European</i>							
Turkish	61	48	0	0	0	3	4
Ceph parents	122	38	0	0	0	1	3
Georgian Jews	74	26	0	0	0	0	3
<i>South American</i>							
Columbian	148	41	0	0	0	3	3

Table 2. Calculated values for Tajima's statistic in four world populations. These are shown together with their *P* values under the assumption of a standard neutral, constant size evolutionary model.

	CEPH	UK	Africa	Japan	PNG
Tajima <i>D</i>	3.799	3.433	2.200	1.148	2.982
<i>P</i>	<10 ⁻³	0.004	0.03	0.168	0.01



A



B

