Letter

The prion protein gene in humans revisited: Lessons from a worldwide resequencing study

Marta Soldevila, Aida M. Andrés,[1] Anna Ramírez-Soriano, Tomàs Marquès-Bonet, Francesc Calafell, Arcadi Navarro, and Jaume Bertranpetit[2]

Unitat de Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain

Ample evidence has accumulated showing that different coding variants of the *PRNP* gene confer differential susceptibility for prion diseases. Here we evaluate the patterns of nucleotide variation in *PRNP* exon 2, which includes all the protein-coding sequence, by resequencing a worldwide sample of 174 humans for 2378 bp. In line with previous studies, we found two main haplotypes differentiated by nonsynonymous substitution in codon 129. Our analyses reveal the worldwide pattern of variation at the *PRNP* gene to be inconsistent with neutral expectations, indicating instead an excess of low-frequency variants, a footprint of the action of either positive or purifying selection. A comparison of neutrality test statistics for *PRNP* with other human genes indicates that the signal of positive selection on *PRNP* is stronger than expected from a possible confounding genome-wide background signal of population expansion. Two main conclusions arise from our analysis. First, the existence of an ancient, stable, balanced polymorphism that has been claimed in a previous study and related to cannibalism can be rejected and is shown to be due to ascertainment bias. Second, our results are consistent with a complex history of selection including mainly positive selection, even if short local periods of balancing selection (Kuru-like episodes), or even a weak purifying selection model, are consistent with our data.

Transmissible spongiform encephalopathies (TSEs), or prion diseases, are a group of rare, subacute and fatal neurodegenerative disorders characterized by accumulation of the abnormal isoform of a host-encoded membrane protein. Human TSEs can be sporadic, acquired, or genetic, and include, among others, Creutzfeldt-Jakob disease (CJD), Kuru (a disease confined to a population in Papua–New Guinea), and variant CJD (vCJD), a concept coined to designate cases potentially caused by the human consumption of cattle suffering from bovine spongiform encephalopathy (BSE).

The human prion protein (PrP) is a product of a single gene located on the short arm of Chromosome 20 (Prusiner 1991). It is encoded by a single exon of *PRNP*, exon 2. Variation in the gene sequence produces protein variants that are causative of genetic TSE diseases; the most common are at codons 200, 178, and 102. Other polymorphisms have been linked to differential susceptibility to the acquired TSE diseases, particularly those in codons 129 and 219 (Palmer et al. 1991; Laplanche et al. 1999; Soldevila et al. 2003).

In the case of codon 219, it has been suggested that the Lys allele acts as a protective factor against sporadic CJD (Shibuya et al. 1998), and it has been shown to be restricted to Asian and Pacific regions (Soldevila et al. 2003).

The common methionine/valine (Met/Val) polymorphism at codon 129 is generally considered to be the most important in genetic susceptibility to prion diseases. Up to 90% of sporadic CJD (sCJD) cases have occurred in individuals who are homozygous for either version of codon 129, Met-Met or Val-Val (Palmer et al. 1991). So far all the vCJD cases reported have been Met-Met homozygotes (Collinge et al. 1996; Valleron et al. 2001; Andrews et al. 2003). Homozygosity at codon 129 is also a key factor in the resistance or susceptibility to the Kuru prion disease, which was shown to have been transmitted during endocannibalistic feasts among the Fore linguistic group in New Guinea (Cervenakova et al. 1998). A worldwide survey of codon 129 revealed marked geographic differences in frequencies of the two variants among continents (Soldevila et al. 2003).

Mead et al. (2003) screened Fore women, who were likely to have participated in cannibalistic feasts and many of whose cohorts had died of Kuru. A statistically significant excess of heterozygotes for the codon 129 polymorphism was found among these women, implying a heterozygote resistance to the disease. Balancing selection in this generation appeared to be the strongest yet documented in any human population (Hedrick 2003). Based on a wide analysis of *PRNP* sequence and haplotype diversity in a worldwide sample, Mead et al. (2003) postulated that variation at this locus had been shaped by strong balancing selection related to prion diseases and cannibalism during the evolution of modern humans. However, as suggested by Kreitman and Di Rienzo (2004), and confirmed by Soldevila et al. (2005), the study of Mead et al. (2003) is affected by a bias in the ascertainment of genetic variants. Mead et al. (2003) first screened *PRNP* polymorphisms by sequencing a limited number of individuals, and then typed the polymorphisms thus identified in a larger worldwide sample of individuals. This led to the exclusion of low-frequency variants and artificially tilted the allele-frequency spectrum toward a pattern consistent with balancing selection, summarized as strongly positive values of statistics such as Tajima's *D*.

Variant CJD in humans is believed to be acquired from cattle infected with BSE; thus, similar selective forces could have been acting in other species. Seabury et al. (2004) studied the selective pressures in the *PRNP* gene in domestic cattle, and they found that purifying selection was the main force driving the evolution of this gene.

[1]Present address: Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[2]Corresponding author.
E-mail jaume.bertranpetit@upf.edu; fax (34) 93 542 28 02.
Article published online ahead of print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.4345506.

The possibility that cannibalism has been a major and global factor in human history, as proposed by Mead et al. (2003), is an issue with wide social implications that deserves proper and careful data and analysis. In fact, selection may have acted distinctly in different places, which implies that the footprint of selection should be sought in the patterns of genetic diversity in local or regional populations. In order to understand the pattern of variation and clarify the selective pressures that have acted on the *PRNP* gene in humans, we undertook a resequencing study of the entire exon 2 of *PRNP* in a sample of 348 chromosomes of humans from populations worldwide.

Results

Codons 129 and 219

Typing the worldwide CEPH-HGDP panel (2128 chromosomes) for codons 129 and 219 confirmed and expanded the results obtained in a smaller previous study (Soldevila et al. 2003). Allele frequencies at codon 129 are geographically heterogeneous (Table 1), with the valine allele ranging, at the continent level, from 5% in East Asia to 65% in the Americas. We note that the very high frequency of 129Val in the Americas is consistent among the five different populations examined in this continent, albeit with a slightly higher frequency in South America.

Codon 219 is found to be monomorphic in Europe, Africa, and the Americas, with the 219Lys allele observed only in Asian and Pacific regions, and very rare in the Middle East/North African population (Table 1). Two homozygotes for the 219Lys allele have been detected, both from Central/South Asian regions, and have been confirmed by sequencing. Hardy-Weinberg equilibrium holds for all populations.

PRNP nucleotide sequence variation

For the 2378 bp sequenced in exon 2 of the *PRNP* gene in 174 humans, 22 variants have been observed, thereof 18 SNPs and five length polymorphisms. Of the latter, three have been identified in the octapeptide repeat region. This region consists of a series of repeats, usually five, named R1 to R4 depending on their sequence, where each codes for eight amino acids, with the exception of R1, which codes for nine. The composition of this region is generally known as R1-R2-R2-R3-R4. We observed one chromosome with a deletion of an R2 repeat, five with the R3-R4 region (also known as region C) deleted, and one with an extra R3-R4 repeat inserted. In fact, this region, in its structure and variation, should be considered as a coding minisatellite. Outside this region, we have detected one single base deletion at position 27495, and another 2-bp deletion at position 26372–26373 of the U29185 reference sequence. All cases of length variation have been analyzed in haploid chromosomes after cloning the PCR products (see Methods).

A total of 12 of the 22 variants are singletons (nine SNPs and three indels), all of which have been confirmed by two independent PCR amplifications and sequencing both strands. A detailed and careful examination of chromatograms has been performed (see Methods). Among the 18 SNPs, transitions (12) are more frequent than transversions (six). Nine of the polymorphisms are found in the coding sequence including the two deletions in the octapeptide repeat region, and seven SNPs. Moreover, 13 polymorphisms, some of them previously unreported, are detected in the 3′-UTR of the *PRNP* exon 2 (Table 2), including two insertions (26372–26373 [TA] and 27495 [T]).

Two of the variants (at 142Ser and 232Arg) deserve special attention, as they have been related to disease. The unconservative amino acid change Gly142Ser might be involved in neurological diseases, as it had been reported for the first time in a North African man with multiple sclerosis and in a Malian woman with viral meningoencephalitis (J.L. Laplanche, unpubl.; see the Official Mad Cow Disease Web site, http://www.madcow.org). In our sample, we have found this substitution in four heterozygotes from Sub-Saharan Africa with a total frequency of 6%. It was also reported by Mead et al. (2003) in individuals of African descent. Variant Met232Arg has been detected in one Japanese; this change was first found as a compound heterozygote with a mutation on codon 180 in a Japanese patient with prion disease (Kitamoto et al. 1993), and a role in causing disease was suggested. However, later reports found this change in healthy individuals (Hitoshi et al. 1993; Hoque et al. 1996), and we have found it in one out of 16 chromosomes from healthy Japanese individuals.

Table 1. Genotype and allele frequency distribution of codon 129 and codon 219

Group	2*N*	Samples failed	129Met/Met	129Val/Val	129Met/Val	129Met	129Val
Africa	254	1	46 (36.5)	21.4 (27)	42.1 (53)	145 (57.5)	107 (42.5)
Middle East/North Africa	356	2	89 (50.6)	20 (11.4)	67 (38)	245 (69.6)	107 (30.4)
Europe	322	1	75 (46.9)	12 (7.5)	73 (45.6)	223 (69.7)	97 (30.3)
Central/South Asia	400	1	98 (49.3)	11 (5.5)	90 (45.2)	286 (71.9)	4.8 (28.1)
East Asia	502	1	227 (90.8)	1 (0.4)	22 (8.8)	476 (95.2)	24 (4.8)
Pacific	78	0	24 (61.5)	4 (10.3)	11 (28.2)	59 (75.6)	19 (24.4)
America	216	2	16 (15.1)	48 (45.3)	42 (39.6)	74 (34.9)	138 (65.1)
Group	**2*N***	**Samples failed**	**219Glu/Glu**	**219Lys/Lys**	**219Glu/Lys**	**219Glu**	**219Lys**
Africa	254	0	127 (100)	0 (0)	0 (0)	254 (100)	0 (0)
Middle East/North Africa	356	2	174 (98.9)	0 (0)	2 (1.1)	350 (99.4)	2 (0.6)
Europe	322	7	154 (100)	0 (0)	0 (0)	308 (100)	0 (0)
Central/South Asia	400	2	193 (97.5)	2 (1)	3 (1.5)	389 (98.2)	7 (1.8)
East Asia	502	1	236 (94.4)	0 (0)	14 (5.6)	486 (97.2)	14 (2.8)
Pacific	78	2	33 (89.2)	0 (0)	4 (10.8)	70 (94.6)	4 (5.4)
America	216	2	106 (100)	0 (0)	0 (0)	212 (100)	0 (0)

Data were obtained by genotyping the entire HGDP-CEPH panel. The total number of chromosomes analyzed was 2128, and the total failures was eight for codon 129 and 16 for codon 219.

Table 2. Haplotypes found in exon 2 of the *PRNP* gene

	R2	DELC	INSC	117	129	142	171	177	219	232	26215	26281	26372	26611	26803	26883	26893	26908	27011	27079	27495	27623	27802		Europe	W-Asia	E-Asia	Africa A	Africa B	Pacific	America
	R	D	I										T																		
S1	2	C	C	A	**A**	G	A	C	G	T	A	G	A	T	T	C	A	A	C	T	T	T	G	**216**	72	36	56	9	16	23	4
S4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	**1**	1						
S5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	**2**	2						
S6	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	**1**	1						
S7	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**3**						3	
S9	.	–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**5**		1		4			
S10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	**1**		1					
S11	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**1**		1					
S12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	**1**			1				
S13	.	.	.	.	.	.	.	.	.	.	.	.	–	.	.	.	.	.	.	.	.	.	.	**1**			1				
S14	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	**1**			1				
S15	.	.	.	.	.	.	.	.	A	.	.	.	–	.	.	.	.	.	.	.	.	.	.	**1**			1				
S17	–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**1**					1		
S19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	**1**				1			
S20	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**4**				2	2		
S26	.	–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	**1**					1		
S27	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	**1**					1		
S2	.	.	.	.	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**87**	27	21	3	8	11	5	12
S3	.	.	.	G	**G**	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	**6**	1	4		1			
S8	.	.	.	.	**G**	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**1**						1	
S16	.	.	.	.	**G**	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**1**			1				
S18	.	.	.	.	**G**	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**1**					1		
S21	.	.	.	.	**G**	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	**1**				1			
S22	.	.	.	.	**G**	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	**2**					2		
S23	.	.	.	.	**G**	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	**1**				1			
S24	.	.	.	.	**G**	.	.	.	.	.	T	.	.	.	.	.	.	G	.	.	.	.	.	**3**					3		
S25	.	.	.	G	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**2**					2		
S28	.	.	.	.	**G**	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	–	.	**1**				1			
																								348	104	64	64	28	40	32	16
Chimp	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.									

The first group of haplotypes contain methionine (A) in position 129, while the second group contains valine (G). The arbitrary names of the haplotypes are indicated on the left and the continental region where the haplotype is found on the right. The total haplotypic frequency is also shown. The alleles in chimpanzee are shown only for the human *PRNP* variable sites.

The diversity and genealogy of *PRNP* haplotypes

The 22 polymorphisms define 28 different haplotypes (Table 2; see Methods for haplotype reconstruction). For the chimpanzee, from a total of 26 divergent positions between the two species, only one position that is polymorphic in the human sequence is shown. Haplotype diversity for the entire data set is relatively low (0.551), and 87.7% of the total haplotype variation can be accounted for by only two haplotypes, S1 (with 129Met) and S2 (with 129Val), which differ only at codon 129. Major haplotype groups derive both from S1 and from S2. Nucleotide diversity (π) is 0.00031; this value is very low compared to other genes that have been sequenced in a sufficient number of diverse individuals (the average for 245 genes in the SeattleSNP database is 0.0009) (see also Stephens et al. 2001; Akey et al. 2004; Livingston et al. 2004). The θ_π estimator is 0.74, which corresponds to the mean number of nucleotide differences between all pairs of sequences in the data set. The θ_w statistic, computed from the number of segregating sites (S) in a given locus and considering only SNPs, has been estimated as 2.8.

Figure 1 shows a median-joining network, describing the most plausible mutational relationships between the 28 haplotypes. The phylogenetic structure around haplotypes S1 and S2 is clearly star-like, with S1 having a slightly higher proportion of haplotypes derived by a single mutational step. Assuming that the phase of haplotypes is correctly resolved, reticulations in the network represent either homoplasy (recurrent mutation events) or recombination events. There are only four single reticulations, involving positions 129–219, 219–26372, 129–26908, and 26893–117. All are transitions and positions 117, 129, 219, and 26893 are in CpGs, making them more likely to be recurrent mutation events, although crossing-over or gene conversion events cannot be ruled out. It is interesting to note that the susceptibility polymorphisms at codons 129 and 219 are involved in three out of four cases, indicating that one or both of them may have a recurrent origin.

Testing for the impact of natural selection

Different neutrality tests have been used. Tajima's *D* statistic is −1.85 for SNPs in the combined sample, which is significantly smaller than expected under neutrality ($P < 0.05$, according to coalescent simulations). When length polymorphisms are included, *D* becomes −2.02 ($P < 0.05$). If statistical significance is evaluated by means of coalescent simulations with recombination ($R = 2.79$; see Methods), statistical significance increases ($P < 0.001$). Values for geographic regions will be discussed below. These values contrast with the positive values found by Mead et al. (2003).

A positive value of *D* points to the possible impact of bal-

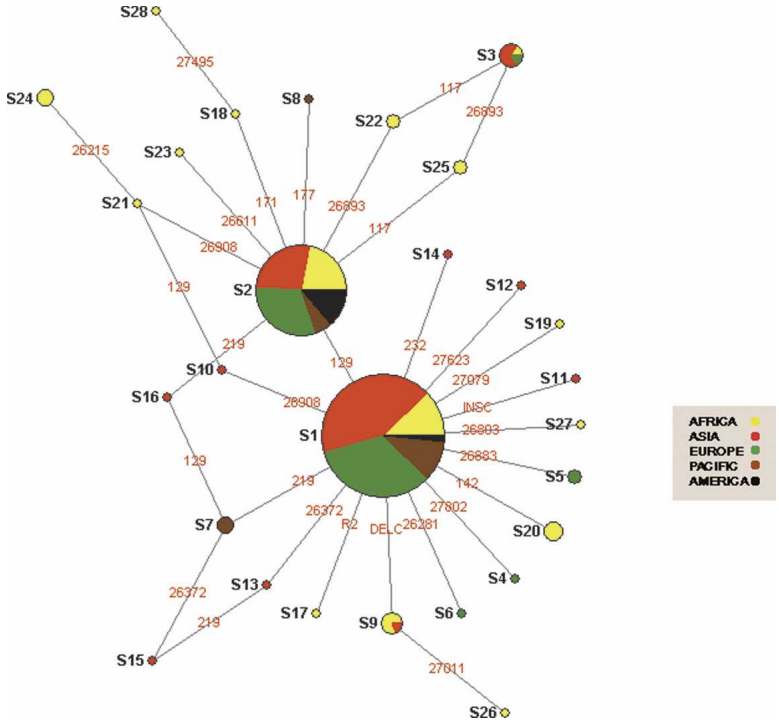


Figure 1. Median-joining network of *PRNP* haplotypes. The circle areas are proportional to the frequency of the haplotypes. The branch lengths are proportional to the number of allele differences they represent. The haplotype number is shown next to the circles, and mutation positions are indicated on the branches linking two haplotypes. Each color inside the circle indicates the presence of the haplotype in a specific continental region.

ancing selection or population subdivision (under simple models of population structure, discussed below) on the patterns of diversity, while a negative value can indicate either recent positive selection, a population expansion, or purifying selection on slightly deleterious alleles (Tajima 1989). Thus, with this and other similar neutrality tests, there is a confusing signal that can be due either to selection or the impact of demographic factors. The assumption of constant population size is anticonservative in the detection of positive natural selection; in fact, the statistical significance of the obtained negative *D*-value is lost if the confidence interval is constructed by simulation under a series of population histories with population expansion (see Figure 4 in Wooding et al. 2004). Another way of evaluating the neutrality test without requiring assumptions on the past is through the comparison with the empirical distribution of existing data sets. Owing to the known ethnical composition and availability of data, the SeattleSNP database has been used. When compared with 132 sequenced genes, a similar *D*-value is found for the *EPHB6* gene, which has been described as having a strong signal of positive selection in Europeans even after correcting the results for multiple tests (Akey et al. 2004).

Negative significant results have been also obtained for Fu and Li's D^* and F^* statistics when just considering SNPs ($D^* = -3.65$, $P < 0.02$; $F^* = -3.54$, $P < 0.02$) (Table 3), and a higher significant deviation from neutral expectations is obtained when length polymorphisms are included ($D^* = -3.77$, $P < 0.002$; $F^* = -3.69$, $P < 0.001$). Fu's *F* test is negative for the combined worldwide sample (−47.28), with highly significant *P*-values using coalescent simulations with and without recombination.

The chimpanzee sequence has been used to infer the ancestral states of human polymorphisms (which were, with just one

Table 3. Neutrality tests for each continental region and for all the samples together (world)

	Africa	Africa A	Africa B	Europe	West Asia	East Asia	Pacific	America[a]	World	World ($R = 2.79$)
S	14	12	11	6	6	5	3	1	23	23
2*N*	68	28	40	104	64	64	32	16	348	348
Tajima's *D*	−1.34	−1.43	−1.32	−1.20	−0.87	−1.64	−0.60	—	−2.02	−2.02
P-value	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	—	Significant	Significant
	$P > 0.10$	$P > 0.10$	$P > 0.10$	$P > 0.10$	$P > 0.10$	$0.10 > P > 0.05$	$P > 0.10$		$P < 0.05$	$P < 0.001$
Fu and Li's D^*	−1.06	−1.29	−0.73	−2.73	−1.53	−0.93	−0.28	—	−3.77	−3.77
P-value	N.S.	N.S.	N.S.	Significant	N.S.	N.S.	N.S.	—	Significant	Significant
	$P > 0.10$	$P > 0.10$	$P > 0.10$	$P < 0.05$	$P > 0.10$	$P > 0.10$	$P > 0.10$		$P < 0.002$	$P < 0.005$
Fu and Li's F^*	−1.38	−1.56	−1.08	−2.62	−1.55	−1.35	−0.43	—	−3.69	−3.69
P-value	N.S.	N.S.	N.S.	Significant	N.S.	N.S.	N.S.	—	Significant	Significant
	$P > 0.10$	$P > 0.10$	$P > 0.10$	$P < 0.05$	$P > 0.10$	$P > 0.10$	$P > 0.10$		$P < 0.001$	$P < 0.001$
Fu's *F*	−17.61	−6.71	−7.85	−3.50	−1.41	−6.60	−0.93	—	−47.28	−47.28
P-value	Significant	Significant	Significant	N.S.	N.S.	Significant	N.S.	—	Significant	Significant
	$P < 0.0005$	$P < 0.006$	$P = 0.003$	$0.10 > P > 0.05$	$P > 0.10$	$P < 0.002$	$P > 0.10$		$P < 0.001$	$P < 0.001$
Fay and Wu's *H*	−0.80	−0.73	−0.92	0.38	0.29	0.49	0.46	—	−1.37	−1.37
P-value	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	—	N.S.	N.S.
	$P > 0.10$	$P > 0.10$	$P > 0.10$	$P > 0.10$	$P > 0.10$	$P > 0.10$	$P > 0.10$		$P > 0.10$	$P > 0.10$

Significance of Tajima's *D*, Fu and Li tests. Fu's *F* and Fay and Wu's *H* are calculated by coalescent simulations using Dnasp 4.00. For the world sample, values of *P* for all neutrality tests are also calculated by taking into account recombination ($R = 2.79$).
N.S. is nonsignificant (>0.05).
[a]Note that only a single SNP is found and these tests have not been computed in this group.

exception, also the most frequent in humans), which allowed us to compute Fay and Wu's H statistic. This is -1.37 ($P > 0.05$) (Table 3), which implies that derived alleles are not found at higher than expected frequencies, which would be the case under positive selection. However, this test would not detect old sweep events (Przeworski 2002).

The McDonald and Kreitman (MK) test of neutrality has been applied to the combined worldwide sample using a range of different primate species as outgroups to compare human polymorphism. While comparisons of humans with the closest species are nonsignificant, those of humans with gibbon and siamang yielded significant results ($P = 0.007$ in both cases).

The pairwise distribution of mutational differences

Figure 2 shows the mismatch distribution obtained from the combined worldwide sample as a smooth curve with the maximum at zero differences followed by a monotonic decrease, such as that predicted for a population that has undergone recent expansion (Rogers and Harpending 1992). The distribution of mutational differences between all pairs of sequences is unimodal both for the total sample and for all geographic groups. A bell-shaped, unimodal curve is consistent with positive selection and contrasts with the bimodal distribution obtained by Mead et al. (2003), and which they interpreted as the result of balancing selection.

Time to the most recent common ancestor (TMRCA) and mutation ages

By resolving the few network reticulations shown in Figure 1 (see Methods), it has been possible to use the Genetree program to estimate the age of the tree and of individual mutations using a coalescent approach. A substitution rate of 0.91×10^{-9} per nucleotide and year has been obtained (see Methods). We have computed the TMRCA assuming a population with constant size and a population growth model. θ_{ML} and the exponential growth parameter β_{ML} have been simultaneously estimated using Genetree. The phylogenetic structure around the two main haplotype clusters is clearly star-like shaped (Fig. 3). The estimation of TMRCA for the entire tree under a constant population model is 380 ± 105 thousand years (ky), which is in the low range of estimations for other human gene genealogies (Harris and Hey 2001; Martinez-Arias et al. 2001). The estimate obtained under a model of population growth, more adapted to human history, is 200 ± 50 ky.

The polymorphism at codon 129 polymorphism is the oldest in the tree, with an age estimated at ~200 ± 100 ky under a constant population model and close to ~100 ± 65 ky under population growth (Fig. 3). Both dates are quite recent for the oldest polymorphism at a human autosomal locus. TMRCA and mutation-age estimates obtained under a population growth model are markedly lower than those obtained under a model of constant population size.

Geographic variation of diversity and selection statistics

When examined by geographic region, the number of segregating sites and heterozygosity indicate that Sub-Saharan Africans harbor the greatest diversity, while Native Americans exhibit by far the greatest homogeneity, with only the two basic haplotypes that differ by the polymorphism at codon 129.

In most cases, the neutrality tests for individual geographic regions are nonsignificant (Table 3), an expected result because of the lack of power caused by small sample size. It is interesting to note that under simple models of population substructure, Tajima's D is expected to be positive; nonetheless, it has been shown that with real data for humans, pooling more populations decreases its value (Ptak and Przeworski 2002 and references therein) but to an extent that would not explain the low values found: The global Tajima's D-value and that for continents are all negative.

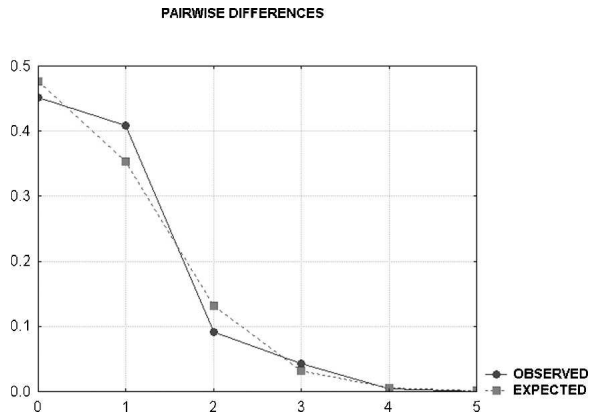


Figure 2. Pairwise difference distribution of all the samples. The results show no difference between observed (circles) and expected (squares) distribution.

In order to distinguish between demographic and selective forces in African and European continental groups, we compared our results with values of Tajima's D obtained for 245 genes (SeattleSNPs database; September 2005) in these two groups. Using this data set as an empirical distribution, both Tajima's D-values are nonsignificant (Africa, $D = -1.34$; $P < 0.065$; Europe, $D = -1.20$; $P < 0.073$), but fall clearly in the left side tail of the distribution. In this case, substructure would not affect the values (Ptak and Przeworski 2002 and references therein). To avoid the bias introduced by sample size, we have recalculated Tajima's D-values for subsamples of our data of the same size as that of the empirical distribution of the two populations in SeattleSNP. Unbiased values show the same trend (Africa, $D = -1.31$; $P < 0.077$; Europe, $D = -0.71$; $P < 0.167$), although with less statistical significance.

A different approach to detect a deficit of polymorphism due to selection that is not biased by the history of human populations is comparing, for the same ethnic group, the values of θ_w (calculated per site) divided by divergence, with those of the Seattle SNPdatabase (Thompson et al. 2004, which considers 159 genes). Results show a low level of variation in Europe (θ_w:div = 0.0450, $P < 0.144$; with resampling to correct for different sample size θ_w:div = 0.0266, $P < 0.015$) but an intermediate value in Africa (θ_w:div = 0.1124, $P < 0.482$; with resampling to correct for different sample size θ_w:div = 0.0881, $P < 0.179$). Sample size has a strong effect on the θ_w-value.

Geographic heterogeneity

F_{ST}-values calculated for the two SNPs in codons 129 and 219 that have been typed for the entire worldwide HGDP-CEPH panel are very different among them: 0.16 for codon 129 and 0.01 for 219. Within continental regions, F_{ST}-values for the same SNPs are very low (Table 4), with the maximum in Amerindian populations (0.09) for codon 129.

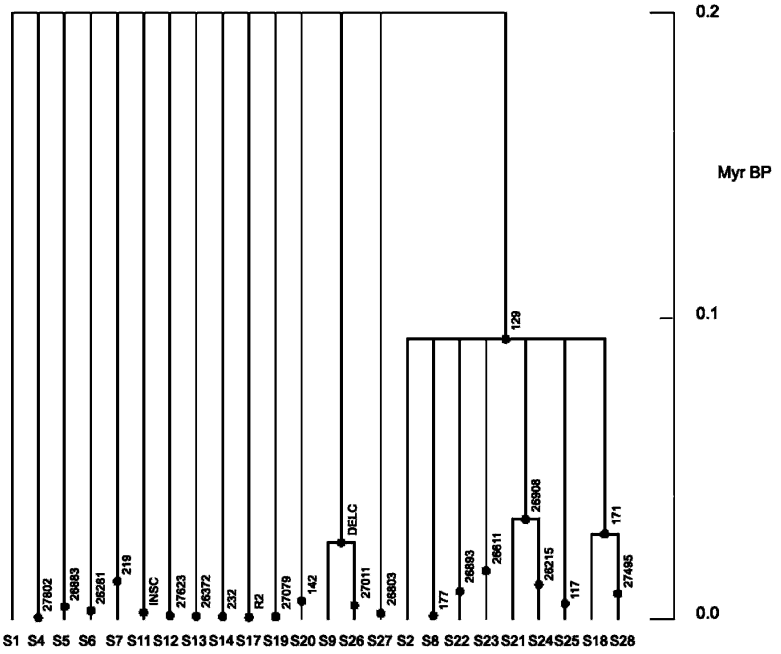


Figure 3. Coalescent tree of the *PRNP* haplotypes under population growth. Numbers along branches represent mutational events. Population growth has been taken into account. The tree was computed using $\theta_{ML} = 7.1$, growth parameter = 3.8, $N = 10{,}000$, and generation time = 25.

The magnitude of the F_{ST}-values found has been empirically compared to two distributions obtained with large numbers of genes: (1) Akey et al. (2002) provided F_{ST}-values for thousands of genes among three geographic regions, and the value obtained for codon 129 is not particularly extreme; (2) Kidd et al. (2004) compiled allele frequencies and F_{ST}-values for 369 SNPs in a diverse population set that is comparable to the one used here and thus a more appropriate reference set. Again, the codon 129 F_{ST}-value (0.16) is lower than that found in 28% of the polymorphisms. This means that the geographical variation of this SNP is not unexpectedly large. A similar analysis has been performed for haplotypes generated from the sequence data (not shown) with equivalent results.

Ascertainment bias

The comparison of our data with that of Mead et al. (2003), in which polymorphisms were ascertained in a limited number of individuals, provides a possibility of illustrating in detail the effect of ascertainment of patterns of variation; this can be done for the complete exon 2, common to both studies. Unfortunately, it is only possible to reconstruct the raw data of the former study (Mead et al. 2001, 2003) for Europeans, where polymorphisms with their frequency are described. These polymorphisms were at the base for typing a larger sample.

The effects of ascertainment bias on three parameters (Tajima's *D* and Fu and Li's *D** and *F**) are shown in Figure 4, where these parameters are given for an increasing number of polymorphisms being added in descending frequency order; thus, the first point includes only the most frequent polymorphism (codon 129), and the remaining are being added until the total of six (specified in the upper part of the figure) detected in the present study; the last value is statistically significant for the three statistics. The Tajima's *D*-value that would correspond to Mead's data is 0.36; this corresponds to only three polymorphisms, and does not include the information provided by low-frequency polymorphisms, which would strongly decrease the three parameters (see Fig. 4).

Discussion

In an effort to better understand the pattern and the worldwide geographic structure of variation at the *PRNP* locus and the extent to which it has been shaped by natural selection, we have analyzed the sequence variation in a worldwide sample of 348 chromosomes for exon 2 of this gene. A resequencing approach has been followed, with careful manual inspection of the sequence traces and cloning of PCR products when needed. Most of the variants have been found at very low frequency, indicating a dearth of variation in the region.

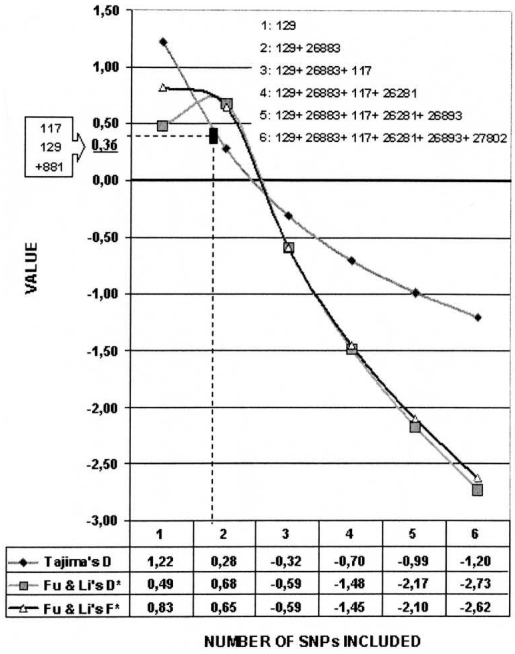
Variation at the known codons involved in susceptibility to prion diseases (129 and 219) shows a geographic stratification, even if heterogeneity (measured as F_{ST}) is not especially high compared to genome-wide values. Two other variants had been related to disease, namely, codons 142 and 232, but their presence in control populations makes them unlikely to be causative of prion diseases.

One of the major points of this study is to unravel the selective forces acting on the *PRNP* gene. No evidence has been found beyond the expected purifying selection in a phylogenetic perspective. In a study by Krakauer et al. (1998), rates of molecular evolution in prion gene have been examined by means of the d_N/d_S ratio, with no evidence for the action of positive selection on *PRNP* in primates. Moreover, Seabury et al. (2004) described purifying selection as the main force driving the evolution of the

Table 4. F_{ST} values for SNPs 129 and 219 in the seven geographical regions

F_{ST}	Sub-Saharan Africa	Middle East/ North Africa	Europe	Central/ South Asia	East Asia	Pacific	America	Global
No. of populations	6	4	7	8	6	2	5	38
2*N*	254	356	322	400	502	78	216	2128
129	0.02	0.03	−0.01	0.01	0.04	0.07	0.09	0.16
219	0.00	0.01	0.00	0.03	−0.01	0.06	0.00	0.01

SNPs are typed for the entire HGDP-CEPH panel.



	1	2	3	4	5	6
Tajima's D	1,22	0,28	-0,32	-0,70	-0,99	-1,20
Fu & Li's D*	0,49	0,68	-0,59	-1,48	-2,17	-2,73
Fu & Li's F*	0,83	0,65	-0,59	-1,45	-2,10	-2,62

Figure 4. Selection tests versus number of SNPs included in the study of Europe. We calculated the three neutrality statistics, starting with the most frequent polymorphic position (1, codon 129) and adding the remaining positions one by one in decreasing order of allele frequency, with the final calculations performed on the full set of SNPs. Comparison with a previous study is shown as a dashed line (for a *D*-value of 0.36). On the *upper right* corner are the SNPs included in this study (six). In the *upper left* box, the three SNPs found in a previous study by Mead et al. (2001, 2003) in the same genome region are shown. Only codons 129 and 117 are common in both studies.

PRNP gene in bovines, with very low d_N/d_S ratios. When comparing divergence between species with polymorphism (in the rate of synonymous vs. nonsynonymous variants) through the McDonald-Kreitman test, contrasting results have been found depending on the compared species, with no signal when comparing the closest species to humans. Thus, at the phylogenetic level, no other signs than purifying selection are evident in the evolution of *PRNP*.

Nonetheless, the forces acting on primate phylogeny may be different from the evolution acting on humans (and, as we discuss below, it may vary among human groups). In an influential paper, Mead et al. (2003) proposed that balancing selection is an important force shaping the patterns of diversity at the *PRNP* gene both locally and globally in the human species, relating it to prion diseases and the role that cannibalistic practices may have had in human history. Although in the local case (referring to the Fore) evidence seems compelling (Hedrick 2003), our analyses have revealed no evidence to support this hypothesis, neither at a global level nor for the main continental regions. The pattern of variation in *PRNP* exon 2, ascertained by sequencing, reveals frequency spectra (allelic partitions) that are not consistent with neutral expectations, but clearly in the contrary direction to balancing selection. The excess of low-frequency variants (yielding, e.g., large negative values of Tajima's *D*) is indicative of either positive selection or purifying selection upon weakly selected variants, but not balancing selection. Whereas the low number of nonsynonymous variants tends to indicate purifying selection, it has been shown that this form of selection does not yield large negative values of Tajima's *D* (Gordo et al. 2002). Values like that obtained for *PRNP* (−2.02), in the lower range of variation for humans, are mainly expected under positive selection even if the genetic effects of past population expansions act in the same direction. Other neutrality statistic tests (Fu and Li *D** and *F**, Fu's *F*) (see Table 3) gave significant negative values, consistent with this interpretation of the results.

As different local selection pressures could have influenced variation at this gene, analyses have been performed for individual geographical regions. In general, the pattern of variation is very similar to that observed for the combined worldwide sample and only the Americas show some special features, difficult to interpret. Human groups are not particularly differentiated; in general, F_{ST} values are intermediate, and even values for codon 129, which is the most ancient SNP, fall in a non-extremely high position when compared to a general F_{ST} distribution (Kidd et al. 2004). Thus, a diversifying pattern of variation is not observed. F_{ST} values are not low enough to argue for balancing selection (Bamshad and Wooding 2003). Regional variation patterns have been compared to the empirical distribution of the large data set of SeattleSNP for Africa and Europe. Neutrality tests (including Tajima's *D*) and the parameter θ_w divided by divergence, fall in the left-hand tail of the distribution, mainly for Europeans, thus excluding the effects of population history in the genetic structure of variation in the *PRNP* gene.

Ascertainment bias has been shown to have influenced the results in previous analyses of the *PRNP* gene (Mead et al. 2003). Thus, it is of great importance to have a full and careful description of all variants to properly apply neutrality tests and avoid ascertainment bias.

Two main conclusions arise from our analysis. First, we can reject the existence of an ancient, stable, balanced polymorphism of the kind that skews the frequency spectrum to an excess of intermediate frequency variants described by Mead et al. (2003). Thus, a general pattern of balancing selection, presumably related to prion diseases and cannibalism, can be rejected in human history and is shown to be due to ascertainment bias. Second, our results are consistent with a positive selection scenario, but not a simple one and are also compatible with a weak purifying selection model. Variability patterns in this gene are most likely caused by a complex history of episodic or fluctuating selection, most likely short periods of positive selection with the associated selective sweeps, followed by drift and/or purifying selection. A complex scenario (Przeworski 2002) or even occasional balancing selection episodes could also be consistent with our results.

Methods

Samples

The HGDP-CEPH Human Genome Diversity Cell Line panel contains a total of 1064 samples from a broad range of different world populations (Cann et al. 2002). From this panel, we selected a total of 174 samples (348 chromosomes) for sequencing *PRNP* exon 2, while the entire panel has been used for genotyping the SNP polymorphisms in codons 129 and 219. For sequencing, samples have been chosen to represent the following large areas and populations and have been grouped in different regions: Africa A ($2N = 28$ from Mbuti and Biaka Pygmy groups), Africa B ($2N = 40$ from San, Mandenka, and Yoruba), Europe

(2*N* = 104 from Basques, Orcadian, Sardinian, Russian, and French), Pacific (2*N* = 32 Non-Austronesian speakers in Papua New Guinea), West Asia (2*N* = 64 from Druze, Bedouin, Palestinian, and Sindhi), East Asia (2*N* = 64 from Japan, Han, Yakut, and Cambodia), and Native Americans (2*N* = 16 from Maya). One chimpanzee (from the Barcelona Zoo) has been sequenced.

SNP genotyping

Codons 129 and 219 have been genotyped using the TaqMan technology from Applied Biosystems (AB) for the entire HGDP-CEPH diversity panel. The Assays-on-Demand service was used to design probes and primers for codon 129, and the Assays-by-Design service was used to design probes and primers for codon 219. All assays have been optimized to work with TaqMan Universal PCR Master Mix and with genomic DNA (10 ng). The total volume used is 5 µL/well (384-well plate). The amplification conditions were as follows: 50°C, 2 min; 95°C, 10 min; followed by 40 cycles of 94°C, 15 sec, and 60°C, 1 min, in an ABI Prism 7900 HT (AB). Fluorescence in each well was measured after PCR, and the results have been analyzed with the SDS software package version 2.1 (AB). The success rate was 99.25% for codon 129 and 98.5% for codon 219 (Table 1).

Sequencing

All site positions referred to in this study are in accordance with a reference sequence from Lee et al. (1998), GenBank accession number U29185. The reference sequence contains a deletion in the octapeptide repeat region (R3-R4, deletion C), thus the fragment amplified is usually 24 bases longer, with a total of 2378 bp. The amplification reactions of exon 2 of the *PRNP* gene have been performed on 10 ng of template DNA in a 20-µL volume by use of AmpliTaq Gold polymerase (PE Biosystems). PCR primers and conditions are available on request. Segments have been sequenced by use of the BigDye Terminator Cycle Sequencing kit from PE Biosystems on an ABI 3700 (AB) DNA sequencer. The amplification primers have been used for cycle-sequencing reactions. All samples have been sequenced for both strands, providing an overlap between the PCR segments. The sequences have been aligned with SEQMAN II 4.03 (DNASTAR) and manually checked. The latter step turned out to be important as one (4.5%) of the 22 polymorphic sites reported in this study would not have been detected based only on automatic allele-calling software.

PCR products with sequencing problems because of heterozygous insertions or deletions (indels) have been cloned with the pMOSBlue blunt ended cloning kit (Amersham Biosciences) following the manufacturer's instructions.

Data analysis

Haplotype inferences and median–joining networks

Haplotype frequencies have been estimated using the Bayesian approach implemented in the Phase 2.0 software. Network 4.1.0.0 software has been used to generate median-joining networks describing possible genealogical relationships among haplotypes in terms of mutational differences (Bandelt et al. 1999).

Neutrality tests and diversity statistics

Several neutrality tests to detect signals of natural selection (Tajima's *D*, Fu and Li's *F** and *D**, Fay and Wu's *H* tests, and MK test) and various diversity statistics (e.g., haplotype diversity or nucleotide diversity) have been calculated using DnaSP 4.00. The significance of the tests has been estimated by means of coalescent simulations as implemented in DNAsp 4.00. These simulations have been performed with and without recombination. The recombination parameter ($R = 4N_e r$) has been set to 2.79 by considering the effective population size N_e = 10,000 (Takahata et al. 1995) and the recombination rate $r = 7.0 \times 10^{-5}$. The latter value derives from a sequenced length of 2378 bp and a local recombination estimate of 2.94 cM/Mb (Kong et al. 2002; STR D20597). The distribution of the observed pairwise nucleotide site differences and the expected values (for no recombination) in growing populations have been obtained using the programs Arlequin (Excoffier et al. 2005) and Network 4.0 (Bandelt et al. 1995).

TMRCA and age of mutations

The time to the most recent common ancestor (TMRCA) and age of mutations were estimated by means of a maxim-likelihood, coalescent-based method (Tavare et al. 1997) implemented in the GeneTree program. The GeneTree software is used to estimate the coalescence times; it generates efficiently likelihood surfaces for $\theta=4N_e\mu$ given a tree and the frequencies of the variants. These estimates have been computed assuming both a constant and an expanding population.

Substitution rate

The substitution rate in exon 2 has been estimated as the number of differences over $2tL$, L being the length of the segment compared and t the divergence time between lineages. With 26 differences with chimpanzee, the divergence is 1.09%, a similar value to that obtained for the whole genome (Chimpanzee Sequencing and Analysis Consortium 2005). If a divergence time of 6 million years is assumed between humans and chimpanzees, the average substitution rate is 0.91×10^{-9} per nucleotide and year for the studied region.

Genetic structure statistics

F_{ST}-values have been calculated using the Arlequin software (Excoffier et al. 2005). It has been performed for the sequence data (2*N* = 348) and also the data for codons 129 and 219 (2*N* = 2128) obtained using the TaqMan assays.

Acknowledgments

Most of the data have been produced in DeCode Genetics (Iceland) thanks to K. Stefánsson and A. Helgason. S. Sigurdadóttir (DeCode) helped in producing the data, and A. Helgason (DeCode) provided fruitful comments to the manuscript. Anna Di Rienzo made many useful comments on the analysis. The raw data for the F_{ST} comparisons have been kindly provided by Mark Shriver (Pennsylvania State University) and Kenneth K. Kidd (Yale University) and the divergence data for the SeattleSNP database by Deborah Nickerson (University of Washington). We also thank Oscar Lao for statistical help. The chimpanzee sample was obtained from the Barcelona Zoo under the agreement with Pompeu Fabra University. This work is supported by DGICYT (BMC2001-0772 and BOS2003-08070) and DURSI (PhD scholarship 2001FI 00632 to M.S. and grant 2001 SGR 00285).

References

Akey, J.M., Zhang, G., Zhang, K., Jin, L., Shriver, M.D. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12:** 1805–1814.

Akey, J.M., Eberle, M.A., Rieder, M.J., Carlson, C.S., Shriver, M.D., Nickerson, D.A., and Kruglyak, L. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* **2:** e286.

Andrews, N.J., Farrington, C.P., Ward, H.J., Cousens, S.N., Smith, P.G., Molesworth, A.M., Knight, R.S., Ironside, J.W., and Will, R.G. 2003. Deaths from variant Creutzfeldt-Jakob disease in the UK. *Lancet* **361:** 751–752.

Bamshad, M. and Wooding, S.P. 2003. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* **4:** 99–111.

Bandelt, H.J., Forster, P., Sykes, B.C., and Richards, M.B. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* **141:** 743–753.

Bandelt, H.J., Forster, P., and Rohl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16:** 37–48.

Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A., et al. 2002. A human genome diversity cell line panel. *Science* **296:** 261–262.

Cervenakova, L., Goldfarb, L.G., Garruto, R., Lee, H.S., Gajdusek, D.C., and Brown, P. 1998. Phenotype-genotype studies in kuru: Implications for new variant Creutzfeldt-Jakob disease. *Proc. Natl. Acad. Sci.* **95:** 13239–13241.

Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437:** 69–87.

Collinge, J., Sidle, K.C., Meads, J., Ironside, J., and Hill, A.F. 1996. Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. *Nature* **383:** 685–690.

Excoffier, L., Laval, G., and Schneider, S. 2005. Arlequin version 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1:** 47–50.

Gordo, I., Navarro, A., and Charlesworth, B. 2002. Muller's ratchet and the pattern of variation at a neutral locus. *Genetics* **161:** 835–848.

Harris, E.E. and Hey, J. 2001. Human populations show reduced DNA sequence variation at the factor IX locus. *Curr. Biol.* **11:** 774–778.

Hedrick, P.W. 2003. A heterozygote advantage. *Science* **302:** 57.

Hitoshi, S., Nagura, H., Yamanouchi, H., and Kitamoto, T. 1993. Double mutations at codon 180 and codon 232 of the PRNP gene in an apparently sporadic case of Creutzfeldt-Jakob disease. *J. Neurol. Sci.* **120:** 208–212.

Hoque, M.Z., Kitamoto, T., Furukawa, H., Muramoto, T., and Tateishi, J. 1996. Mutation in the prion protein gene at codon 232 in Japanese patients with Creutzfeldt-Jakob disease: A clinicopathological, immunohistochemical and transmission study. *Acta Neuropathol. (Berl)* **92:** 441–446.

Kidd, K.K., Pakstis, A.J., Speed, W.C., and Kidd, J.R. 2004. Understanding human DNA sequence variation. *J. Hered.* **95:** 406–420.

Kitamoto, T., Ohta, M., Doh-ura, K., Hitoshi, S., Terao, Y., and Tateishi, J. 1993. Novel missense variants of prion protein in Creutzfeldt-Jakob disease or Gerstmann-Straussler syndrome. *Biochem. Biophys. Res. Commun.* **191:** 709–714.

Kong, A., Gudbjartsson, D.F., Sainz, J., Jonsdottir, G.M., Gudjonsson, S.A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., et al. 2002. A high-resolution recombination map of the human genome. *Nat. Genet.* **31:** 241–247.

Krakauer, D.C., Zanotto, P.M., and Pagel, M. 1998. Prion's progress: Patterns and rates of molecular evolution in relation to spongiform disease. *J. Mol. Evol.* **47:** 133–145.

Kreitman, M. and Di Rienzo, A. 2004. Balancing claims for balancing selection. *Trends Genet.* **20:** 300–304.

Laplanche, J.L., Hachimi, K.H., Durieux, I., Thuillet, P., Defebvre, L., Delasnerie-Laupretre, N., Peoc'h, K., Foncin, J.F., and Destee, A. 1999. Prominent psychiatric features and early onset in an inherited prion disease with a new insertional mutation in the prion protein gene. *Brain* **122 (Pt 12):** 2375–2386.

Lee, I.Y., Westaway, D., Smit, A.F., Wang, K., Seto, J., Chen, L., Acharya, C., Ankener, M., Baskin, D., Cooper, C., et al. 1998. Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. *Genome Res.* **8:** 1022–1037.

Livingston, R.J., von Niederhausern, A., Jegga, A.G., Crawford, D.C., Carlson, C.S., Rieder, M.J., Gowrisankar, S., Aronow, B.J., Weiss, R.B., and Nickerson, D.A. 2004. Pattern of sequence variation across 213 environmental response genes. *Genome Res.* **14:** 1821–1831.

Martinez-Arias, R., Calafell, F., Mateu, E., Comas, D., Andres, A., and Bertranpetit, J. 2001. Sequence variability of a human pseudogene. *Genome Res.* **11:** 1071–1085.

Mead, S., Mahal, S.P., Beck, J., Campbell, T., Farrall, M., Fisher, E., and Collinge, J. 2001. Sporadic—but not variant—Creutzfeldt-Jakob disease is associated with polymorphisms upstream of PRNP exon 1. *Am. J. Hum. Genet.* **69:** 1225–1235.

Mead, S., Stumpf, M.P., Whitfield, J., Beck, J.A., Poulter, M., Campbell, T., Uphill, J.B., Goldstein, D., Alpers, M., Fisher, E.M., et al. 2003. Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. *Science* **300:** 640–643.

Palmer, M.S., Dryden, A.J., Hughes, J.T., and Collinge, J. 1991. Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease. *Nature* **352:** 340–342.

Prusiner, S.B. 1991. Molecular biology of prion diseases. *Science* **252:** 1515–1522.

Przeworski, M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* **160:** 1179–1189.

Ptak, S.E. and Przeworski, M. 2002. Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet.* **18:** 559–563.

Rogers, A.R. and Harpending, H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9:** 552–569.

Seabury, C.M., Honeycutt, R.L., Rooney, A.P., Halbert, N.D., and Derr, J.N. 2004. Prion protein gene (*PRNP*) variants and evidence for strong purifying selection in functionally important regions of bovine exon 3. *Proc. Natl. Acad. Sci.* **101:** 15142–15147.

Shibuya, S., Higuchi, J., Shin, R.W., Tateishi, J., and Kitamoto, T. 1998. Codon 219 Lys allele of PRNP is not found in sporadic Creutzfeldt-Jakob disease. *Ann. Neurol.* **43:** 826–828.

Soldevila, M., Calafell, F., Andres, A.M., Yague, J., Helgason, A., Stefansson, K., and Bertranpetit, J. 2003. Prion susceptibility and protective alleles exhibit marked geographic differences. *Hum. Mutat.* **22:** 104–105.

Soldevila, M., Calafell, F., Helgason, A., Stefansson, K., and Bertranpetit, J. 2005. Assessing the signatures of selection in PRNP from polymorphism data: Results support Kreitman and Di Rienzo's opinion. *Trends Genet.* **21:** 389–391.

Stephens, J.C., Schneider, J.A., Tanguay, D.A., Choi, J., Acharya, T., Stanley, S.E., Jiang, R., Messer, C.J., Chew, A., Han, J.H., et al. 2001. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293:** 489–493.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585–595.

Takahata, N., Satta, Y., and Klein, J. 1995. Divergence time and population size in the lineage leading to modern humans. *Theor. Pop. Biol.* **48:** 198–221.

Tavare, S., Balding, D.J., Griffiths, R.C., and Donnelly, P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* **145:** 505–518.

Thompson, E.E., Kuttab-Boulos, H., Witonsky, D., Yang, L., Roe, B.A., and Di Rienzo, A. 2004. CYP3A variation and the evolution of salt-sensitivity variants. *Am. J. Hum. Genet.* **75:** 1059–1069.

Valleron, A.J., Boelle, P.Y., Will, R., and Cesbron, J.Y. 2001. Estimation of epidemic size and incubation time based on age characteristics of vCJD in the United Kingdom. *Science* **294:** 1726–1728.

Wooding, S., Kim, U.K., Bamshad, M.J., Larsen, J., Jorde, L.B., and Drayna, D. 2004. Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. *Am. J. Hum. Genet.* **74:** 637–646.

Received June 27, 2005; accepted in revised form November 7, 2005.

The prion protein gene in humans revisited: Lessons from a worldwide resequencing study

Marta Soldevila, Aida M. Andrés, Anna Ramírez-Soriano, et al.

Genome Res. 2006 16: 231-239
Access the most recent version at doi:10.1101/gr.4345506