

Qin Zan · Bo Wen · Yungang He · Yi Wang  
Shuhua Xu · Ji Qian · Daru Lu · Li Jin

## Complete sequence data support lack of balancing selection on PRNP in a natural Chinese population

Received: 17 September 2005 / Accepted: 23 January 2006 / Published online: 25 March 2006  
© The Japan Society of Human Genetics and Springer-Verlag 2006

**Abstract** The M129V mutation in the human prion protein gene (PRNP) is the primary site linked to susceptibility to prion diseases in humans. The heterozygous state of this allele has been proven to be more resistant to prion-related diseases such as Creutzfeldt–Jakob disease. Based on a study of the common genetic variations, it has been proposed that balancing selection has played a critical role in shaping the distribution of 129V in worldwide populations. Kreitman and Di-Rienzo [Kreitman and Di-Rienzo (2004) *Trend Genet* 20:300–304] challenged this hypothesis by pointing out that the exclusion of polymorphisms with low frequency may introduce an ascertainment bias and, in turn, lead to a wrong conclusion. By studying sequence variations in samples from the Human Genome Diversity Project, Soldevila et al. [Soldevila et al. (2005) *Trends Genet* 21:389–391] showed that this ascertainment bias does exist. We argue that the evidence presented by Soldevila et al. may be compromised by the samples tested since they were selected from a population that may be sub-

structured. In this study, we re-evaluated the hypothesis of balancing selection in a natural Chinese population using a much longer segment encompassing the entire genomic region of the PRNP gene (15 kb). We showed that the pattern of genetic variation in PRNP is not consistent with the presence of balancing selection in this gene.

**Keywords** Chinese · Balancing selection · PRNP M129V

### Introduction

Human prion protein, when conformationally transformed into another state, is responsible for Creutzfeldt–Jakob disease (CJD) and other prion diseases. Among detected mutations in PRNP, the gene encoding this protein, a methionine to valine mutation at the 129th amino acid (M129V) was found to be prevalent in all ethnic groups. M129V heterozygotes have been shown to be intrinsically protected from complete induction of the prion conformation change and therefore might be resistant to the disease (Rujescu et al. 2002; Lee et al. 2001).

By studying the common genetic variations in a 4.3 kb segment flanking M129V, Mead et al. (2003) showed positive Tajima's *D* (Tajima 1989) in worldwide populations using a selected group of variations in PRNP. They proposed that balancing selection played a critical role in shaping the worldwide distribution of 129V, including Fore-speaking New Guinean, African, European, and East Asian populations. In addition, it has been shown that the frequency of 129V in East Asians is low (Erginel-Unaltuna et al. 2001; Jeong et al. 2004; Yu et al. 2004) and the Tajima's *D* is also lower, though positive, in East Asians than in other populations (Mead et al. 2003).

Kreitman and Di-Rienzo (2004) challenged the conclusions of Mead et al. 2003 by pointing out that the exclusion of polymorphisms with low frequency may

Q. Zan · B. Wen · Y. He · Y. Wang · S. Xu · J. Qian  
D. Lu · L. Jin

State Key Laboratory of Genetic Engineering and Center for Anthropological Studies, School of Life Sciences and Morgan-Tan International Center for Life Sciences, Fudan University, Shanghai, People's Republic of China

L. Jin  
CAS-MPG Partner Institute of Computational Biology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai, People's Republic of China

L. Jin  
Center for Genome Information, Department Environmental Health, University of Cincinnati, Cincinnati, OH, USA

L. Jin (✉)  
Institute of Genetics, Fudan University,  
220 Handan Rd., 200433 Shanghai,  
People's Republic of China  
E-mail: ljin007@gmail.com  
Tel.: +86-21-65642800

introduce ascertainment bias, which may tilt the estimation of the allele-frequency spectrum towards the strongly positive Tajima's  $D$  values seen in Mead's work. By analyzing a 2.5 kb segment of PRNP in samples selected from Human Genome Diversity Project [Centre d'Etude du Polymorphisme Humain (HGDP-CEPH); Cann et al. 2002; Soldevila et al. 2005] Kreitman and Di-Rienzo (2004) showed that Tajima's  $D$ 's are generally negative in worldwide populations, and that the smallest  $D$  was found in East Asians. However, their conclusion may be compromised by the fact that the East Asian individuals in their study were sampled from multiple populations with different ethnic backgrounds, and the presence of population substructures may lead to a biased estimation of  $D$  values from neutrality expectation (Simonsen et al. 1995; Hammer et al. 2003).

In this report, we reevaluated the presence of balancing selection on PRNP based on sequence data from the entire genomic region (15 kb) in a large sample of natural Han Chinese (92 chromosomes) from Boxing County, Shandong, PR China. A negative Tajima's  $D$  was observed when all genetic variations were included in the analysis, contradicting the hypothesis of balancing selection on PRNP. Furthermore, a systematic population study in ten Han Chinese populations validates the previous observation of low 129V frequency in East Asians.

## Materials and methods

Genomic DNA was extracted by standard techniques from the blood of 436 unrelated healthy donors sampled from Shanghai, Shandong, Liaoning, Yunnan, Sichuan, Xinjiang, and Beijing. Genotyping of the codon 129 polymorphism was conducted by PCR-RFLP using the enzyme *Bsa*AI. We sequenced the entire genomic locus of the PRNP gene (GenBank: AL133396). Overall, 46 individuals from Boxing County, Shandong were selected for sequencing using an ABI PRISM 3100 Genetic Analyzer (Foster City, CA). Primer sequences will be provided upon request. We used Tajima's  $D$  test (Tajima 1989) and Fu and Li's test (Fu and Li 1993) to examine departure from the assumption of neutral evolution.

## Results

Three (0.6%) heterozygotes and 433 (99.4%) homozygotes for the allele 129M were observed in 436 Chinese samples from ten populations. The frequency of allele 129V was about 0.3% in Chinese. The frequencies of genotypes and of 129V in the populations tested are presented in Table 1.

We sequenced a 15 kb-segment encompassing the entire PRNP gene in 46 samples from Boxing County, Shandong. A total of eight polymorphic sites was found, including M129V. Unlike M129V, the other seven variants are all located in non-coding regions. These eight PRNP variations are in complete linkage disequilibrium and therefore form a perfect haplotype block, with all pairwise  $|D'|$  values equal to 1. Overall, eight haplotypes were observed and their frequencies were estimated using Phase 2.0 (Stephens et al. 2001) and are listed in Table 2.

The hypothesis of balancing selection is characterized by a deep split in the genealogy and an excess of high-frequency polymorphisms flanking the sites being selected (Mead et al. 2003). Using Tajima's  $D$  (Tajima 1989) and a related statistic, Fu and Li's  $D$  (Fu and Li 1993), an excess of high-frequency polymorphisms results in large positive values of the statistics and is indicative of balancing selection. The Tajima's  $D$  value (Tajima 1989) in the Boxing County population was  $-1.78$  ( $0.10 > P > 0.05$ ), and Fu and Li's  $D$  (Fu and Li 1993) was  $-1.92$  ( $0.10 > P > 0.05$ ), both inconsistent with the hypothesis of balancing selection, in which positive  $D$  is expected. In contrast, the negative  $D$  values may reflect the roles of positive selection and population expansion although the observation is not statistically significant. In addition, the possible presence of population structure in these populations can be excluded based on the lack of deviation from Hardy-Weinberg expectation at all eight sites studied (data not shown).

The reduced median network of the haplotypes was inferred using NETW4101 (Forster et al. 2001) (<http://www.fluxus-engineering.com>) (see Fig. 1). In Fig. 1, the area of each circle is proportional to the frequency of the haplotype it represents. The haplotype carrying the 129V allele (haplotype G) was derived from the M-car-

**Table 1** Frequency of human PRNP M129V polymorphism in ten Chinese populations

Population (no. of individuals)	Frequency (%) MM (%)	Frequency (%) MV (%)	Frequency (%) VV (%)	Frequency (%) V (%)
Shanghai suburb (49)	49 (100)	0 (0)	0 (0)	0
Shandong Qingdao (49)	49 (100)	0 (0)	0 (0)	0
Shandong Weihai (49)	49 (100)	0 (0)	0 (0)	0
Liaoning Dalian (30)	30 (100)	0 (0)	0 (0)	0
Yunnan (49)	49 (100)	0 (0)	0 (0)	0
Sichuan (49)	49 (100)	0 (0)	0 (0)	0
Shandong Jinan (49)	46 (93.9)	3 (6.1)	0 (0)	3.1
Xinjiang (30)	30 (100)	0 (0)	0 (0)	0
Beijing (36)	36 (100)	0 (0)	0 (0)	0
Shandong Boxing (46)	45 (97.8)	1 (2.2)	0 (0)	1.1

**Table 2** PRNP haplotype frequencies in Boxing County population

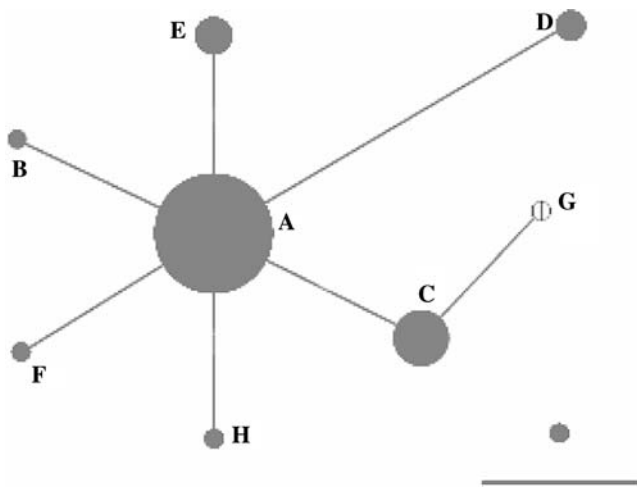
Haplotype (%)	Polymorphism sites and surrounding sequences							
	-8,489 GAGA (A/G)CTTT	-7,952 ACCA (A/T)TGAC	-3,179 CTTC (A/G)TCAC	-2,660 CAAA (C/T)AGTG	-2,498 TACC (A/G)TATA	385 CTAC (A/G)TGCT <sup>a</sup>	930 AGTA (A/G)ACCT	978 AGCC (A/G)TTGC
A (80.4%)	T	A	G	T	A	A	G	G
B (1.1%)	T	T	G	T	A	A	G	G
C (9.8%)	T	A	A	T	A	A	G	G
D (2.2%)	T	A	G	C	G	A	G	G
E (3.3%)	G	A	G	T	A	A	G	G
F (1.1%)	T	A	G	T	A	A	A	G
G (1.1%)	T	A	A	T	A	G	G	G
H (1.1%)	T	A	G	T	A	A	G	A

<sup>a</sup>Site 385 represents M129V

rying haplotype C by a single mutation (M129V). A deep split between the 129V-carrying and 129M-carrying lineages, as expected under the assumption of balancing selection, was not observed in our samples. Haplotype A is the most frequent in Chinese and is probably the founding haplotype in East Asian populations. The star-like structure of the genealogy of M-carrying lineages and the negative  $D$  statistics are consistent with population expansion in East Asia. We also examined the mismatch distribution of PRNP haplotypes in our samples (data not shown) and again it does not support the presence of balancing selection as shown by Mead et al. (see Supporting Online Material Fig. S1 in Mead et al. 2003).

## Discussion

Since Tajima's  $D$  and other related statistics for detecting the presence of selection are based on the



**Fig. 1** Reduced median network of PRNP haplotypes. The size of the circle is proportional to the frequency of that haplotype within the sample. Haplotype G represents the haplotype with the 129V allele, while the other circles represent haplotypes with 129M alleles. Each line connecting two haplotypes represents mutation steps. The *circle* and the *line* in the lower right of the figure indicate the size that represents one individual and one step distance, respectively

frequency spectrum of all genetic variations, Kreitman and Di-Rienzo (2004) pointed out that an analysis of a subset of variations may constitute an ascertainment bias that may lead to a wrong conclusion, as exemplified by the study of Mead et al. (2003) on PRNP in which only more frequent variations were included. This notion was supported by a study of a selection of PRNP variations by Soldevila et al. (2005) in which all variations within a 2.4 kb region were included. In the latter study, Tajima's  $D$ 's are generally negative and the smallest  $D$  was found in the East Asian population. However, the samples used in the study of Soldevila et al. (2005) were obtained from the HGDP-CEPH collection (Soldevila et al. 2005). It should be noted that the sample sizes for each population in the collection are small (<http://www.cephb.fr/HGDP-CEPH-Panel>) especially from East Asians (ten lymphoblastoid cell lines in most populations). Pooling of samples from different ethnic groups may generate a substructured population in which negative  $D$  values (Hammer et al. 2003) can be observed. Therefore, using the data of Soldevila et al. (2005) to support the notion of Kreitman and Di-Rienzo (2004) requires careful examination.

In this study, we showed a negative Tajima's  $D$  value in a 15 kb segment encompassing the entire genomic region of PRNP in a natural population. Our result therefore constitutes stronger support of Kreitman's notion since (1) we studied a natural population instead of a substructured sample, (2) our sample size is larger, and (3) a much larger segment was surveyed. However, our observation does not preclude the possible absence of a driving force for balancing selection in East Asia, although the marginally significant presence of balancing selection suggested by Mead et al. (2003) underscores the necessity of using larger samples and all segregating sites in the analysis.

We also examined the possibility of selection using several different approaches including Fu and Li's  $D$  (Fu and Li 1993) and reduced median networks (Forster et al. 2001). In either case, our observations are inconsistent with the hypothesis of balancing selection, at least in Chinese.

**Acknowledgements** This study was supported by Chinese High-Tech Program (2001AA224021 and 2002BA711A10) and Shanghai Science and Technology Development Fund (00DJ14003).

---

## References

- Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL (2002) A human genome diversity cell line panel. *Science* 296:261–262
- Erginel-Unaltuna N, Peoc'h K, Komurcu E, Acuner TT, Issever H, Laplanche JL (2001) Distribution of the M129V polymorphism of the prion protein gene in a Turkish population suggests a high risk for Creutzfeldt–Jakob disease. *Eur J Hum Genet* 9:965–968
- Forster P, Torroni A, Renfrew C, Röhl A (2001) Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* 18:1864–1881
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709
- Hammer MF, Blackmer F, Garrigan D, Nachman MW, Wilder JA (2003) Human population structure and its effects on sampling Y chromosome sequence variation. *Genetics* 164:1495–1509
- Jeong BH, Nam JH, Lee YJ, Lee KH, Jang MK, Carp RI, Lee HD, Ju YR, Jo SA, Park KY, Kim YS (2004) Polymorphisms of the prion protein gene (PRNP) in a Korean population. *J Hum Genet* 49:319–324
- Kreitman M, Di Rienzo A (2004) Balancing claims for balancing selection. *Trend Genet* 20:300–304
- Lee HS, Brown P, Cervenakova L, Garruto RM, Alpers MP, Gajdusek DC, Goldfarb LG (2001) Increased susceptibility to kuru of carriers of the PRNP 129 methionine/methionine genotype. *J Infect Dis* 183:192–196
- Mead S, Stumph MPH, Whitfield J, Beck JA, Poulter M, Campbell T, Uphill JB, Goldstein D, Alpers M, Fisher EMC, Collinge J (2003) Balancing selection at the prion protein gene consistent with prehistoric kuru-like epidemics. *Science* 300:640–643
- Rujescu D, Meisenzahl EM, Giegling I, Kirner A, Leinsinger G, Hegerl U, Hahn K, Möller HJ (2002) Methionine homozygosity at codon 129 in the prion protein is associated with white matter reduction and enlargement of CSF compartments in health volunteers and schizophrenic patients. *NeuroImage* 15:200–206
- Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141:413–429
- Soldevila M, Calafell F, Helgason A, Stefansson K, 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 M, Smith N, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Yu SL, Jin L, Sy MS, Mei FH, Kang SL, Sun GH, Tien P, Wang FS, Xiao GF (2004) Polymorphisms of the PRNP gene in Chinese populations and the identification of a novel insertion mutation. *Eur J Hum Genet* 12:867–870