A *Trans*-Specific Polymorphism in *ZC3HAV1* Is Maintained by Long-Standing Balancing Selection and May Confer Susceptibility to Multiple Sclerosis

R. Cagliani,[1] F. R. Guerini,[2] M. Fumagalli,[1] S. Riva,[1] C. Agliardi,[2] D. Galimberti,[3] U. Pozzoli,[1] A. Goris,[4] B. Dubois,[4] C. Fenoglio,[3] D. Forni,[1] S. Sanna,[5] I. Zara,[6] M. Pitzalis,[7] M. Zoledziewska,[7] F. Cucca,[5,7] F. Marini,[1] G. P. Comi,[3] E. Scarpini,[3] N. Bresolin,[1,3] M. Clerici,†[,8,9] and M. Sironi†*[,1]

[1]Scientific Institute IRCCS E. Medea, Bosisio Parini, Lecco, Italy

[2]Laboratory of Molecular Medicine and Biotechnologies, Don C. Gnocchi Foundation ONLUS, IRCCS, Milano, Italy

[3]Dino Ferrari Centre, Department of Neurological Sciences, University of Milan, Fondazione Ca' Granda IRCCS Ospedale Maggiore Policlinico, Milan, Italy

[4]Laboratory for Neuroimmunology, Department of Neurosciences, Katholieke Universiteit Leuven, Leuven, Belgium

[5]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy

[6]CRS4, Advanced Genomics Computing Technology, Parco tecnologico della Sardegna, Pula, Cagliari, Italy

[7]Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy

[8]Chair of Immunology, Department of Biomedical Sciences and Technologies LITA Segrate, University of Milano, Milano, Italy

[9]Fondazione Don C. Gnocchi, IRCCS, Milano, Italy

†These authors contributed equally to this work.

***Corresponding author:** E-mail: manuela.sironi@bp.lnf.it.

Associate editor: Anne Stone

Research article

Abstract

The human *ZC3HAV1* gene encodes an antiviral protein. The longest splicing isoform of *ZC3HAV1* contains a C-terminal PARP-like domain, which has evolved under positive selection in primates. We analyzed the evolutionary history of this same domain in humans and in *Pan troglodytes*. We identified two variants that segregate in both humans and chimpanzees; one of them (rs3735007) does not occur at a hypermutable site and accounts for a nonsynonymous substitution (Thr851Ile). The probability that the two *trans*-specific polymorphisms have occurred independently in the two lineages was estimated to be low ($P = 0.0054$), suggesting that at least one of them has arisen before speciation and has been maintained by selection. Population genetic analyses in humans indicated that the region surrounding the shared variants displays strong evidences of long-standing balancing selection. Selection signatures were also observed in a chimpanzee population sample. Inspection of 1000 Genomes data confirmed these findings but indicated that search for selection signatures using low-coverage whole-genome data may need masking of repetitive sequences. A case–control study of more than 1,000 individuals from mainland Italy indicated that the Thr851Ile SNP is significantly associated with susceptibility to multiple sclerosis (MS) (odds ratio [OR] = 1.47, 95% confidence intervals [CI]: 1.08–1.99, $P = 0.011$). This finding was confirmed in a larger sample of 4,416 Sardinians cases/controls (OR = 1.18, 95% CI: 1.037–1.344, $P = 0.011$), but not in a population from Belgium. We provide one of the first instances of human/chimpanzee *trans*-specific coding variant located outside the major histocompatibility complex region. The selective pressure is likely to be virus driven; in modern populations, this variant associates with susceptibility to MS, possibly via the interaction with environmental factors.

Key words: *ZC3HAV1*, *trans*-specific polymorphism, balancing selection, multiple sclerosis.

Introduction

The zinc-finger antiviral protein (ZAP) was originally identified from a rat cDNA library due to its conferring resistance to murine leukemia viruses (Gao et al. 2002). Subsequent studies indicated that ZAP also inhibits several viruses of the alphaviridae (Bick et al. 2003; Zhang et al. 2007) and filoviridae (Muller et al. 2007) families. ZAP acts through direct binding to the viral RNA and recruits the processing exosome, eventually leading to viral RNA degradation (Guo et al. 2004, 2007).

The human ortholog of ZAP is encoded by *ZC3HAV1*, an interferon-inducible gene located on chromosome 7 (7q34). The gene codes for two major isoforms generated by alternative splicing (Kerns et al. 2008). The products share a common N-terminus carrying four CCCH zinc-finger motifs, although only the longest isoform displays a carboxy-terminal poly(ADP-ribose) polymerase (PARP)–like domain. The CCCH zinc-finger motifs are directly involved in viral RNA binding; conversely, the precise function of the PARP-like domain is unknown, but it enhances the activity of ZC3HAV1 against alphaviruses (Kerns et al. 2008).

A recent evolutionary study in primates indicated that the PARP-like domain, but not the CCCH zinc-finger domains, has been subjected to recurrent episodes of positive selection (Kerns et al. 2008). These data, albeit unexpected given the role of the N-terminal zinc fingers in viral recognition (Guo et al. 2004), are consistent with the notion whereby genes involved in immune response have been common targets of natural selection along primate and mammalian history (Barreiro and Quintana-Murci 2010).

Viruses have exerted a selective pressure on several human genes, and a certain degree of overlap has been noticed between genes targeted by virus-driven selective pressure and loci involved in the pathogenesis of autoimmune diseases including multiple sclerosis (MS) (Fumagalli et al. 2010), a demyelinating autoimmune disease of the central nervous system. This parallelism and the observation that a subset of alleles for such diseases represent selection targets (Fumagalli, Pozzoli, et al. 2009; Barreiro and Quintana-Murci 2010) contribute to a long-term debate concerning the underlying selective patterns and pressures responsible for the maintenance of autoimmune susceptibility variants in human populations (Barreiro and Quintana-Murci 2010; Sironi and Clerici 2010).

Here, we analyzed the selective pattern of the *ZC3HAV1* PARP-like domain in human populations and in *Pan troglodytes*. Our data indicate that long-term balancing selection has maintained a human/chimpanzee *trans*-specific Thr/Ile polymorphism in *ZC3HAV1*, and this variant is associated with higher risk of MS in Italian populations.

Materials and Methods

Samples and Sequencing

Human genomic DNA from HapMap subjects (20 Yoruba, YRI, 20 European, CEU, and 20 Asians, AS; 24 and 15 African American, AA, and South American, SAI individuals, respectively) was obtained from the Coriell Institute for Medical Research. All analyzed regions were polymerase chain reaction (PCR) amplified and directly sequenced. PCR products were treated with ExoSAP-IT (USB Corporation, Cleveland OH), directly sequenced on both strands with a Big Dye Terminator sequencing Kit (v3.1 Applied Biosystems) and run on an Applied Biosystems ABI 3130 XL Genetic Analyzer (Applied Biosystems). Sequences were assembled using AutoAssembler version 1.4.0 (Applied Biosystems) and inspected manually by two distinct operators. The genomic DNA of 9 *P. troglodytes* was obtained from the Gene Bank of Primates, Primate Genetics, Germany (http://dpz.eu/index.php).

Chimpanzee subspecies was determined by the analysis of mitochondrial DNA (mtDNA) and Y chromosome as described (Stone et al. 2002) as well as by application of STRUCTURE analysis (Pritchard et al. 2000) to resequenced autosomal regions. In particular, the mtDNA hypervariable region I was sequenced for all individuals using previously described primer pairs (Morin et al. 1994) and compared with published sequences (Morin et al. 1994) to infer the maternal origin of each individual. Three regions (sY19, sY84, and sY123) on the nonrecombining portion of the Y chromosome (Stone et al. 2002) were also sequenced for the three male chimpanzee individuals in our sample and compared with previous data from *P. troglodytes* subspecies (Stone et al. 2002). Since the paternal origin of the six females in our sample could not be determined, we used data from 26 autosomal biallelic markers for STRUCTURE analysis (Pritchard et al. 2000). In particular, one SNP was randomly selected for each of the chromosomal genomic regions we resequenced in *P. troglodytes* ($n = 16$). Ten additional variants were included and derived from genotyping of these same individuals at ten genes (one SNP per gene in *PTPN22*, *CD4*, *IFIH1*, *LY9*, *AICDA*, *PON1*, *OAS1*, *SFTPD*, *CTLA4*, and *PIAS1*), which we are currently analyzing in this species (manuscript in preparation). STRUCTURE was used in the “admixture” mode so that individuals are allowed to have ancestry from multiple populations. We used a model of correlated allele frequencies with 100,000 burn-in iterations and 1,000,000 follow-on Markov chain Monte Carlo iterations. The program was run 20 times for each value of K ranging from 1 to 4, and for each run, the posterior probability of $K = 1, 2, 3$, or 4 was calculated. Results of these runs indicated that a model with $K = 1$ has much higher probability compared with the other values of K (supplementary fig. S1, Supplementary Material online).

All chimpanzee gene regions were resequenced as described above. All primers used in the study are available as supplementary table S1 (Supplementary Material online).

Case/Control Association Study

The case/control population from mainland Italy comprised 507 patients (333 females and 174 males) and 523 age- and sex-matched healthy individuals (333 females and 190 males) with the same geographic origin. MS patients were recruited at the MS Centre of Don Gnocchi Foundation in Milan and at Department of Neurological Sciences, University of Milan. All patients gave informed consent according to protocols approved by the local Ethics Committees. All patients and controls were Italian of European ancestry. Patients underwent a standard battery of examinations, including medical history, physical and neurological examination, screening laboratory test, brain magnetic resonance imaging. All MS subjects fulfilled McDonald’s criteria (McDonald et al. 2001). Median age was 40.5 (standard deviation [SD]: 11.1) and 42.2 (SD: 20.5) years for MS and controls, respectively. Genotyping of rs3735007 was performed by direct resequencing, as described above, using genomic DNA extracted from peripheral blood.

The case/control populations from Sardinia consist of 2,273 patients and 1,917 healthy individuals recruited from all over the island and for whom at least three of the four grandparents were of Sardinian origin (Sanna et al. 2010). Patients were all of the relapsing–remitting form of MS, and they have been diagnosed and selected using the McDonald criteria (McDonald et al. 2001). The female to male ratio was equal to 2.1:1 in the assessed samples. The control group of healthy individuals is composed of 1,917 adult volunteer blood donors, unrelated to each other and to the patients and also 231 affected family-based controls (AFBAC) derived from 231 MS family trios. AFBAC allele and haplotype frequencies are constructed using the two alleles in each trio that are not transmitted from the parents to the affected child. These familial pseudocontrols are matched to the cases for ethnic origin and are thus robust to population stratification (Thomson 1995). All individuals studied and all analyses on their samples were done according to the Declaration of Helsinki and were approved by the local medical ethics and institutional review committees. Individuals were genotyped using the Affymetrix 6.0 array, and standard quality control filters were applied (single nucleotide polymorphism [SNP] call rate $>$ 98%, Hardy–Weinberg equilibrium, $\mathrm{HWE}_{P\ \mathrm{value}} > 10^{-6}$, minor allele frequency $>$ 1%, lack of excess of Mendelian errors or discrepancies in genotyped trios or DNA duplicates) (Sanna et al. 2010). Among the genotyped and quality control assessed SNPs, rs1047129 marker showed the highest linkage disequilibrium (LD) with rs3735007 and was used for replication. At this variant, genotypes were successfully called for 2,271 patients, 1,915 controls and were unambiguously determined for 231 AFBAC.

As for the Belgian sample, patients of European descent fulfilled Poser or McDonald criteria (McDonald et al. 2001; Poser 2006) for the diagnosis of MS and were compared with controls from the same population. All individuals gave appropriate informed consent, and the study was approved by the Ethics Committee of the University Hospitals Leuven. Genotyping of rs3735007 was performed using a Taqman Assay-on-Demand on a 7300 Sequence Detection System (Applied Biosystems) according to the manufacturer's instructions.

Compliance to HWE was evaluated by a χ^2 test. The association of SNP genotypes with disease was calculated using PLINK (Purcell et al. 2007).

The Wittke-Thompson test for HWE deviation was performed as described by the authors (Wittke-Thompson et al. 2005). Equations are parametrized in q (susceptibility allele frequency), α (risk in nonsusceptible homozygotes), β (heterozygote relative risk), γ (homozygote relative risk), and K_p (trait prevalence in the general population). We obtained maximum likelihood (ML) estimates for these parameters minimizing the goodness-of-fit test statistic (as reported in Wittke-Thompson et al. (2005)) using the Broyden-Fletcher-Goldfarb-Shanno method. Using an estimate of K_p, the procedure was repeated with a general model estimating q, β, and γ, and for constrained specific models, estimating q and gamma (dominant: beta $=$ γ; recessive: $\beta = 1$, $\gamma > 1$; additive: $\beta = (\gamma + 1)/2$, $\gamma > 1$; and multiplicative: $\beta = \mathrm{sqrt}(\gamma)$, $\gamma > 1$). Given the different number of parameters in the general model, the Akaike information criteria were used for the best-fit model selection. A P value was then calculated for the minimal value of the test statistic using a χ^2 distribution with 1 or 2 degrees of freedom (df) for the general and constrained models, respectively.

Calculations were carried out in the R environment (R Development Core Team 2008).

Population Genetic Analyses

Tajima's D (D_T) (Tajima 1989), Fu and Li's D^* and F^* (Fu and Li 1993) statistics as well as diversity parameters θ_W (Watterson 1975) and π (Nei and Li 1979) were calculated using libsequence (Thornton 2003). Calibrated coalescent simulations were performed using the cosi package (Schaffner et al. 2005) and its best-fit parameters for YRI, AA, CEU, and AS populations with 10,000 iterations. For SAI, a previously reported demographic model (Ray et al. 2010) was used and included in the cosi best-fit model. Demographic parameters for YRI, AA, CEU, and AS implemented in cosi are described in Schaffner et al. (2005). Coalescent simulation was also run using additional demographic models (Marth et al. 2004; Voight et al. 2005; Gutenkunst et al. 2009). In all cases, they were conditioned on mutation and recombination rates. Estimates of the population recombination rate parameter ρ were obtained from resequencing data with the use of the Web application MAXDIP (http://genapps.uchicago.edu/maxdip/) (Hudson 2001) and converted in centimorgan per megabase. The lower recombination rate obtained from the five populations in each region was used in coalescent simulations; these amounted to 0.31 cM/Mb for ZC3HAV1 exons 10–12 and 0.0025 cM/Mb for the region covering exon 13. P values of summary statistics for the analyzed ZC3HAV1 regions were obtained by comparing the observed D_T, D^* and F^* with the distribution in the 10,000 simulated genealogies.

The ML-ratio HKA test was performed using the MLHKA software (Wright and Charlesworth 2004), as previously proposed (Fumagalli, Cagliani, et al. 2009). For human populations, 16 reference loci were randomly selected among loci included in the National Institute of Environmental Health Sciences (NIEHS) Genome Project, that are shorter than 20 kb and that have been resequenced in the four populations; the only criterion was that Tajima's D values were consistent with neutrality (i.e., Tajima's D is higher than the fifth and lower than the 95th percentiles in the distribution of NIEHS genes). As for chimpanzee, the MLHKA was performed using the 16 resequenced regions as reference loci.

Genotype data for 2 kb regions from 238 resequenced human genes were derived from the NIEHS SNPs Program web site (http://egp.gs.washington.edu). In particular, we selected genes that had been resequenced in populations of defined ethnicity including Europeans, Yoruba, African American, and Asians (NIEHS panel 2). Similarly, genotype data for 5-kb windows were obtained from the NIEHS Program web site; in this case, windows were randomly selected with the only requirement that they contain at least five segregating sites in all analyzed populations.

Data from the Pilot 1 phase of the 1000 Genomes Project were retrieved from the dedicated web site (http://www.1000genomes.org/). Low-coverage SNP genotypes were organized in a MySQL database. A set of programs was developed to retrieve genotypes from the database and to analyze them according to selected regions/populations. These programs were developed in C++ using the GeCo++ (Cereda et al. 2011) and the libsequence (Thornton 2003) libraries.

Genotype information was obtained for *ZC3HAV1* and for 2,000 randomly selected RefSeq genes.

Sliding-window analysis was performed on overlapping 2.5-kb windows moving with a step of 250 bp. For each window, we calculated θ_W divided by the total number of fixed differences and π. As for site frequency spectrum–based statistics, these were calculated using low-coverage data for the two *ZC3HAV1* gene regions and for 2,000 windows randomly derived from the genes mentioned above. These windows were 2 kb in size and were used to obtain the distribution of D_T, F^* and D^* in the three human populations. Masking of transposable elements was performed through University of California–San Cruz annotation tables (Repeating Elements by RepeatMasker track).

For the comparison of nucleotide diversity estimates derived from Sanger resequencing and low-coverage whole-genome data, we analyzed all genes included in the NIEHS panel 2 ($n = 238$). Fully resequenced (i.e., without resequencing gaps) genic regions were identified and divided into continuous subregions depending on their annotation in the NIEHS data as unique DNA or transposable elements. This procedure yielded a total of 5,271 unique regions (average size = 659 bp) and 5,192 repetitive regions (average size = 249 bp). θ_W was calculated as described above for YRI, CEU, and AS using both NIEHS and 1000 Genomes sequencing data. Correlations between nucleotide diversity estimates were calculated using Kendall's correlation coefficients (τ), a nonparametric statistic used to measure the degree of correspondence between two rankings. The reason for using this test is that even in the presence of ties, the sampling distribution of τ satisfactorily converges to a normal distribution for values of n larger than 10 (Salkind 2007). A normal approximation with continuity correction to account for ties was used for *P* value calculations (Kendall 1976).

Haplotype Analysis and Time to the Most Recent Common Ancestor Calculation

Haplotypes were inferred using PHASE version 2.1 (Stephens et al. 2001; Stephens and Scheet 2005). In order to check for consistency, 100, 250, and 500 iterations were performed with the last runs ten times longer than other runs (-X10 flag) to increase estimate accuracy. Haplotypes for individuals resequenced in this study are available as supplementary table S2 (Supplementary Material online). LD analyses were performed using the Haploview (v. 4.1) (Barrett et al. 2005) using SNP data from HapMap and default parameters ($\mathrm{HWE}_{P\ \mathrm{value}}$ cutoff = 0.001, minimum genotype percentage = 75%, maximum number of Mendel errors = 1, minor minimum allele frequency = 0.001). Haplotype blocks were identified through the confidence interval method (Gabriel et al. 2002). The significance of LD was assessed by means of a standard chi-square statistic as implemented in graphical overview of linkage disequilibrium (GOLD) (Abecasis and Cookson 2000) and Bonferroni corrected for the number of tested SNP pairs. GOLD groups low-frequency alleles to avoid spurious results: We used the default grouping strategy at 7%. An additional analysis of recombination was performed using the LDhat program (McVean et al. 2002), which uses the composite likelihood method of Hudson (Hudson 2001), extended to finite-sites models, to calculate the population recombination rate ρ ($4N_er$) given an estimate of θ_W. The program then uses a likelihood permutation test to address the presence of recombination and identifies SNP pairs that display more or less LD than expected assuming a constant recombination rate over the entire region. Rare variants (i.e., with minor allele frequency lower than 5%) are omitted from the analysis as uninformative.

Median-joining networks to infer haplotype genealogy were constructed using NETWORK 4.5 (Bandelt et al. 1999). Estimate of the time to the most common ancestor (TMRCA) was obtained using a phylogeny-based approach implemented in NETWORK 4.5 using a mutation rate based on the number of fixed differences between chimpanzee and humans. An additional TMRCA estimate derived from application of a ML coalescent method implemented in GENETREE (Griffiths and Tavare 1994, 1995). The method assumes an infinite-site model without recombination; therefore, haplotypes and sites that violate these assumptions need to be removed: In the case of ZC3HAV1, we eliminated three variants. Again, the mutation rate μ was obtained on the basis of the divergence between human and chimpanzee and under the assumption both that the species separation occurred 6 Ma (Glazko and Nei 2003) and of a generation time of 25 years. The migration matrix was derived from previous estimated migration rates (Schaffner et al. 2005). Using μ and θ maximum likelihood (θ_{ML}), we estimated the effective population size parameter (N_e); this result equaled 22,400. With these assumptions, the coalescence time, scaled in $2N_e$ units, was converted into years. For the coalescence process, 10^6 simulations were performed. A third TMRCA estimate was obtained by applying a previously described method (Evans et al. 2005) that calculates the average pairwise difference between all chromosomes and the MRCA: This value was converted into years on the basis of mutation rate retrieved as above. The SD for this estimate was calculated as previously described (Thomson et al. 2000).

Prediction of SNP Effect

Genomic evolutionary rate profiling (GERP) and Grantham scores for rs3735007 were obtained from the NIEHS Exome Variant Server (http://snp.gs.washington.edu/niehsExome/). Prediction of SNP effects using PolyPhen (Sunyaev et al. 2001), PolyPhen2 (Adzhubei et al. 2010), SIFT (Ng and Henikoff 2003), Panther (Thomas and Kejariwal 2004), SNPs&GO (Calabrese et al. 2009), SNAP (Bromberg and Rost 2007), and MutPred (Li et al. 2009) were obtained from the dedicated websites: http://blocks.fhcrc.org/sift/SIFT.html, http://genetics.bwh.harvard.edu/pph/, http://genetics.bwh.harvard.edu/pph2/, http://www.pantherdb.org/tools/csnpScoreForm.jsp, http://snps-and-go.biocomp.unibo.it/snps-and-go, http://www.rostlab.org/services/SNAP, http://mutpred.mutdb.org. Prediction of protein ubiquitination probabilities was calculated using the UbPRED (http://

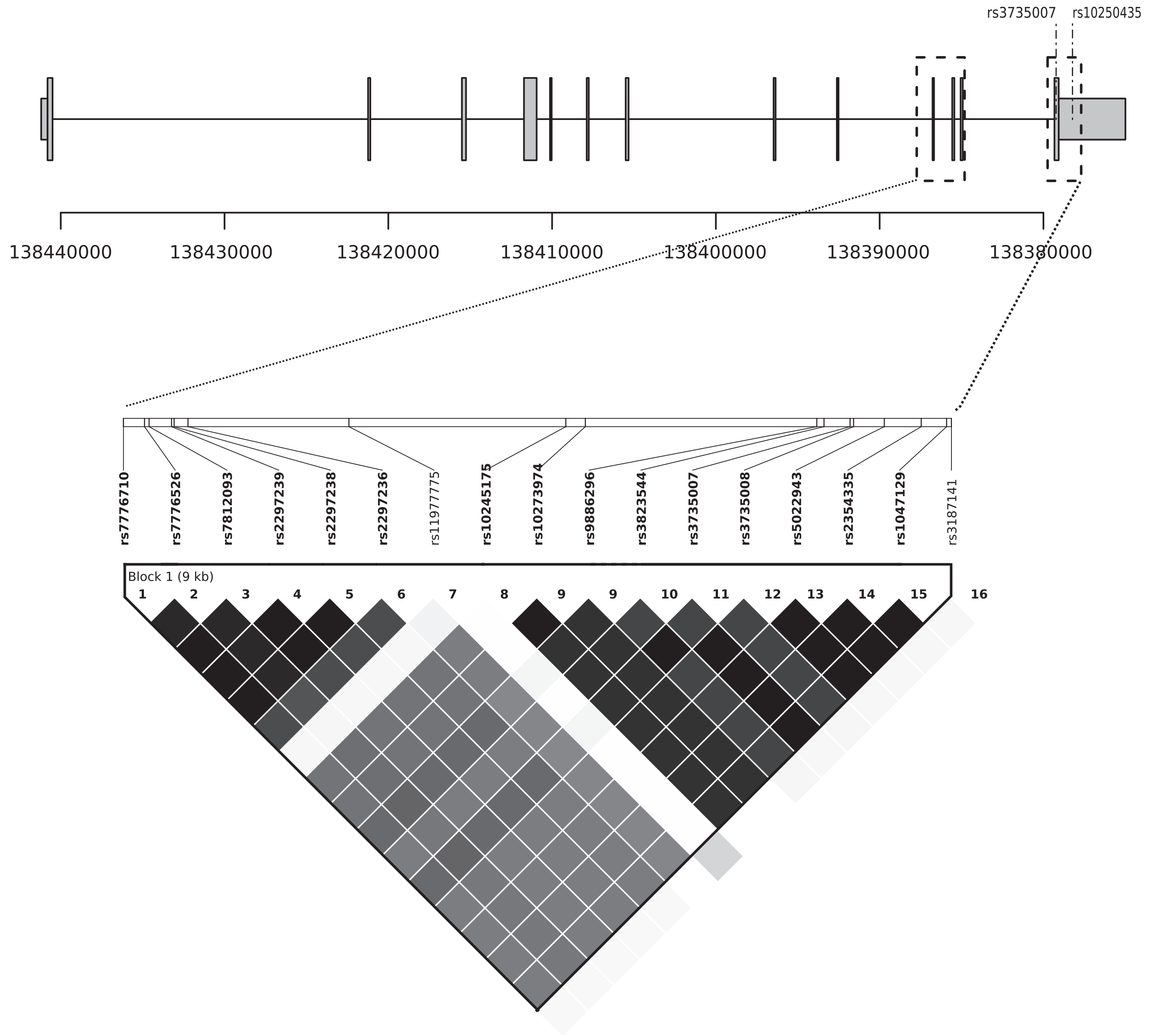

Fig. 1. Schematic diagram of the exon–intron structure of *ZC3HAV1*. Exons are indicated with the gray boxes. The two regions we resequenced are denoted by the hatched lines. The LD (r^2) plot refers to CEU, and data were derived from HapMap. The two *trans*-specific SNPs are shown.

www.ubpred.org) and CKSAAP_UbSite (http://protein.cau.edu.cn/cksaap_ubsite/index.php) servers.

Results

Nucleotide Diversity and Neutrality Tests in Human Populations

The PARP-like domain of *ZC3HAV1* is encoded by exons 10–13 (Kerns et al. 2008). The region covering these exons is in relatively tight LD in most human populations (fig. 1 and supplementary fig. S2, Supplementary Material online); indeed, LD extends beyond exon 13 and also covers part of the nearby *ZC3HAV1L* gene. (supplementary fig. S2, Supplementary Material online). In order to analyze the evolutionary history of the PARP-like domain, we resequenced two distinct genomic regions covering exons 10–12 (region 1, 2.40 kb) and 13 (region 2, 2.05 kb) (fig. 1) in five human populations (Yoruba, YRI; African Americans, AA; Europeans, CEU; East Asians, AS; and South Americans, SAI). The number of segregating variants identified in each human population for the two regions is reported in table 1. The statistical significance of LD between SNP pairs was evaluated using these data and indicated stronger LD in the region encompassing exon 13 (supplementary fig. S3, Supplementary Material online). An additional analysis of recombination in the two regions was performed using a composite likelihood method implemented in LDhat (McVean et al. 2002). In both regions and for the five populations analyzed separately, values of ρ equal or very close to 0 were obtained, and the likelihood permutation test revealed no evidence of recombination. No SNP pair in either region displayed significantly higher or lower LD than expected given the SNP frequencies and the estimate of $\rho \sim 0$.

We calculated nucleotide diversity by means of θ_W (Watterson 1975) and π (Nei and Li 1979) for the two regions we resequenced in *ZC3HAV1* and for a large number

Table 1. Nucleotide Diversity and Neutrality Tests for the *ZC3HAV1* Gene Regions.

				θ_W ($\times 10^{-4}$)		Π ($\times 10^{-4}$)		Tajima's *D*		Fu and Li's *D**		Fu and Li's *F**		Fu and Li's *D*	Fu and Li's *F*
Gene Region	Population	N^a	S^b	Value	Rank[c]	Value	Rank[c]	Value $(P)^d$	Rank[c]	Value $(P)^d$	Rank[c]	Value $(P)^d$	Rank[c]	Value $(P)^d$	Value $(P)^d$
1 (exons 10–12)	YRI	40	11	10.77	0.77	14.93	0.39	1.17 (0.025)	0.96	0.35 (0.2)	0.70	0.73 (0.081)	0.84	0.22 (0.25)	0.60 (0.11)
	AA	48	13	12.20	0.81	16.18	0.47	0.99 (0.035)	0.94	−0.51 (0.53)	0.40	−0.012 (0.28)	0.62	−0.17 (0.38)	−0.20 (0.21)
	CEU	40	12	11.75	0.93	15.20	0.59	0.91 (0.18)	0.80	0.97 (0.14)	0.83	1.11 (0.12)	0.87	0.93 (0.15)	1.00 (0.16)
	AS	40	11	10.77	0.93	10.75	0.44	−0.0047 (0.47)	0.51	−1.28 (0.14)	0.19	−1.02 (0.19)	0.27	−0.41 (0.35)	−0.40 (0.35)
	SAI	30	8	8.41	n.a.	12.62	n.a.	1.52 (0.086)	n.a.	0.056 (0.51)	n.a.	0.59 (0.32)	n.a.	1.33 (0.066)	1.570 (0.059)
2 (exon 13)	YRI	40	19	21.77	0.98	40.20	0.95	2.78 (<0.001)	>0.99	1.65 (<0.001)	>0.99	2.39 (<0.001)	>0.99	1.79 (<0.001)	2.58 (<0.001)
	AA	48	22	24.16	0.99	41.35	0.95	2.31 (0.0011)	>0.99	0.70 (0.12)	0.86	1.49 (0.01)	0.99	1.11 (0.057)	1.87 (0.003)
	CEU	40	18	20.62	0.98	39.33	0.95	2.96 (<0.0012)	>0.99	1.25 (0.057)	0.95	2.15 (0.002)	>0.99	1.35 (0.04)	2.30 (0.002)
	AS	40	17	19.48	0.99	38.63	0.94	3.19 (<0.0014)	0.99	1.21 (0.06)	0.95	2.21 (0.0025)	0.99	1.30 (0.04)	2.35 (0.002)
	SAI	30	16	19.68	n.a.	10.04	n.a.	−1.66 (0.29)	n.a.	1.56 (<0.001)	n.a.	0.64 (0.32)	n.a.	1.71 (<0.001)	0.70 (0.30)

NOTE.—n.a., not available.

[a] Sample size (chromosomes).

[b] Number of segregating sites.

[c] Percentile rank relative to a distribution of 2 kb windows from NIEHS genes.

[d] *P* value calculated by coalescent simulations.

Table 2. MLHKA Test for the Two Analyzed *ZC3HAV1* Regions.

		MLHKA							
		YRI		CEU		AS		AA	
Region	Fixed Sub[a]	k^b	*P*	k^b	*P*	k^b	*P*	k^b	*P*
1	21	2.06	0.28	2.74	0.053	2.38	0.069	2.17	0.17
2	15	3.15	0.012	3.94	0.0019	4.07	0.0012	5.29	0.0004

[a] Number of fixed substitutions (human/chimpanzee).

[b] Selection parameter ($k > 1$ indicates an excess of polymorphism relative to divergence).

of windows 2 kb in size (see Materials and Methods) deriving from 238 human genes resequenced by the NIEHS SNP Discovery Program (http://egp.gs.washington.edu/) in CEU, AA, YRI, and AS (data for SAI are not available). The empirical distributions obtained from these windows were used to calculate the percentile rank of θ_W and π values for the *ZC3HAV1* regions. As shown in table 1, no exceptional diversity was observed for exons 10–12; conversely, extremely high values for both θ_W and π were observed for the region covering exon 13 in all populations. Randomly selected 2-kb windows may contain few segregating sites, possibly introducing a large variance in the distribution of diversity values. Thus, we also performed the same calculations using 5-kb windows sampled from the same genes; very similar results were obtained in these analyses (supplementary table S3, Supplementary Material online).

Polymorphism levels also depend on local mutation rates; therefore, under neutral evolution, the amount of within- and between-species diversity is expected to be similar at all loci in the genome. The multilocus HKA test was developed to verify this expectation (Wright and Charlesworth 2004). Specifically, we applied a multilocus MLHKA (maximum likelihood HKA) test by comparing polymorphism and divergence levels at the *ZC3HAV1* regions with 16 NIEHS genes resequenced in YRI, AA, CEU, and AS. As specified in the Materials and Methods, these genes were randomly selected among NIEHS loci shorter than 20 kb that showed no evidence of natural selection (Fumagalli, Cagliani, et al. 2009). No significant excess of intra- versus interspecies diversity was observed for region 1 (table 2). Conversely, *ZC3HAV1* region 2 displays a significant excess of polymorphism compared with divergence in all populations (table 2). It is worth noting that *ZC3HAV1* is located on chromosome 7 in tandem orientation with a homolog, *ZC3HAVL1*. Yet, this latter displays no PARP-like domain and, therefore, no alignment with *ZC3HAV1* across exons 10–13. Thus, conversion events between the two genes are not expected to occur in the region we analyzed.

Unusually, high levels of genetic diversity might indicate the action of balancing selection, as neutral variation tends to be maintained with the selected alleles. Another hallmark of balancing selection is a distortion of the SFS toward intermediate frequency alleles. Thus, to gain further insight into the evolutionary history of the two *ZC3HAV1* regions, we applied SFS-based tests, such as Tajima's *D* (D_T) (Tajima 1989) and Fu and Li's *D** and *F** (Fu and Li 1993). We evaluated the statistical significance D_T, *D** and *F** by means both of coalescent simulations that incorporate demographic scenarios (see Material and Methods) and of empirical comparison with the distribution of these statistics calculated over 2-kb windows. Most statistics failed to reject the null hypothesis of selective neutrality for the *ZC3HAV1* region 1 (table 1). Conversely, SFS-based statistics calculated for region 2 yielded significantly high values both in coalescent simulations and empirical comparisons for YRI, CEU, and AS. As for SAI, coalescent simulations yielded a significantly low value for D_T and a significantly high result for D^*. Very similar results were obtained with SFS-based statistics that incorporate out-group information (from orangutan in this case), namely Fu and Li's F and D, whose statistical significance was calculated through coalescent simulations (table 1). Comparable percentile ranks were also obtained using the distribution of 5-kb reference windows (supplementary table S3, Supplementary Material online) and by applying different demographic scenarios in coalescent simulations (see Materials and Methods) (supplementary table S4, Supplementary Material online).

Overall, these data suggest that balancing selection has been acting on *ZC3HAV1* exon 13.

Identification of *Trans*-Specific Polymorphisms and Population Genetic Analysis in Chimpanzees

In order to analyze the evolutionary history of the *ZC3HAV1* exon 13 region in chimpanzee and to assess the presence of *trans*-specific polymorphisms, we resequenced nine *P. troglodytes* individuals.

A total of nine segregating sites were identified, two of them being shared with humans. Although one of these SNPs (rs10250435) occurs at a CpG dinucleotide, suggesting that it may be accounted for by independent mutations arisen after speciation, rs3735007 (C/T) does not involve a CpG and represents a nonsynonymous substitution (Thr851Ile). The probability of observing n shared SNPs in a genomic region where segregating sites are observed in two species is given by the hypergeomtric density (Clark 1997). Nineteen and 6 non-CpG SNPs segregate in humans and chimpanzees, respectively, in the *ZC3HAV1* resequenced region. Thus, the probability of observing one polymorphism at the same position over a length of 1,969 non-CpG nucleotides amounts to 0.055; given that roughly 2/3 of substitutions are accounted for by transitions in humans (Nachman and Crowell 2000), the probability of observing the same alleles in the two species amounts to ~0.036 for transitions (which is the case of rs3735007). The same calculation can be performed for SNPs that occur at CpG dinucleotides (five and three in humans and chimpanzees, respectively): In this case, by considering 66 CpG dinucleotides in the human/chimpanzee ancestor, the probability of observing a shared polymorphism involving a CpG amounts to 0.199; among substitutions at CpGs, roughly 75% are transitions (Nachman and Crowell 2000), resulting in a probability of 0.15 to observe the same alleles in the two species at the same CpG site. Thus, the combined probability that the two *trans*-specific polymorphisms have occurred independently in the two lineages is low ($P = 0.0054$).

The presence of human–chimpanzee *trans*-specific polymorphisms suggests that balancing selection signatures should be detectable in *P. troglodytes* populations, as well. To test this possibility, we performed population genetic analyses in this species. Since we had no information concerning the geographic origin and subspecies type of these individuals, we addressed this issue first. Analysis of mtDNA indicated that all of them had *P.t. verus* maternal origin; similarly, the three male chimpanzees had Y chromosomes suggesting *P.t. verus* paternal ancestry. STRUCTURE analysis of these individuals using 26 autosomal loci (see Materials and Methods) indicated that a model with only one cluster ($k = 1$) had much higher probability compared to models with $K = 2$, 3, or 4 (supplementary fig. S1, Supplementary Material online), suggesting that all individuals belong to the same subspecies (i.e., *P.t. verus*). In order to assess whether *ZC3HAV1* shows unusual levels of nucleotide variability, we resequenced 16 genomic regions in these same individuals to be used as an empirical comparison. All regions were similar in size to *ZC3HAV1* region 2 and are described in the supplementary table S5 (Supplementary Material online). Ten of these regions were intergenic and previously proposed to be neutrally evolving (Frisse et al. 2001); six more intergenic regions were randomly selected with the only requirement that no gene was annotated within 200 kb. Calculation of θ_W and π indicated that region 2 in *ZC3HAV1* shows the highest values of diversity compared with the 16 reference regions (fig. 2). This result was confirmed by application of the MLHKA test, which yielded significant results ($k = 2.95$, $P = 0.049$). Conversely, SFS-based statistics (D_T, F^* and D^*) were not exceptionally high in this region (supplementary table S5, Supplementary Material online). In order to analyze the haplotype genealogy for the *ZC3HAV1* region 2 in chimpanzees, we constructed a median-joining network; this indicated the presence of two major clades, with one basal branch carrying the *trans*-specific polymorphism (fig. 2). Haplotype frequency in our *P. troglodytes* sample was very different between the two clades, resulting in no excess of intermediate frequency alleles. This explains the failure of SFS-based tests to reject neutrality.

Haplotype Genealogy and Estimation of the TMRCA

In analogy to the analysis we performed in chimpanzees, we analyzed the genealogy of human haplotypes identified in the five populations. The topology of the median-joining network displays two major clades (denoted as 1 and 2 in fig. 3) containing common haplotypes and separated by long-branch lengths. The Thr851Ile polymorphism (variant 4 in the network) defines the major haplotype in clade 1 (fig. 3). All populations tend to display a similar frequency of haplotypes in the two clades with the exception of SAI chromosomes that mainly cluster in clade 1. This observation explains the low value of D_T we obtained for this population (table 1).

In line with the relatively tight LD in the region, the network displayed no reticulation, and only a couple of recurrent

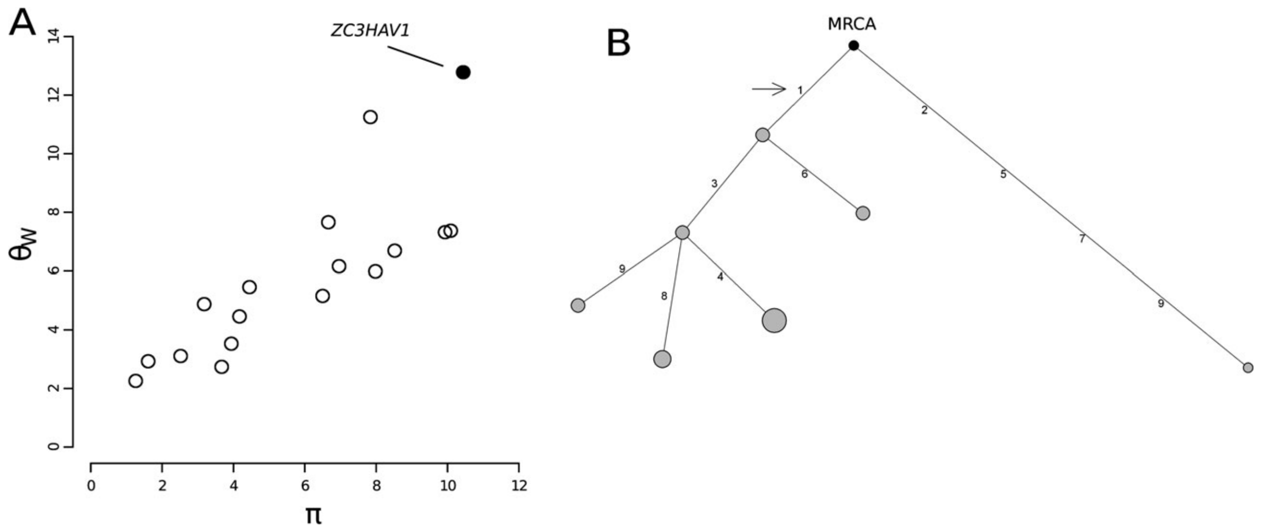

Fig. 2. Nucleotide diversity and haplotype analysis in chimpanzee. (*A*) Plot of θ_W and π values for the 16 regions we analyzed (white circles) and for *ZC3HAV1* region 2 (black). (*B*) Network analysis of region 2 in chimpanzee: the MRCA is indicated by the black circle; the position of the *trans*-specific variant is shown (arrow).

mutations occur in the genealogy. Under these conditions, estimation of the TMRCA is robust. To this aim, we applied three approaches; the first is a phylogeny-based method that calculates the average pairwise difference between the haplotype clusters and a root (Bandelt et al. 1999). Using a mutation rate based on the number of fixed differences with chimpanzee and a separation time of 6 myr (Glazko and Nei 2003), we estimated TMRCAs of 5.96 myr (SD: 1.42 myr). The second approach is implemented in GENETREE and is based on an ML coalescent analysis (Griffiths and Tavare 1994, 1995). Using this method, the TMRCA of the *ZC3HAV1* haplotype lineages amounted to 4.36 myr (SD: 1.05 myr) (fig. 3). A third TMRCA of 6.12 myr (SD: 2.2 myr) was obtained by the application of a previously proposed method based on the mean nucleotide diversity between all chromosomes and an MRCA (Evans et al. 2005).

These TMRCA estimates are much deeper than expected under neutrality (Tishkoff and Verrelli 2003; Garrigan and Hammer 2006).

Sliding-Window Analysis with 1000 Genomes Pilot Project Data

Due to the combined action of mutation and recombination, signatures of long-standing balancing selection are expected to extend over relatively short genomic regions (Charlesworth 2006). We took advantage of the availability of data from the 1000 Genomes Pilot project (1000 Genomes Project Consortium et al. 2010) to verify this expectation by analyzing sequence variation along the entire *ZC3HAV1* transcription unit. The low-coverage 1000 Genomes approach, which generated whole-genome sequencing data of 179 individuals with different ancestry (YRI, CEU, and AS), is estimated to have relatively low power to detect singleton SNPs or rare variants (1000 Genomes Project Consortium et al. 2010). Thus, sliding-window analysis and comparison with Sanger resequencing data also serve the purpose of testing the suitability of low-coverage data for large-scale identification of balancing selection targets. Therefore, we calculated π and the ratio of θ_W over human–chimpanzee divergence in sliding windows moving along *ZC3HAV1*. It should be noted that sliding-window analyses have an inherent multiple testing problem that is difficult to correct because of the nonindependence of windows. In order to partially account for this limitation, we applied the same procedure to 2,000 randomly selected human genes, and the distribution of π and θ_W per divergence was obtained for the corresponding windows. This allowed calculation of the 95th percentile and visualization of regions that exceed this threshold; nonetheless, these analyses should be regarded as mainly exploratory and descriptive.

Using the 1000 Genomes data for the three populations, peaks of both π and θ_W per divergence were observed for *ZC3HAV1* exons 10–12 and 13, the two regions we resequenced (supplementary fig. S4, Supplementary Material online). In both cases, nucleotide diversity exceeded the 95th percentile. Yet, contrary to our results based on Sanger resequencing, higher values of both diversity estimates were obtained for region 1. Further analyses indicated that, as expected, the 1000 Genomes Project detected more variants in region 1 than we did in our smaller sample (supplementary table S6, Supplementary Material online). The opposite situation occurred in region 2, where we detected a total of 24 variants compared with 13 in the low-coverage data. Inspection of these SNPs revealed that most of them ($n = 8$) have intermediate frequency in all populations we analyzed and are located within an AluSx element in the 3′ untranslated region of *ZC3HAV1* (supplementary table S6, Supplementary Material online). Indeed, a large portion of "inaccessible sites" in the low-coverage 1000 Genomes data maps to repetitive sequences (1000 Genomes Project Consortium et al. 2010); sliding-window analysis of *ZC3HAV1* after masking of transposable elements revealed a similar or higher nucleotide diversity in region 2 compared with region 1 (supplementary fig. S5, Supplementary Material online).

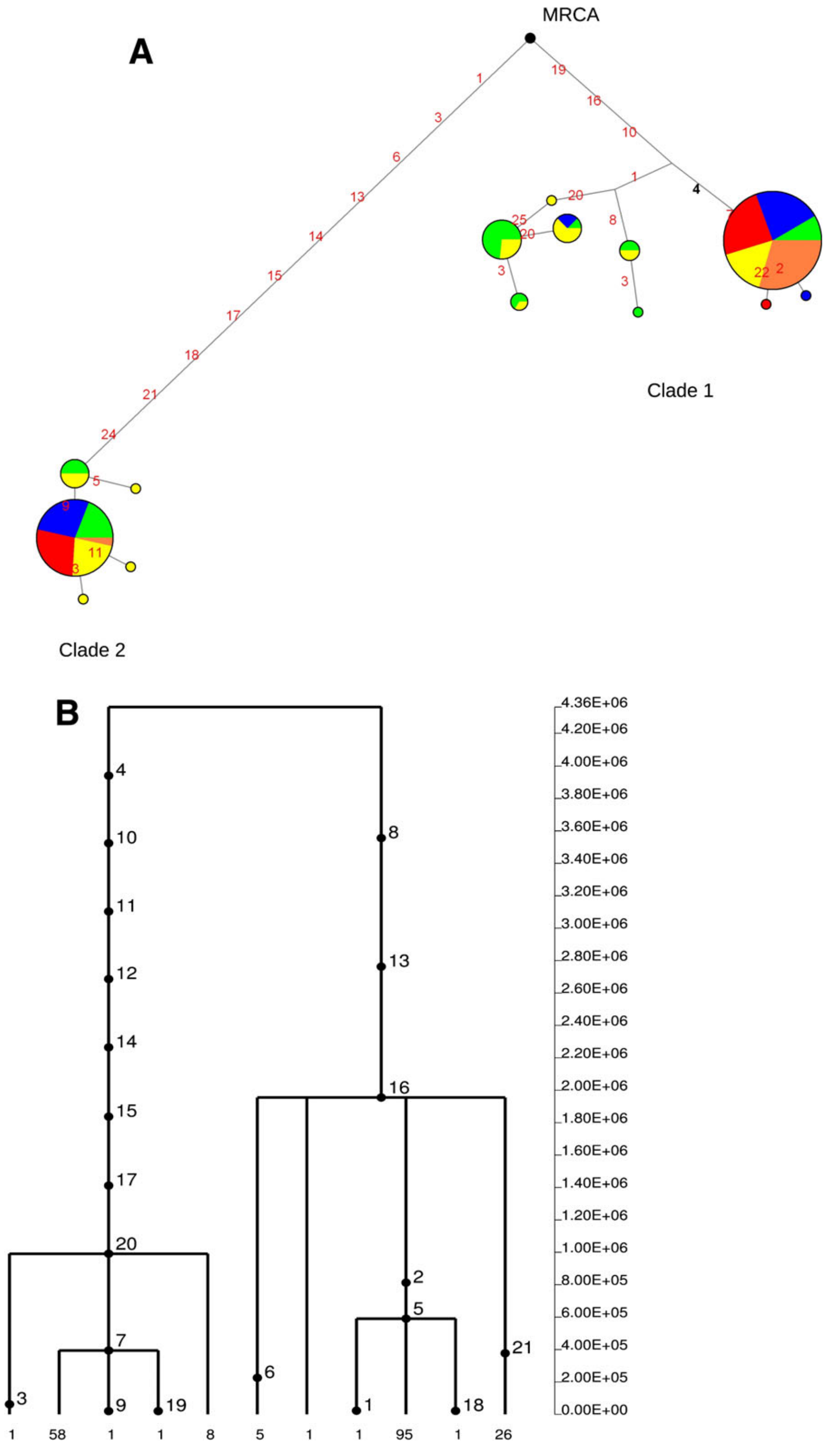

Fig. 3. Median-joining network and GENETREE analysis of *ZC3HAV1* haplotypes. (*A*) Median-joining network: each node represents a different haplotype, with the size of the circle proportional to frequency. Nucleotide differences between haplotypes are indicated on the branches of the network. Circles are color coded according to population (green: YRI, yellow: AA, blue: CEU, red: AS, and orange: SAI). The MRCA is also shown (black circle). The relative position of mutations along a branch is arbitrary. (*B*) GENETREE: mutations are represented as black dots and named for their physical position along the regions. The absolute frequency of each haplotype is also reported. Note that mutation numbering does not correspond to that reported in (*A*) (see Materials and Methods).

In order to verify whether this report bias represents a general feature of low-coverage data, we calculated θ_W using both the NIEHS and the 1000 Genomes data for a large number of windows deriving from either unique or repetitive sequences (see Materials and Methods). In YRI, CEU, and AS, Kendall's correlation coefficient (τ) between

Table 3. Association Analysis of *ZC3HAV1* Variants in Multiple Sclerosis.

				Genotype Counts				
SNP	A1[a]	A2[b]	Sample Origin	MS	HC	P[c]	P (recessive)[d]	OR (95 CI)[e]
rs3735007	C	T	Mainland Italy	132/228/147	101/275/147	0.016	0.011	1.47 (1.08–1.99)
rs1047129	T	C	Sardinia	746/1098/426	629/1100/417	0.037	0.011	1.18 (1.03–1.34)
rs3735007	C	T	Belgium	182/401/236	96/205/128	0.91	0.73	1.05 (0.81–1.37)

[a] Allele 1.

[b] Allele 2.

[c] Fisher's exact test *P* value for a genotypic model.

[d] Fisher's exact test *P* value for a recessive model.

[e] OR for a recessive model with 95% CI.

θ_W estimates obtained using the two data sets was higher for unique compared with repetitive sequences (for unique and repetitive sequences, CEU: $\tau = 0.46$ and 0.38; YRI: $\tau = 0.55$ and 0.48; and AS: $\tau = 0.69$ and 0.60, all P values < 0.001). Quantile–quantile plots of θ_W values calculated from the 1000 Genomes low-coverage data and from NIEHS Sanger resequencing experiments indicated that higher estimates of nucleotide diversity are obtained from the latter, and the discrepancy is stronger for transposable elements compared with unique regions (supplementary fig. S6, Supplementary Material online).

We next used 1000 Genomes data to calculate D_T, F^* and D^* for the two *ZC3HAV1* region we resequenced. These values were compared with those deriving from 2,000 windows (2.5 kb in size) randomly drawn from the same 2,000 genes used above. Given the results obtained above, SFS-based statistics were calculated for both repeat-masked and nonmasked windows. As expected, all statistics tended to display higher values compared with those calculated from our resequencing experiments. Using both masked and unmasked regions, D_T and F^* calculated for the *ZC3HAV1* region 2 showed values equal to or above than the 99th percentile, confirming the rejection of neutrality (supplementary table S7, Supplementary Material online). No significantly high value was observed for region 1 (supplementary table S7, Supplementary Material online).

Association Analysis for Multiple Sclerosis

In order to test the possibility that, in analogy to other antiviral response genes (O'Brien et al. 2010; Cagliani et al. 2011), variants in ZC3HAV1 affect the susceptibility to MS, we analyzed SNPs located in region 2 in two Italian case/control cohorts. In particular, the Thr851Ile polymorphism (rs3735007) was genotyped in 507 MS and 523 healthy controls (HC) from mainland Italy. A significant deviation from HWE with an excess of homozygotes was observed in cases but not in controls ($P = 0.026$ for MS subjects). Wittke-Thompson et al. (2005) have developed a test to verify whether deviations from HWE can be explained by an underlying genetic model for the trait being analyzed rather than by other effects. Using a K_p (prevalence of MS in the general population) of 0.0007 for Mainland Italy (Pugliatti et al. 2006), the best fitting of the genotypic proportions in cases and controls was obtained with a recessive model. The goodness-of-fit test was not significant ($\chi^2 = 2.27$, $P = 0.32$, df = 2) for this model, and the estimated γ (homozygote relative risk) was 1.38. These data indicate that a recessive model with only genetic effects adequately explains HWE deviation in the Italian sample. Comparison of genotype frequencies in MS and HC indicated a significant difference (Fisher's exact P value = 0.016, 2 df), and for a recessive model with CC homozygotes predisposed to disease, the odds ratio [OR] was 1.47 (95% confidence intervals [CI]: 1.08–1.99, $P = 0.011$) (table 3). In order to replicate this finding, we obtained the genotype distribution of rs1047129, which is in tight LD with rs3735007 ($r^2 = 0.87$), from a case–control genome-wide association study for MS performed in a population of 4,416 individuals from Sardinia (2,270 patients and 2,146 controls). rs1047129 is located on the major branch leading to clade 2 (variant 24 in fig. 3), and its T allele is in phase with the rs3735007 C allele. No deviation from HWE was observed for rs1047129. As shown in table 3, the genotype frequency of this SNP significantly differed in cases and controls (for a recessive model with TT homozygotes predisposed to disease, OR = 1.18, 95% CI: 1.037–1.344, $P = 0.011$).

Finally, we analyzed rs3735007 in MS and HC individuals from Belgium. In this population, we observed no deviation from HWE and no difference in the genotype distribution of cases and controls (table 3).

Discussion

Genes coding for protein products involved in both adaptive and innate immune responses have likely been subject to an extraordinary selective pressure during the evolutionary history of mammals. Although some of these genes have been targets of relatively recent selective events in human populations (reviewed in Barreiro and Quintana-Murci (2010)), other loci may have evolved in response to long-term pressures. In this latter situation, polymorphisms may be maintained in populations for a period of time, which is statistically inconsistent with neutrality and even survive speciation events (Charlesworth 2006).

A large body of evidence suggests that long-standing balancing selection has shaped diversity at the human HLA class I and class II genes and maintained *trans*-specific polymorphisms at these loci (Lawlor et al. 1988; Mayer et al. 1988; Gongora et al. 1996; Charlesworth 2006). For HLA genes, natural selection is thought to promote diversity at the peptide-binding cleft of antigen presenting molecules and to be pathogen driven (Prugnolle et al. 2005). Outside the major histocompatibility complex (MHC), no human/chimpanzee *trans*-specific amino acid variant has been reported, and the scant descriptions of long-standing balancing selection in humans have involved innate immunity genes (Cagliani et al. 2008, 2010). This testifies the extraordinary relevance of adaptive immunity for organism survival but suggests that searching for long-standing signatures of natural selection might contribute to the identification of non-HLA loci and variants with a prominent role in immune response. Our data suggest that the Thr851Ile variant in the PARP-like domain of ZC3HAV1 has been maintained as a *trans*-specific variant in human and chimpanzee populations. Several evidences support this conclusion and argue against the possibility that the polymorphism results from concurrent mutation in the two species. First, as calculated above, the probability that the variant represents an independent mutation event in the human and chimpanzee lineages is low and much more so when the presence of a second shared SNP (albeit involving a CpG dinucleotide) is accounted for. Therefore, at least one of the two shared polymorphisms is likely to have arisen before speciation. The lower mutability of rs3735007 compared with the CpG SNP and its potential functional significance (it accounts for the Thr851Ile variant) make it a better candidate as a selection target. Second, all tests supported the action of balancing selection on ZC3HAV1 region 2 in human populations. Third, using different methods, we estimated TMRCAs of the haplotype genealogy ranging from 4.36 to 6.12 myr. These are much deeper than expected under neutrality, although they display large variances and differ sensibly from one another. This is expected as relatively few segregating sites are used for time estimates, despite high local diversity. Several studies have provided different estimates for the human/chimpanzee lineage split (with ranges from 4 to 7 myr; Glazko and Nei 2003; Kumar et al. 2005), again with wide confidence intervals. Therefore, although caution should be used when speculating on date estimates, the TMRCAs we obtained for *ZC3HAV1* exon 13 might be consistent with the Thr851Ile variant having arisen before human/chimpanzee lineage split. Finally, analysis of the chimpanzee sample also suggested the action of natural selection, as demonstrated by the higher nucleotide diversity observed in *ZC3HAV1*. The average estimates obtained for θ_W (0.054%) and π (0.056%) for putatively neutrally evolving regions are comparable to although slightly lower than previously reported data for *P.t. verus* populations both in intergenic and in genic regions (Gilad et al. 2003; Yu et al. 2003; Verrelli et al. 2006, 2008; Claw et al. 2010). This discrepancy may be due to the relatively limited number of individuals we resequenced. Still, the values we calculated for *ZC3HAV1* are higher than those obtained for all other regions, suggesting the action of natural selection in the maintenance of diversity at this locus. Moreover, haplotype analysis indicated that, in analogy to the observation in humans, the *trans*-specific SNP separates the two major haplotype clades.

Despite its limitations, sliding-window analysis of ZC3HAV1 using the 1000 Genomes low-coverage whole-genome data also suggested that the balancing selection signature is located in the region surrounding exon 13 (supplementary figs. S4 and S5, Supplementary Material online). To our knowledge, low-coverage data have never been used for the detection of balancing selection signatures. Our comparison with Sanger resequencing experiments suggests that unequal coverage of repetitive elements (1000 Genomes Project Consortium et al. 2010) may bias diversity estimates across the genome. Indeed, using a large number of gene windows resequenced by both the 1000 Genomes Project and by the NIEHS Program, we show that the presence of transposable elements biases nucleotide diversity estimates obtained through next generation sequencing approaches. Thus, large-scale efforts aimed at describing natural selection through the use of low-coverage data should be better performed after repeat masking.

It is worth noting that, despite having an extremely deep coalescent time, the *ZC3HAV1* region surrounding exon 13 displays high LD (supplementary fig. S3, Supplementary Material online) and virtually no evidence of recombination (as also evident from the haplotype genealogy; fig. 3). This is partially explained by the region being short; as mentioned above, signatures of long-standing balancing selection are expected to have a limited genomic span (in the order of a few hundred bases), given that they are eroded over time by both mutation and recombination, as determined by both simulations and real-data analyses (Wiuf et al. 2004; Bubb et al. 2006). Thus, the power to detect long-term balancing selection is higher in regions with low recombination rates (Wiuf et al. 2004), as the one we describe herein. In particular, Bubb et al. (2006) suggested that, in order to be detectable, long-standing balancing selection must involve at least two physically linked loci that are both selection targets (i.e., a balanced haplotype). This would allow accumulation of neutral variability over longer regions due to selection against most recombination events in the interval. Whether another variant in the region, in addition to Thr851Ile, represents a selection target remains to be evaluated, but this possibility might reconcile the strong LD we detected with the deep coalescence time of the haplotype genealogy.

The possible functional significance of the Thr851Ile variant is difficult to infer, as the role of the PARP-like domain itself is presently unknown. In order to evaluate the possible functional effect of the amino acid replacement, we applied several prediction algorithms. The GERP (Cooper et al. 2005) and Grantham scores (Grantham 1974) (−6.1 and 89, respectively) indicated that the position is not highly conserved and that the substitution is moderately conservative. Application of the PolyPhen (Sunyaev et al. 2001), PolyPhen2 (Adzhubei et al. 2010), SIFT (Ng and Henikoff 2003), Panther (Thomas and Kejariwal 2004), and SNPs&GO (Calabrese et al. 2009) algorithms suggested that the variant has relatively limited functional impact (not shown). Conversely, SNAP (Bromberg and Rost 2007) indicated the

SNP to be nonneutral, although with very low reliability, and the use of MutPred (Li et al. 2009) yielded a relatively high general score (0.586) and a borderline probability ($P = 0.054$) that the variant might affect ubiquitination at a nearby Lysine residue (K849). Although potentially interesting given the known exploitation of the host's ubiquitin system for immune evasion, this result should be taken with caution as prediction of ubiquitination sites (using UbPRED, Radivojac et al. 2010 and CKSAAP_UbSite, Chen et al. 2011, not shown) yielded low scores for K849. Overall, a conservative interpretation of the results for prediction algorithms is that the SNP has no major effect on the function of ZC3HAV1. We consider that this finding is not in contrast with the hypothesis that rs3735007 represents the selection target; indeed, selection might have operated not by hampering or reducing protein function but rather by modifying its affinity or binding specificity for viral components, a possibility that is difficult to test by the use of available prediction algorithms. In fact, a previous analysis of the entire *ZC3HAV1* coding region indicated that the PARP-like domain has been a target of selection during primate evolution, as well, suggesting that it serves some important function for survival or reproduction (Kerns et al. 2008). In line with these considerations, it has been demonstrated that the long *ZC3HAV1* isoform containing the PARP-like domain has stronger activity against alphaviruses compared with the shorter form (Kerns et al. 2008). Therefore, a straightforward interpretation of our data envisages a situation where the *trans*-specific polymorphism affects either the specificity of viral recognition by *ZC3HAV1* or the binding of a viral inhibitor to *ZC3HAV1*, two alternatives that were previously proposed to account for the positive selection signature of this region in primates (Kerns et al. 2008). One interesting possibility is a genetic conflict between primate hosts and viral pathogens involving the *ZC3HAV1* PARP-like domain on one side and the viral macrodomains on the other. Macrodomains are conserved in several organisms and are generally referred to as X domains in the context of viral-encoded proteins (Gorbalenya et al. 1991). Apart from alphaviruses, few other viruses, including coronaviruses and rubella, possess X domains (Egloff et al. 2006; Neuvonen and Ahola 2009). Cellular macrodomains have the ability of inhibiting PARPs through a direct interaction (Ouararhni et al. 2006), and a recent report identified three macrodomains encoded by the highly pathogenetic SARS-CoV (Tan et al. 2009). The authors suggested that viral macrodomains might have evolved to counteract the antiviral activity of ZC3HAV1 (Tan et al. 2009). Whether *ZC3HAV1* has any activity against coronaviruses and whether this occurs through the PARP-like domain remains to be elucidated. Nonetheless, these observations suggest that ZAP might display an antiviral activity that extends beyond alphaviruses and filoviruses and may include more common human pathogens, such as rubella and coronaviruses.

Although a specific causative agent has never been identified, several evidences suggest that viral infections may cause or at least contribute to the development of MS. MS is an autoimmune disease which results from the exposure of genetically predisposed individuals to environmental factors which, in turn, trigger a breakdown in T-cell tolerance to myelin antigens. In line with the possible involvement of viral agents in the pathogenesis of MS, a recent report indicated that a polymorphism which decreases the enzymatic activity of *OAS1*, an antiviral response gene, is associated with MS predisposition and disease severity (O'Brien et al. 2010; Cagliani et al. 2011). This concept is supported by the observation that a subset of human genes involved in the pathogenesis of MS have been targets of virus-driven selection (Fumagalli et al. 2010). Data herein indicate that, in two independent and genetically distinct populations from Italy, variants in *ZC3HAV1* confer a risk to develop MS. The reasons why this finding was not replicated in a Belgian sample remain to be elucidated and may derive either from insufficient statistical power of the study or from the interaction between *ZC3HAV1* polymorphisms with environmental factors. Such an interaction occurs between the hepatitis A virus (HAV) and polymorphisms in its receptor: HAV infection is protective against atopy only in subjects carrying a polymorphic 6-amino acid insertion in the *HAVCR1* mucin-like domain (McIntire et al. 2003). Indeed, gene-by-environment (GXE) interactions have been recently shown to represent a common attribute of complex diseases (Romanoski et al. 2010). In the case of MS, the relevance of GXE interaction remains to be evaluated, but several epidemiological studies (reviewed in Ebers (2008)) indicate that diverse environmental risk factors modulate disease prevalence. Also, it is worth mentioning that rs3735007 is included in a relatively large LD block that partially covers the nearby *ZC3HAV1L* gene (supplementary fig. S2, Supplementary Material online), putatively encoding an antiviral protein with still uncharacterized function. Thus, we cannot rule out the possibility that the association we detected with MS in the Italian cohorts is accounted for by a variant located in *ZC3HAV1L*. The probability is low that a nonsynonymous SNP in this latter gene is responsible, as the only one (rs17856272) displays very low frequency in CEU and limited LD ($r^2 = 0.038$) with rs3735007. Still, other variants with regulatory function located within the LD block might account for the association with MS and need to be examined in further studies.

In summary, our data describe one of the first examples *trans*-specific coding SNP maintained by balancing selection and located outside the MHC and warrant further analyses on the role of this variant in MS susceptibility.

Supplementary Material

Supplementary figures S1–S6 and tables S1–S7 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

M.C. is supported by grants from the Istituto Superiore di Sanita "Programma Nazionale di Ricerca sull' AIDS," the EMPRO and AVIP EC WP6 Projects, the nGIN EC WP7 Project, the Japan Health Science Foundation, 2008 Ricerca Finalizzata (Italian Ministry of Health), 2008 Ricerca Corrente (Italian Ministry of Health), Progetto FIRB RETI: Rete Italiana Chimica Farmaceutica CHEM-PROFARMA-NET (RBPR05NWWC) and Fondazione CARIPLO. D.G. is supported by grant Giovani Ricercatori 2007 (D.lgs 502/92) from Italian Ministry of Health. B.D. is a Clinical Investigator of the Fund for Scientific Research Flanders (FWO-Vlaanderen) and is supported by the Biogen Idec Chair Translational Research in Multiple Sclerosis and by the Bayer Chair on Fundamental Genetic Research regarding the Neuroimmunological Aspects of Multiple Sclerosis. B.D. and A.G. are supported by Onderzoeksfonds K.U.Leuven/Research Fund K.U.Leuven (OT/11/087), Wetenschappelijk Onderzoek Multiple Sclerose (WOMS)—Vlaanderen and the Belgian Neurological Society. A.G. is supported by the Belgian Charcot Foundation. We wish to thank Dr Christian Roos from the Gene Bank of Primates, Primate Genetics (Germany), for sending chimpanzee genomic DNA.

References

Abecasis GR, Cookson WO. 2000. GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* 16:182–183.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods.* 7:248–249.

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

Barreiro LB, Quintana-Murci L. 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet.* 11:17–30.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.

Bick MJ, Carroll JW, Gao G, Goff SP, Rice CM, MacDonald MR. 2003. Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *J Virol.* 77:11555–11562.

Bromberg Y, Rost B. 2007. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* 35:3823–3835.

Bubb KL, Bovee D, Buckley D, et al. (12 co-authors). 2006. Scan of human genome reveals no new loci under ancient balancing selection. *Genetics* 173:2165–2177.

Cagliani R, Fumagalli M, Biasin M, Piacentini L, Riva S, Pozzoli U, Bonaglia MC, Bresolin N, Clerici M, Sironi M. 2010. Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. *Hum Genet.* 128:577–588.

Cagliani R, Fumagalli M, Guerini FR, et al. (17 co-authors). 2011. Identification of a new susceptibility variant for multiple sclerosis in OAS1 by population genetics analysis. *Hum Genet.* 131:87–97.

Cagliani R, Fumagalli M, Riva S, Pozzoli U, Comi GP, Menozzi G, Bresolin N, Sironi M. 2008. The signature of long-standing balancing selection at the human defensin beta-1 promoter. *Genome Biol.* 9:R143.

Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. 2009. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat.* 30:1237–1244.

Cereda M, Sironi M, Cavalleri M, Pozzoli U. 2011. GeCo++: a C++ library for genomic features computation and annotation in the presence of variants. *Bioinformatics* 27:1313–1315.

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* 2:e64.

Chen Z, Chen YZ, Wang XF, Wang C, Yan RX, Zhang Z. 2011. Prediction of ubiquitination sites by using the composition of k-spaced amino acid pairs. *PLoS One* 6:e22930.

Clark AG. 1997. Neutral behavior of shared polymorphism. *Proc Natl Acad Sci U S A.* 94:7730–7734.

Claw KG, Tito RY, Stone AC, Verrelli BC. 2010. Haplotype structure and divergence at human and chimpanzee serotonin transporter and receptor genes: implications for behavioral disorder association analyses. *Mol Biol Evol.* 27:1518–1529.

Cooper GM, Stone EA, Asimenos G, NISC Comparative Sequencing Program, Green ED, Batzoglou S, Sidow A. 2005. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 15:901–913.

Ebers GC. 2008. Environmental factors and multiple sclerosis. *Lancet Neurol.* 7:268–277.

Egloff MP, Malet H, Putics A, et al. (12 co-authors). 2006. Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains. *J Virol.* 80:8493–8502.

Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM, Tishkoff SA, Hudson RR, Lahn BT. 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* 309:1717–1720.

Frisse L, Hudson RR, Bartoszewicz A, Wall JD, Donfack J, Di Rienzo A. 2001. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am J Hum Genet.* 69:831–843.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.

Fumagalli M, Cagliani R, Pozzoli U, Riva S, Comi GP, Menozzi G, Bresolin N, Sironi M. 2009. Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Res.* 19:199–212.

Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Bresolin N, Clerici M, Sironi M. 2010. Genome-wide identification of susceptibility alleles for viral infections through a population genetics approach. *PLoS Genet.* 6:e1000849.

Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Riva S, Clerici M, Bresolin N, Sironi M. 2009. Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *J Exp Med.* 206:1395–1408.

Gabriel SB, Schaffner SF, Nguyen H, et al. (18 co-authors). 2002. The structure of haplotype blocks in the human genome. *Science* 296:2225–2229.

Gao G, Guo X, Goff SP. 2002. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* 297:1703–1706.

Garrigan D, Hammer MF. 2006. Reconstructing human origins in the genomic era. *Nat Rev Genet.* 7:669–680.

1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073.

Gilad Y, Bustamante CD, Lancet D, Paabo S. 2003. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am J Hum Genet.* 73:489–501.

Glazko GV, Nei M. 2003. Estimation of divergence times for major lineages of primate species. *Mol Biol Evol.* 20:424–434.

Gongora R, Figueroa F, Klein J. 1996. The HLA-DRB9 gene and the origin of HLA-DR haplotypes. *Hum Immunol.* 51:23–31.

Gorbalenya AE, Koonin EV, Lai MM. 1991. Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Lett.* 288:201–205.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* 185:862–864.

Griffiths RC, Tavare S. 1994. Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci.* 344:403–410.

Griffiths RC, Tavare S. 1995. Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Math Biosci.* 127:77–98.

Guo X, Carroll JW, Macdonald MR, Goff SP, Gao G. 2004. The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. *J Virol.* 78:12781–12787.

Guo X, Ma J, Sun J, Gao G. 2007. The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. *Proc Natl Acad Sci U S A.* 104:151–156.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5:e1000695.

Hudson RR. 2001. Two-locus sampling distributions and their application. *Genetics* 159:1805–1817.

Kendall MG. 1976. Rank correlation methods. London: Griffin.

Kerns JA, Emerman M, Malik HS. 2008. Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genet.* 4:e21.

Kumar S, Filipski A, Swarna V, Walker A, Hedges SB. 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc Natl Acad Sci U S A.* 102:18842–18847.

Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P. 1988. HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* 335:268–271.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 25:2744–2750.

Marth GT, Czabarka E, Murvai J, Sherry ST. 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* 166:351–372.

Mayer WE, Jonker M, Klein D, Ivanyi P, van Seventer G, Klein J. 1988. Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO J.* 7:2765–2774.

McDonald WI, Compston A, Edan G, et al. (16 co-authors). 2001. Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann Neurol.* 50:121–127.

McIntire JJ, Umetsu SE, Macaubas C, et al. (13 co-authors). 2003. Immunology: hepatitis A virus link to atopic disease. *Nature* 425:576.

McVean G, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160:1231–1241.

Morin PA, Moore JJ, Chakraborty R, Jin L, Goodall J, Woodruff DS. 1994. Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* 265:1193–1201.

Muller S, Moller P, Bick MJ, Wurr S, Becker S, Gunther S, Kummerer BM. 2007. Inhibition of filovirus replication by the zinc finger antiviral protein. *J Virol.* 81:2391–2400.

Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* 76:5269–5273.

Neuvonen M, Ahola T. 2009. Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. *J Mol Biol.* 385:212–225.

Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31:3812–3814.

O'Brien M, Lonergan R, Costelloe L, et al. (12 co-authors). 2010. OAS1: a multiple sclerosis susceptibility gene that influences disease severity. *Neurology* 75:411–418.

Ouararhni K, Hadj-Slimane R, Ait-Si-Ali S, Robin P, Mietton F, Harel-Bellan A, Dimitrov S, Hamiche A. 2006. The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity. *Genes Dev.* 20:3324–3336.

Poser CM. 2006. Revisions to the 2001 McDonald diagnostic criteria. *Ann Neurol.* 59:727–728.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, Balloux F. 2005. Pathogen-driven selection and worldwide HLA class I diversity. *Curr Biol.* 15:1022–1027.

Pugliatti M, Rosati G, Carton H, Riise T, Drulovic J, Vecsei L, Milanov I. 2006. The epidemiology of multiple sclerosis in Europe. *Eur J Neurol.* 13:700–722.

Purcell S, Neale B, Todd-Brown K, et al. (11 co-authors). 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81:559–575.

R Development Core Team. 2008. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, Goebl MG, Iakoucheva LM. 2010. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins* 78:365–380.

Ray N, Wegmann D, Fagundes NJ, Wang S, Ruiz-Linares A, Excoffier L. 2010. A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Mol Biol Evol.* 27:337–345.

Romanoski CE, Lee S, Kim MJ, et al. (13 co-authors). 2010. Systems genetics analysis of gene-by-environment interactions in human cells. *Am J Hum Genet.* 86:399–410.

Salkind NJ, editor. 2007. Encyclopedia of measurement and statistics. Thousand Oaks (CA): Sage Publications.

Sanna S, Pitzalis M, Zoledziewska M, et al. (41 co-authors). 2010. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nat Genet.* 42:495–497.

Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. 2005. Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* 15:1576–1583.

Sironi M, Clerici M. 2010. The hygiene hypothesis: an evolutionary perspective. *Microbes Infect.* 12:421–427.

Stephens M, Scheet P. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet.* 76:449–462.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 68:978–989.

Stone AC, Griffiths RC, Zegura SL, Hammer MF. 2002. High levels of Y-chromosome nucleotide diversity in the genus pan. *Proc Natl Acad Sci U S A.* 99:43–48.

Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P. 2001. Prediction of deleterious human alleles. *Hum Mol Genet.* 10:591–597.

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

Tan J, Vonrhein C, Smart OS, Bricogne G, Bollati M, Kusov Y, Hansen G, Mesters JR, Schmidt CL, Hilgenfeld R. 2009. The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog.* 5:e1000428.

Thomas PD, Kejariwal A. 2004. Coding single-nucleotide polymorphisms associated with complex vs. mendelian disease:

evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci U S A.* 101:15398–15403.

Thomson G. 1995. Mapping disease genes: family-based association studies. *Am J Hum Genet.* 57:487–498.

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci U S A.* 97: 7360–7365.

Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* 19:2325–2327.

Tishkoff SA, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet.* 4:293–340.

Verrelli BC, Lewis CM Jr, Stone AC, Perry GH. 2008. Different selective pressures shape the molecular evolution of color vision in chimpanzee and human populations. *Mol Biol Evol.* 25:2735–2743.

Verrelli BC, Tishkoff SA, Stone AC, Touchman JW. 2006. Contrasting histories of G6PD molecular evolution and malarial resistance in humans and chimpanzees. *Mol Biol Evol.* 23:1592–1601.

Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci U S A.* 102:18508–18513.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 7:256–276.

Wittke-Thompson JK, Pluzhnikov A, Cox NJ. 2005. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet.* 76:967–986.

Wiuf C, Zhao K, Innan H, Nordborg M. 2004. The probability and chromosomal extent of trans-specific polymorphism. *Genetics* 168:2363–2372.

Wright SI, Charlesworth B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168:1071–1076.

Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li WH. 2003. Low nucleotide diversity in chimpanzees and bonobos. *Genetics* 164:1511–1518.

Zhang Y, Burke CW, Ryman KD, Klimstra WB. 2007. Identification and characterization of interferon-induced proteins that inhibit alphavirus replication. *J Virol.* 81:11246–11255.