Long-Term Balancing Selection at the Antiviral Gene *OAS1* in Central African Chimpanzees

William Ferguson,[1,2] Shira Dvora,[1] Ronald W. Fikes,[1] Anne C. Stone,[3] and Stéphane Boissinot*[,1,2]

[1]Department of Biology, Queens College, the City University of New York

[2]Graduate School and University Center, the City University of New York

[3]School of Human Evolution and Social Change, Arizona State University

***Corresponding author:** E-mail: stephane.boissinot@qc.cuny.edu.

Associate editor: Yoko Satta

Abstract

Oligoadenylate synthetases (OAS) are interferon-induced enzymes that participate in the first line of defense against a wide range of viral infection in animals. Upon activation by viral double-stranded RNA, OAS synthesizes (2–5) oligoadenylates, which activate RNase L, leading to the nonspecific degradation of cellular and viral RNA. Some association studies in humans suggest that variation at one of the *OAS* genes, *OAS1*, could be influencing host susceptibility to viral infection. We assessed the diversity of *OAS1* in hominoid primates with a focus on chimpanzees. We found that the *OAS1* gene is extremely polymorphic in Central African chimpanzee and exhibits levels of silent and replacement diversity much higher than neutral regions of the chimpanzee genome. This level of variation strongly suggests that balancing selection is acting on *OAS1*, and indeed, this conclusion was validated by several tests of neutrality. We further demonstrated that balancing selection has been acting at this locus since the split between chimpanzees, humans, and gorillas (~8.6 Ma) and caused the persistence of two deeply divergent allelic lineages in Central African chimpanzees. These two groups of *OAS1* alleles differ by a large number of amino acids (a.a.), including several a.a. putatively involved in RNA binding. It is therefore very likely that variation at the *OAS1* locus affects the innate immune response of individuals to specific viral infection. Our data strongly suggest that interactions between viral RNA and OAS1 are responsible for the maintenance of ancestral polymorphisms at this locus for at least 13.2 My.

Key words: balancing selection, OAS1, chimpanzee.

Introduction

Organisms and their microbial pathogens are engaged in a permanent conflict that has dramatically affected the evolution of their genes and genomes (Woolhouse et al. 2002; Hill 2006; Worobey et al. 2007). The pattern of evolution of immunity genes can tell us a great deal about the nature of host–pathogen interactions because these interactions leave the signature of natural selection on immunity genes. For instance, some host-response genes show a high rate of replacement substitutions relative to synonymous substitutions resulting from the succession of selective sweeps of resistance alleles. Some other genes, such as genes of the major histocompatibilty complex (MHC), are maintained in a polymorphic equilibrium either because heterozygote genotypes have a selective advantage or because variations in frequency of pathogenic strains cause cyclic changes in the frequency of resistance alleles (Figueroa et al. 1988; Lawlor et al. 1988; Mayer et al. 1988; Edwards et al. 1997). The evolutionary analysis of immunity genes is of crucial biomedical importance because polymorphisms at those genes can account for differences in susceptibility to infection (Worobey et al. 2007). For instance, deletion Δ32 in the *CCR5* gene confers resistance to infection with HIV-1 (Liu et al. 1996), and *Human Leucocyte antigen* class I alleles influence the clinical outcome of exposure to Dengue fever virus (Stephens et al. 2002). However, for most infections, in particular viral infections, the genetic factor(s) responsible for differences in susceptibility have not been identified.

Oligoadenylate synthetases (OAS) are interferon-inducible enzymes that are found in all animals, from sponges to mammals. OASs participate in the first line of defense against a wide range of viral infections (for reviews on OAS, see Goodbourn et al. 2000; Hovanessian and Justesen 2007). Upon infection, *OAS* genes are upregulated by interferon and are activated by a double-stranded (ds) RNA-mediated process. Production of dsRNA by a virus is believed to be the trigger of oligoadenylate synthesis in infected cells. Activated OAS proteins synthesize, from ATP, 2′,5′-linked oligoadenylates. Oligoadenylate molecules activate RNase L, which, in turn, degrades cellular and viral RNA thus preventing viral replication. In addition, the activation of RNase L by oligoadenylates leads to a stimulation of transcription for a number of genes that suppress viral replication (Malathi et al. 2005). In humans, the *OAS* gene family has three enzymatically active copies, *OAS1*, *OAS2*, and *OAS3*, resulting from duplication events that occurred before the origin of mammals (Kumar et al. 2000). *OAS1* is upregulated during the early stage of infection by a wide range of viruses and has been shown to play an antiviral role during infection with encephalomyocarditis virus, vaccinia virus, respiratory syncytial virus, and HIV-1

(Gribaudo et al. 1991; Schroder et al. 1994; Diaz-Guerra et al. 1997; Behera et al. 2002). Because OAS proteins interact directly with viral RNAs, it is plausible that changes in the OAS proteins can affect the protection conferred by the OAS pathway. Although the interactions between OAS proteins and dsRNAs seem relatively nonspecific, it has been suggested that polymorphisms at the *OAS1* gene can affect host susceptibility to specific viral infection. Two studies have reported an association between single nucleotide polymorphisms (SNPs) in *OAS1* and the outcome of infection with hepatitis C, West Nile virus, and the virus that causes severe acute respiratory syndrome (SARS) (Knapp et al. 2003; Hamano et al. 2005; Lim et al. 2009). In another study, an SNP at a splice acceptor site in the *OAS1* gene was shown to be responsible for variations of the enzymatic activity of OAS1, thus demonstrating that a polymorphism in the *OAS1* gene has a direct phenotypic effect (Bonnevie-Nielsen et al. 2005). However, the significance of these studies is difficult to assess without a better understanding of the evolutionary history of the *OAS1* gene in human and related species.

Because *OAS1* plays a major role in antiviral resistance, we decided to examine its evolution and diversity in hominoid primates. *OAS1* was sequenced in all species of hominoid primates with a focus on the common chimpanzee, for which individuals from three of the four subspecies were analyzed. We found that the OAS1 gene is extraordinarily variable in Central African chimpanzees and that some polymorphisms are older than the split between chimp, gorilla, and human, ~8.6 Ma (Steiper and Young 2006). We determined that this pattern of variation results from the action of long-term balancing selection.

Materials and Methods

Sampling

We obtained DNA from 37 chimpanzees (*Pan troglodytes*), including representative of the Central African subspecies, *P. t. troglodytes* (7 individuals), the Eastern African subspecies, *P. t. schweinfurthii* (3 individuals), and the Western African subspecies *P. t. verus* (26 Individuals). An additional individual carrying a "Nigerian" mitochondrial DNA (mtDNA) haplotype, possibly belonging to the *P. t. ellioti* subspecies, was sequenced but is not included in our analysis as inferences made from a single individual are not meaningful. The list of individuals and their origin is available as supplementary table 1 (Supplementary Material online). A number of the individuals used in this study had been previously analyzed for their mtDNA, Y chromosome, and autosomal microsatellites (supplementary table 1, Supplementary Material online). When necessary, the subspecific origin of individuals was confirmed by sequencing of the mtDNA control region. In addition, we analyzed eight humans of various ethnic origins (*Homo sapiens*), seven bonobos (*Pan paniscus*), three gorillas (*Gorilla gorilla*), two orangutans (*Pongo pygmaeus*), five olive baboons (*Papio anubis*), and one Rhesus macaque (*Macaca mulatta*).

Sequencing and Haplotype Determination

The coding sequence of the p42 isoform (containing exons 1, 2, 3, 4, and 5) of the *OAS1* gene was obtained for all individuals. Each exon was amplified independently by polymerase chain reaction (PCR) using primers located in introns. *OAS1* belongs to a multigenic family, yet the three copies (*OAS1*, *2*, and *3*) originated before the mammalian radiation and are so divergent that primers located in introns are unambiguously locus specific. The PCR products were treated with Shrimp alkaline phosphatase and exonuclease I, followed by phenol/chloroform extraction and ethanol precipitation. The PCR products were directly sequenced using the Big-Dye terminator chemistry by Macrogen (Seoul, Korea). Haplotypes were inferred by the method of Stephens et al. (2001) implemented in the program PHASE. Individuals with the lowest probability of correct haplotype call were selected for further experimental haplotyping. Experimental haplotyping was accomplished by amplifying long fragments that encompass polymorphisms for which the linkage is unclear. Individual clones were then sequenced allowing the determination of the chromosomal phase by comparison with the diploid sequence. The molecular information obtained for the individuals with the lowest PHASE call was then incorporated into the PHASE analysis, and an additional round of statistical inference was performed. This procedure was repeated until 95% of the individuals had a PHASE call with a confidence interval higher than 95%. Because each exon is amplified independently, we verified that they all belonged to the same gene and not to some unidentified OAS1-related locus. To this end, we sequenced the *OAS1* cDNA in two individuals carrying the two major classes of alleles (see Results). Chimpanzee cell cultures were purchased from the Coriell repository. The two cultures we obtained were from a bonobo (*Pan paniscus*) and from a Western African chimpanzee (*P. t. verus*). Cultures were propagated using Dulbecco's modified eagles medium with specific percentages of heat-inactivated fetal bovine serum. The cells were grown until confluent and then induced overnight using recombinant interferon-α. Total RNA was extracted using Trizol, followed by phenol/chloroform extractions and ethanol precipitations. The RNA was treated with DNase, and its integrity was checked on a 2% agarose gel stained with ethidium bromide. cDNA was generated by reverse transcription of the total RNA with Moloney murine leukemia virus reverse transcriptase and then using DNA polymerase for the second strand synthesis. The most common isoform of *OAS1* (isoform p42) was amplified by PCR, purified, and sequenced. The list of primers used for amplification is available on request from the corresponding author. Sequences are available in the EMBL database under accession number FM163403.1–FM163432.1.

Data Analysis

Sequences were aligned and manipulated using the BioEdit platform (Hall 1999). Because *OAS1* belongs to a multigenic family, we tested the possibility that gene conversion between *OAS1* and *OAS2* or *OAS3* could have affected the

sequence of *OAS1* alleles using the method of Sawyer (1989) implemented in the GENECONV program. The nucleotide diversity in *OAS1* was estimated using the parameters π and Watterson's θ. Tajima's *D* (Tajima 1989) and Fu and Li's (Fu and Li 1993) were calculated to assess the effect of selection on polymorphisms. Statistical significance of these parameters was assessed by coalescent simulations with the software HKA (http://genfaculty.rutgers.edu/hey/software#HKA) with 10,000 iterations. We tested for selection using the Hudson–Kreitman–Aguade test (HKA test; Hudson et al. 1987). This test is based on the assumption that, under neutrality, polymorphism and divergence should be the same across the genome. The test compares the ratio of polymorphism to divergence between a gene of interest and a neutral region of the genome. If the difference between the two ratios is significant using a goodness-of-fit test, we can reject the hypothesis of neutrality. To take into account interloci variation, we performed a multilocus HKA (implemented in J. Hey's HKA software) using 50 noncoding regions sequenced by Yu et al. (2003). We also tested the hypothesis that under neutrality, the ratio of nonsynonymous to synonymous polymorphism within species is identical to the ratio of nonsynonymous to synonymous fixed differences (i.e., MK test; McDonald and Kreitman 1991). We calculated the neutrality index (N.I.) (Rand and Kann 1996) defined as N.I. = (number of polymorphic replacement/number of fixed replacement)/(number of polymorphic silent sites/number of fixed silent sites). Under neutrality, this ratio should be equal to one. The extent of linkage disequilibrium (LD) was assessed by calculating the correlations (R^2) for all pairwise comparisons between SNPs. Significant correlations were determined with χ^2 tests using a Bonferroni correction. The four-gamete test (Hudson and Kaplan 1985) was also used to estimate the minimum number of recombination events in our sample. Population genetics calculations were performed using the DnaSP program. Synonymous (ds) and nonsynonymous (d*n*) distances, as well as their ratio ω, were estimated by maximum likelihood using the model of Yang and Nielsen (2000) implemented in PAML (Yang 2000). The time to the most recent common ancestor (TMRCA) and its deviation was assessed using two methods. First, we used the method implemented in NETWORK 4.5 (Bandelt et al. 1999) based on the average distance from the root (Morral et al. 1994). Usually, mutation rates are inferred using the number of fixed difference between the species of interest and an outgroup. This approach can, however, result in an underestimation of the mutation rate in the case of very deep allelic divergence in the focal species. Thus, we used two estimates, one based on the average divergence between chimpanzee OAS1 and orangutan (used as an outgroup) and one calculated from the number of fixed differences between human (a species for which we found no divergent alleles) and orangutan, assuming a separation time of orangutan and African apes of 18.3 My (Steiper and Young 2006). Additional mutation rates were estimated using baboon or macaque as outgroup, but they did not have any significant impact on the analysis (data not shown). A second estimate of the TMRCA was calculated using a maximum-likelihood coalescent method implemented in GENETREE (Griffiths and Tavare 1994). Mutation rates were inferred as described above, and a generation time of 20 years was assumed. Using these assumptions, the coalescence time and its standard deviation (SD), scaled in $2N_e$ units, was converted into time. Recombinant haplotypes were removed from the TMRCA calculations.

Results

We assessed the variation of the OAS1 gene in a sample of 37 chimpanzees (*Pan troglodytes*) representing three of the four recognized subspecies: the Western African chimpanzee, *P. t. verus* (26 individuals), the Central African chimpanzee, *P. t. troglodytes* (7 individuals), and the Eastern African chimpanzee, *P. t. schweinfurthii* (3 individuals). For comparison, *OAS1* sequences were obtained from seven bonobos (*P. paniscus*), three gorillas (*G. gorilla*), two orangutans (*P. pygmaeus*), and eight humans from various ethnic origins.

Patterns of Nucleotide Diversity in *OAS1*

We identified a total of 29 SNPs at 28 positions in common chimpanzees (fig. 1). Twenty of those SNPs result in changes at the amino acid level. As the *OAS1* gene is 335 amino acids long, it means that 6 % of all amino acid positions are variable in this species, which is remarkably high for a mammalian gene. Because this number of mutations is unusually elevated, we verified experimentally that the exons that were amplified independently by PCR belong to the same protein-coding unit. We obtained cell lines from a bonobo and a Western chimpanzee. Based on the sequencing of exons, the bonobo culture comes from an individual carrying an *OAS1* allele, which is 99.5% identical at the amino acid level to class II alleles (see below), and the Western African chimpanzee is homozygous for the most common allele found in Western chimpanzee, the A allele (fig. 1). Using a cDNA library generated from the two cell lines, we amplified the p42 isoform of *OAS1* and sequenced it directly. In each species, we obtain a single *OAS1* sequence identical to the one obtained by sequencing of the exons. Thus, we are confident that our PCR amplifications amplify specifically exons from the *OAS1* locus. We then examined the possibility that gene conversion between *OAS1* and *OAS2* or *OAS3* could have affected the sequence of *OAS1* alleles. Using GENECONV, we were unable to detect a single significant instance of gene conversion between *OAS1* and any other member of the *OAS* gene family. Considering that *OAS1*, *OAS2*, and *OAS3* result from duplication events that predate the origin of mammals (Kumar et al. 2000) and are consequently quite different, any event of gene conversion would have been readily detected by GENECONV. Thus, we can be confident that polymorphisms in *OAS1* result from novel mutations in *OAS1* and not from the effect of gene conversion.

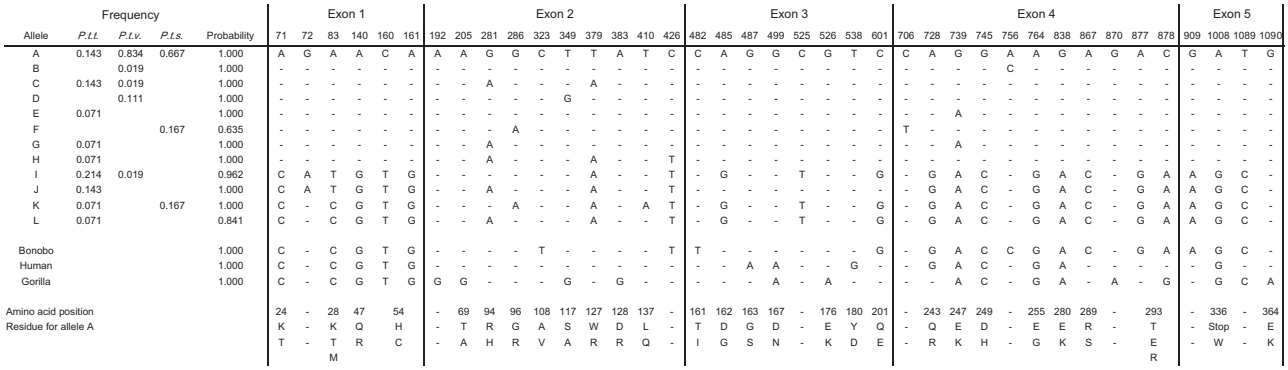

Fig. 1. Variation at the *OAS1* gene in common chimpanzees. The human, gorilla, and pygmy chimpanzee sequences are *OAS1* consensus sequences. The PHASE call probability of each haplotype is indicated.

The nucleotide diversity was calculated for each chimpanzee subspecies using two estimators, π, which estimates the average number of pairwise differences among sequences, and Watterson's θ, which estimates diversity from the number of segregating sites (table 1). Using both estimators, we uncovered extremely high levels of silent diversity in Central ($\pi = 1.01\%$) and Eastern African chimpanzees (0.65%). We also uncovered an even higher level of nucleotide diversity at replacement sites in these two subspecies ($\pi = 1.18\%$ and 0.79% in Central and Eastern African Chimpanzees, respectively), which is consistent with the large number of nonsynonymous SNPs. We compared the silent diversity in *OAS1*, which best reflects the neutral history of the gene, with estimates of diversity in 50 noncoding sequences analyzed by Yu et al. (2003) and 10 noncoding regions analyzed in Fischer et al. (2006) and Gilad et al. (2003) (fig. 2; Eastern chimpanzees are not included because of small sample size). The silent diversity at *OAS1* in Central African chimpanzee falls clearly outside the range of diversity reported for these 60 putatively neutral regions of the chimpanzee genome. *OAS1* diversity at both silent and replacement sites is considerably lower in West African chimpanzee and is well within the range of neutral diversity reported for this subspecies (fig. 2). The diversity in bonobo, gorilla, and human is also lower than in Central and Eastern African chimpanzees (table 1). Thus, the elevated diversity observed in Central and East African chimpanzee is specific for these subspecies and suggests that *OAS1* has experienced radically different evolutionary history among chimpanzee subspecies.

The extraordinarily high level of nucleotide variation in Central and Eastern African chimpanzee strongly suggests that balancing selection, that is, a form of selection that maintains polymorphisms in populations, is acting at the *OAS1* locus. We examined this possibility by testing several predictions associated with balancing selection. If a locus is evolving under balancing selection, we expect an excess of polymorphisms at intermediate frequency in populations. Such an excess of intermediate frequency polymorphism will inflate the parameter π but will not affect θ, which depends solely on the number of segregating sites. Under neutrality, the two estimators of diversity π

Table 1. Summary Statistics for OAS1 Coding Sequences.

Polymorphism	Taxon	N^a	S^b	π^c(%)	θ^d(%)	Tajima's D^e	Fu and Li's D^f
Silent	***Pan troglodytes verus***	**52**	**6**	**0.09**	**0.51**	**−2.24**	**−4.50**
	***P. t. verus* (without Lee)**	**50**	**1**	**0.01**	**0.06**	**−1.10**	**−1.85**
	Pan troglodytes troglodytes	**14**	**5**	**1.01**	**0.61**	**2.46**	**1.33**
	Pan troglodytes schweinfurthii	**6**	**5**	**0.65**	**0.85**	**−1.37**	**−1.40**
	Pan paniscus	**14**	**1**	**0.15**	**0.12**	**0.54**	**0.75**
	Gorilla gorilla	**6**	**3**	**0.44**	**0.36**	**1.12**	**1.40**
	Homo sapiens	**8**	**0**	**0.00**	**0.00**	**NA**	**NA**
Replacement	***P. t. verus***	**52**	**17**	**0.11**	**0.50**	**−2.40**	**−4.79**
	***P. t. verus* (without Lee)**	**50**	**3**	**0.04**	**0.09**	**−1.17**	**−1.74**
	P. t. troglodytes	**14**	**19**	**1.18**	**0.79**	**1.96**	**1.07**
	P. t. schweinfurthii	**6**	**18**	**0.79**	**1.00**	**−1.33**	**−1.35**
	P. paniscus	**14**	**1**	**0.03**	**0.04**	**−0.27**	**0.76**
	G. gorilla	**6**	**2**	**0.16**	**0.16**	**1.03**	**1.28**
	H. sapiens	**8**	**1**	**0.05**	**0.07**	**1.44**	**0.89**

NOTE.—*P* values < 0.05 are indicated in bold. NA, not applicable.

[a] Number of chromosomes.

[b] Number of segregating sites.

[c] Nucleotide diversity.

[d] Watterson's estimator of nucleotide diversity.

[e] Tajima's *D* (Tajima 1989).

[f] Fu and Li's *D* (Fu and Li 1993).

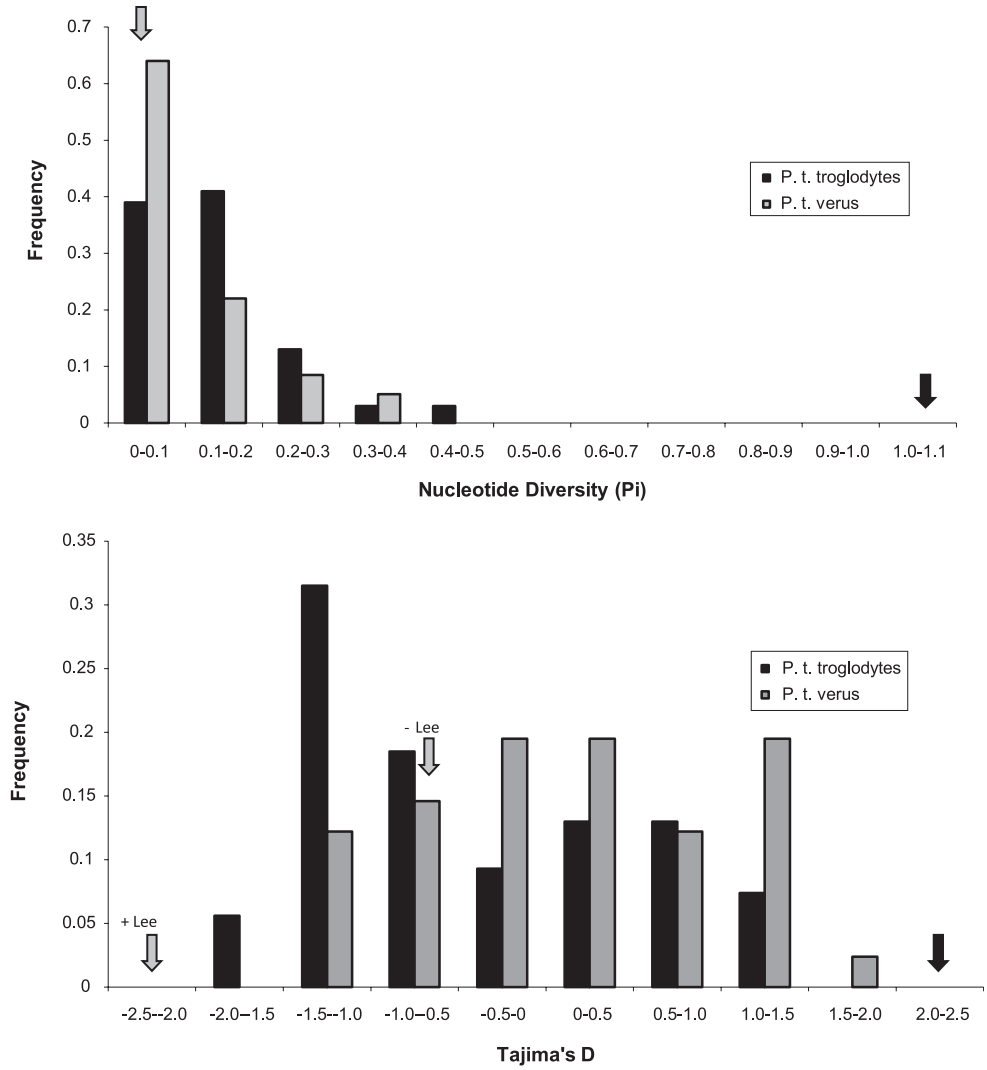

Fig. 2. Frequency distribution of nucleotide diversity (top panel) and Tajima's *D* (bottom panel) in noncoding DNA of the Central (*Pan troglodytes troglodytes*) and Western (*Pan troglodytes verus*) African chimpanzees. The distributions are derived from the analyses of Gilad et al. (2003), Yu et al. (2003), and Fischer et al. (2006). The values of nucleotide diversity and Tajima's *D* for *OAS1* are represented by black (Central chimpanzees) or gray (Western chimpanzees) arrows. For the Western African populations, Tajima's *D* was calculated with (+Lee) or without (−Lee) the single individual (named Lee) carrying a very divergent haplotype (see text).

and θ are expected to be equal, which can be tested using Tajima's *D* statistic. In Central African chimpanzee, Tajima's *D* has a statistically significant positive value ($D = 2.46$; $P < 0.05$) that is higher than any of the values calculated for the 60 noncoding regions of reference (fig. 2; lower panel). A positive value of *D* is consistent with an excess of intermediate-frequency alleles and supports the action of balancing selection, but it can also be interpreted as evidence of changes in population size (e.g., a population bottleneck) or fine-scale population structure. However, demographic events will affect all loci, whereas the effect of natural selection will be locus specific. In Central chimpanzees, the allele frequency spectrum in neutral region of the genome is generally skewed toward rare alleles and negative *D* values (fig. 2; lower panel). Thus, the significantly elevated value of Tajima's *D* at *OAS1* is more likely to be caused by balancing selection than by the demographic history of Central chimpanzees. Tajima's *D* is significantly negative ($D = -2.24$; $P < 0.05$) in the Western chimpanzee, but this is caused by the presence of a single individual (Lee; table 1) carrying a divergent haplotype (haplotype I) most commonly found in Central chimpanzees. If this individual is excluded from the analysis, Tajima's *D* remains negative ($D = -1.10$; $P > 0.10$), but the hypothesis of neutrality cannot be rejected. We also used the test developed by Fu and Li (1993) that compares the number of internal ("old") and external ("young") mutations on a gene genealogy to detect selection. If balancing selection is acting on a locus, an excess of internal mutations relative to external mutations should be detected and should result in positive values of the parameter *D*. This is exactly what we observe in Central African chimpanzee where an excess of internal mutations was found ($D = 1.33$). Although this value falls within the range of values expected under neutrality (−2.43 to 1.45, based on 10,000 coalescent iterations), it is an outlier when compared with the genome average

calculated using the data of Yu et al. (2003) (−1.78 to 1.15). In Western African chimpanzees, a significantly negative value of D was calculated ($D = -4.50$; $P < 0.05$), but this was caused by the presence in the sample of the individual carrying a very divergent haplotype. When this individual is removed, the hypothesis of neutrality is no longer rejected.

We further tested the hypothesis that *OAS1* is evolving under balancing selection in Central and Eastern African chimpanzee by performing a Hudson–Kreitman–Aguadé test (Hudson et al. 1987), also known as HKA test. This test is based on the assumption that, under neutrality, polymorphism and divergence should be the same across all loci. The test compares the ratio of polymorphism to divergence between a gene of interest and a collection of presumably neutral regions of the genome. If the difference between the ratios is significant using a goodness-of-fit test, we can reject the hypothesis of neutrality. As neutral regions of reference, we used the 50 noncoding regions sequenced by Yu et al. (2003), using the baboon as an outgroup. The test was performed using the "HKA" software, distributed by Dr Jody Hey, which allows for calculation of statistical significance using coalescent simulations. We determined that the level of variation at *OAS1* in Central chimpanzees significantly exceeds what is expected given the divergence between chimpanzee and baboon and is not compatible with neutrality, thus confirming the balancing selection hypothesis (sum of deviations = 90.64; $P = 0.0003$). In addition, the test of maximum cell value showed that *OAS1* is an outlier (cell value = 41.53; $P < 0.0001$), and its removal from the data set produced nonsignificant tests. In Western chimpanzees, we were also able to reject the hypothesis of neutrality (sum of deviations = 77.48; $P = 0.0058$), but this result was due to the presence of the same individual (Lee) mentioned above. When Lee is excluded from the analysis, the hypothesis of neutrality is no longer rejected (sum of deviations = 45.38; $P = 0.62$). Note that the results of the HKA test need to be taken with caution because one of the assumptions of the HKA test, the absence of intralocus recombination, does not hold in our case (see below).

Finally, we tested the prediction that if *OAS1* is evolving under balancing selection, a significant excess of polymorphic replacement relative to silent ones is expected (table 2). Under neutrality, the amount of intraspecific polymorphism is correlated to the level of between species divergence, and this should hold for both silent and replacement sites. We compared the level of polymorphism at silent and replacement sites with the level of fixation between chimpanzee and baboon at silent and replacement sites by calculating N.I. (Rand and Kann 1996). In Central and Eastern African chimpanzees, the N.I. reveals a more than 2-fold excess of replacement polymorphisms relative to silent ones. Although these differences are not statistically significant ($P = 0.077$ and 0.139 for the Central and Eastern chimpanzees, respectively), the excess of replacement polymorphisms is consistent with the action of balancing selection in these two subspecies.

Table 2. Polymorphism and Divergence for Silent (*S*) and Replacement (*R*) Sites in Chimpanzees.

	Polymorphisms		Divergence from Baboon			
	R	*S*	*R*	*S*	N.I.	*P*[a]
Pan troglodytes troglodytes	21	5	15	12	2.500	0.077
Pan troglodytes schweinfurthii	19	5	16	12	2.156	0.139
Pan troglodytes verus	19	6	15	12	2.044	0.152
P. t. verus (without Lee)	3	1	24	14	1.361	1.000
Pan paniscus	1	1	23	15	1.087	1.000

[a] *P* values calculated using a two-tailed Fisher exact test.

LD and Age of the Haplotypes

The combination of SNPs results in 13 haplotypes in chimpanzees (fig. 1). A LD analysis (fig. 3*A*) reveals a striking pattern of LD between SNPs located in exon 1 and SNPs located in exons 3 and 4. In fact, most SNPs in exons 1 and 4 are in complete LD, although exon 1 is separated from exon 4 by more than 8 Kb of sequence. Conversely, there is little LD between the SNPs in exons 1 and 2, which are 1.3 Kb apart, and between exons 2 and 3, which are separated by a 2.2 Kb intron. This pattern suggests that recombination has affected the evolution of this gene, and indeed, we estimated using the four gametes test (Hudson and Kaplan 1985) that a minimum of three recombination events is necessary to explain the data. Thus, the nonrandom association of SNPs is not attributable to a lack of recombination, as LD decays in *OAS1* over distances shorter than 1.3 Kb (i.e., the distance between exons 1 and 2). Instead, this pattern suggests that selection could be favoring LD, possibly because of epistatic interactions between polymorphisms in exons 1 and 4.

Based on the combinations of polymorphic a.a. in exons 1 and 4, haplotypes cluster into two clearly distinct classes separated by a long branch on a network analysis (fig. 4): class I (haplotypes A to H) and class II (haplotypes I to L). Class I and II alleles are found at similar frequencies in Central African chimpanzees (0.43 and 0.57, respectively), but class I alleles dominate in Western African chimpanzees (0.98) as a single individual (Lee) carries a class II allele. We calculated the ratio of nonsynonymous to synonymous divergence ω between these classes (excluding recombinant haplotypes), and we found that it was significantly higher than 1 (mean estimate of $\omega = 1.82$). A value of ω higher than 1 indicates that some form of selection in favor of amino acid changes is acting on a gene and is also consistent with the action of balancing selection. We then used two different approaches to determine the TMRCA of the two classes of haplotypes. First, we used the approach implemented in NETWORK to estimate the TMRCA. We used two mutation rates, one based on the average divergence between chimpanzee and orangutan (see Materials and Methods section) and one based on the number of fixed differences between orangutan and human and a divergence time of 18.3 My. These calibrations yielded TMRCA estimates of 13.2 My (SD = 2.7 My) and

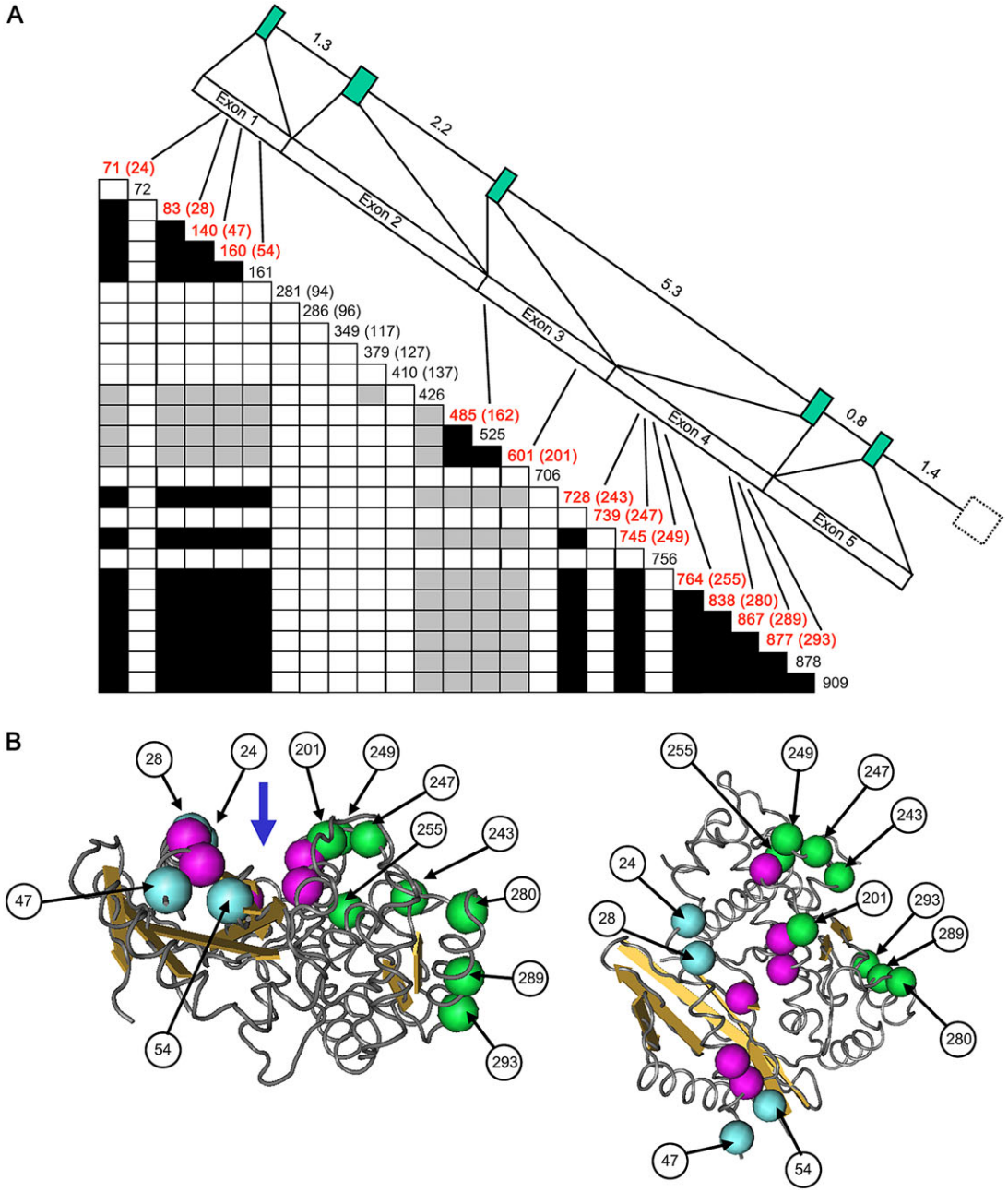

FIG. 3. (*A*) Plot of the association between SNPs across the *OAS1* gene. Black boxes indicate significant LD at the 1% level and gray boxes at the 5% level. The numbers correspond to the nucleotide positions and the number in parenthesis to the amino acid positions. Numbers in red correspond to the a.a. that are mapped on the secondary structure of *OAS1* (bottom panel). The diagram above the plot shows the five exons (in green) constitutive of the p42 isoform of *OAS1*. The numbers above the diagram indicates the length of the introns in Kilobase. (*B*) Secondary structure of the OAS1 protein. The structure was previously resolved by Hartmann et al. (2003). The blue arrow indicates the position of the positively charged groove believed to act as a dsRNA-binding site. The red balls indicate the position of a.a. that were mutated experimentally by Hartmann et al. (2003) to demonstrate their contribution to the protein's ability to bind dsRNA. The blue and the green balls correspond to polymorphic a.a. in exon 1 and in exons 3 and 4, respectively.

16.2 My (SD = 3.3 My), respectively. Second, we used the coalescent approach implemented in GENETREE. This approach assumes an infinite site model without recombination; thus we removed recombinant haplotypes. Using the mutation rates described above and a generation time of 20 years, we obtained estimates for the TMRCA of 15.25 My (SD = 3.1 My; $\Theta_{ML} = 11.29$; $N_e \sim 20{,}000$) and 18.0 My (SD = 3.7 My), which are close to the estimates obtained using the NETWORK approach. These values are older than the split between human, chimpanzee, and gorilla, estimated at 8.6 My, indicating that the origin of the two classes of alleles predate the separation between hominoids.

Discussion

We found that the *OAS1* gene is extremely polymorphic in the Central and Eastern African subspecies of the

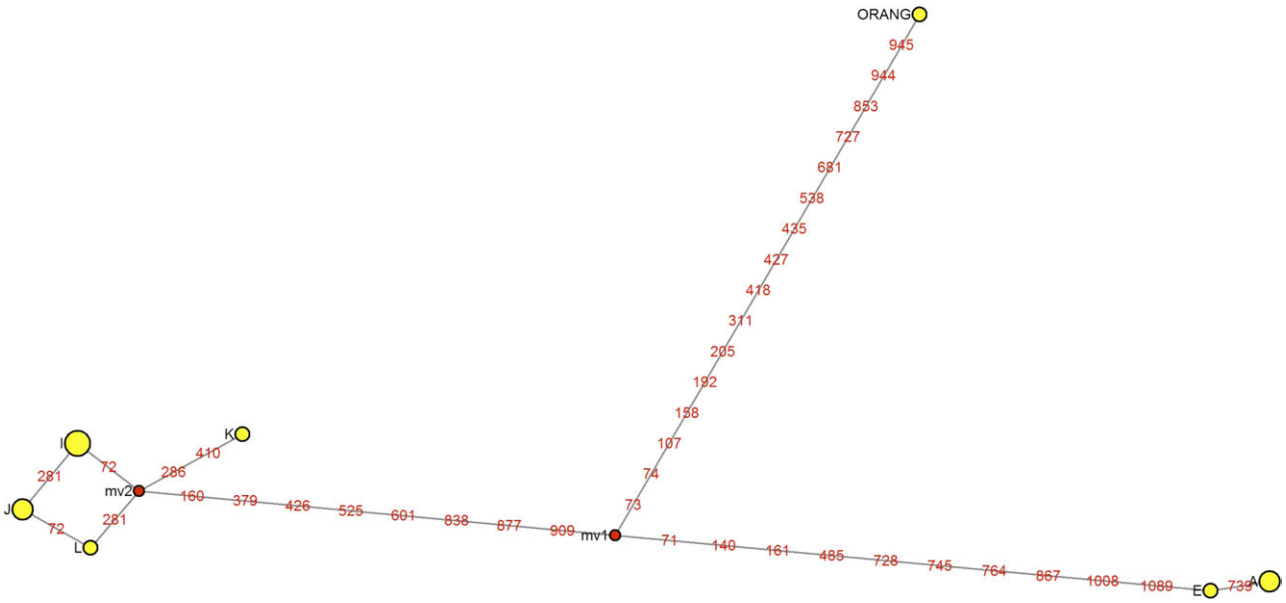

Fig. 4. Genealogy of the *OAS1* haplotypes in Central African chimpanzee reconstructed through a reduced median network. Each yellow circle corresponds to a different haplotype. The ORANG node corresponds to the orangutan OAS1.

chimpanzee and is, in fact, much more variable than neutral regions of the chimpanzee genome. The level of polymorphism we report at *OAS1* is, to our knowledge, the highest in hominoids outside the MHC. Such a high level of polymorphism is extremely unusual and strongly suggests that *OAS1* has evolved under balancing selection. We provide several lines of evidence that support this hypothesis, including a significant excess of intermediate frequency polymorphism, a significant deviation from neutrality as revealed by the HKA test, an excess of replacement polymorphisms relative to silent ones, and a high ratio of nonsynonymous to synonymous substitutions. We further estimate that long-term balancing selection has favored the persistence of two deeply divergent haplotypic lineages for an extended period of evolutionary times, ranging from 13.2 to 18.0 My, depending on the estimates. Thus, the allelic variation at *OAS1* is older than the split between humans, chimpanzees, and gorillas that occurred 8.6 Ma (Steiper and Young 2006), making this polymorphism the oldest one in hominoids outside the *MHC* (Lawlor et al. 1988; Mayer et al. 1988). Balanced polymorphisms are apparently rare in nature, and their persistence across speciation events is even more uncommon (Charlesworth 2006). This is because the signature of balancing selection can easily be erased by recombination and drift. The evolution of novel alleles that sweep the population and replace preexisting polymorphisms (De Groot et al. 2002) as well as drastic reductions in population size can similarly reduce the diversity of loci that have evolved under balancing selection. In hominoids, the only well-established cases of long-term balancing selection resulting in trans-specific polymorphism involve genes of the *MHC* (Lawlor et al. 1988; Mayer et al. 1988), the *TRIM5* locus (Cagliani, Fumagalli, Biasin, et al. 2010), and possibly the *PRIM2* locus (Hodgkinson and Eyre-Walker 2010). The importance of balancing selection in maintaining polymorphisms in population has been a matter of debate since the origin of the field of molecular evolution. Although balancing selection is unlikely to be common (Bubb et al. 2006), recent studies have suggested that it could have had a significant impact on human evolution. For instance, Andres et al. (2009) performed a scan of the human genome and identified conservatively 60 genes harboring the signature of balancing selection. Candidate gene approaches have also detected the signature of balancing selection on a number of loci and the list of genes believed to evolve under balancing selection has considerably increased in recent year (Bamshad et al. 2002; Verrelli et al. 2002; Baysal et al. 2007; Cagliani et al. 2008; Fumagalli et al. 2009; Cagliani, Fumagalli, Riva, et al. 2010). However, in all those cases, the balanced polymorphisms are comparatively much younger than the one at the *MHC* and *OAS1*. The scarcity of ancient balanced polymorphisms in human might result from the demographic history of our species. Humans went through a population bottleneck that caused a reduction in polymorphisms, including balanced ones. Other hominoids (e.g., chimpanzees) do not seem to have suffered a population bottleneck to the same extent humans have (Stone et al. 2002; Fischer et al. 2006; Becquet et al. 2007) and are more likely to have retained ancestral polymorphisms. In particular, Central African chimpanzees have maintained a long-term effective population size much larger than the Western and Eastern subspecies (Hey 2010; Wegmann and Excoffier 2010), which could explain the persistence of balanced polymorphisms in Central populations and the near fixation of class I alleles in Western chimpanzees. In Western chimpanzee, a class II allele was found in a single individual (Lee), whereas all the other individuals carried class I alleles. It is possible that the class II allele in Western chimpanzee results from recent gene flow

from Central to Western chimpanzee, but more population data will be necessary to answer this question. It is also possible that the differences in frequency of class I and II alleles between chimpanzee subspecies results from the exposure of chimpanzees to different viruses in Western and Central Africa. This was previously suggested to explain differences in variation at HIV susceptibility loci (MacFie et al. 2009) among chimpanzee subspecies.

The observation that *OAS1* is evolving under balancing selection implies some functional differences between class I and II haplotypes. We examined this possibility by mapping the polymorphic amino acids (a.a.) on the crystal structure of the OAS1 protein (fig. 3*B*) that was previously resolved by Hartmann et al. (2003) (fig. 3*B*). The dsRNA-binding domain of the OAS1 protein corresponds to a positively charged groove on the back face of the structure (indicated by a blue arrow on fig. 3*B*). It is believed that the binding of dsRNA to this groove leads to conformational changes of the protein necessary for its activation. We found that the four polymorphic a.a. in exon 1 (a.a. 24, 28, 47, and 54) as well as a.a. 201 in exon 3 are located directly in the positively charged grove and could very well be participating in the dsRNA-binding activity of the protein. Interestingly, all these changes involve positively charged a.a. (R, H, and K; fig. 1), and it is plausible that changes at these sites affect the charge and binding property of the protein, although this needs experimental validation. This raises the intriguing possibility that balancing selection at *OAS1* is driven by interactions between the OAS1 protein and viral dsRNA. Thus, it is plausible that class I and II alleles confer different level of protection during viral infection or protect the host against different viral strains. This hypothesis is consistent with studies that found an association between polymorphisms in *OAS1* and differences in susceptibility to SARS, hepatitis C, and West Nile virus (Knapp et al. 2003; Hamano et al. 2005; Lim et al. 2009). This is also consistent with studies in mice that have shown that variation at one *OAS1* homolog, *Oas1b*, accounts for difference in susceptibility against flavivirus infection (Mashimo et al. 2002; Perelygin et al. 2002) among mice strains. Interestingly, we recently discovered that long-term balancing selection is also acting on this gene in mice (Ferguson et al. 2008). However, this model does not account for the persistence of polymorphisms in exon 4 (a.a. 243, 247, 249, 255, 280, 289, and 293). These residues are located on the surface of the protein in a region with no well-defined function but which could play a role in the folding of the protein (Hartmann et al. 2003), and there is no reason to believe that they are involved in dsRNA binding. In fact, an experimental mutation at position 247 (a.a. 246 in Hartmann et al. 2003) does not affect the dsRNA-binding affinity of the protein. The exact role of this region and the impact of these polymorphisms on the function of the protein will require further experiments.

Supplementary Material

Supplementary table 1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to thank those who provided primate samples: the New Iberia Primate Center (supported by an National Center for Research Resources Grant No. U42 RR015087 to the University of Louisiana at Lafayette, New Iberia Research Center), the Primate Foundation of Arizona, the Southwest Foundation for Biomedical Research, the Riverside Zoo, the Detroit Zoo, the Cincinnati Zoo, the Georgia State Language Research Center, the Yerkes National Primate Research Center, Primarily Primates, and the Centre International de Recherches Médicales de Franceville (CIRMF). We thank two anonymous reviewers for their comments on an earlier version of the manuscript. The work was conducted in part with equipment from the Core Facilities for Imaging, Cellular, and Molecular Biology at Queens College. This research was supported by National Institute of Health grant AI058968-01A1 and PSC-CUNY grant 69552-00 38 to S.B. R.W.F. was supported by the National Science Foundation Undergraduate Research Mentoring program at Queens College (award 0731613 to S.B.).

References

Andres AM, Hubisz MJ, Indap A, et al. (12 co-authors). 2009. Targets of balancing selection in the human genome. *Mol Biol Evol.* 26:2755–2764.

Bamshad MJ, Mummidi S, Gonzalez E, et al. (11 co-authors). 2002. A strong signature of balancing selection in the 5′ cis-regulatory region of CCR5. *Proc Natl Acad Sci U S A.* 99:10539–10544.

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

Baysal BE, Lawrence EC, Ferrell RE. 2007. Sequence variation in human succinate dehydrogenase genes: evidence for long-term balancing selection on SDHA. *BMC Biol.* 5:12.

Becquet C, Patterson N, Stone AC, Przeworski M, Reich D. 2007. Genetic structure of chimpanzee populations. *PLoS Genet.* 3:0617–0625.

Behera A, Kumar M, Lockey R, Mohapatra S. 2002. 2′-5′ Oligoadenylate synthetase plays a critical role in interferon-gamma inhibition of respiratory syncytial virus infection of human epithelial cells. *J Biol Chem.* 277:25601–25608.

Bonnevie-Nielsen V, Field LL, Lu S, Zheng DJ, Li M, Martensen PM, Nielsen TB, Beck-Nielsen H, Lau YL, Pociot F. 2005. Variation in antiviral 2′,5′-oligoadenylate synthetase (2′5′AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene. *Am J Hum Genet.* 76:623–633.

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, 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.

Cagliani R, Fumagalli M, Riva S, Pozzoli U, Fracassetti M, Bresolin N, Comi GP, Sironi M. 2010. Polymorphisms in the CPB2 gene are maintained by balancing selection and result in haplotype-preferential splicing of exon 7. *Mol Biol Evol.* 27:1945–1954.

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* 2:e64.
De Groot N, Otting N, Doxiadis G, Balla-Jhagjhoorsingh S, Heeney J, van Rood J, Gagneux P, Bontrop R. 2002. Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. *Proc Natl Acad Sci U S A.* 99:11748–11753.
Diaz-Guerra M, Rivas C, Esteban M. 1997. Inducible expression of the 2-5A synthetase/RNase L system results in inhibition of vaccinia virus replication. *Virology* 277:220–228.
Edwards SV, Chesnut K, Satta Y, Wakeland EK. 1997. Ancestral polymorphism of Mhc class II genes in mice: implications for balancing selection and the mammalian molecular clock. *Genetics* 146:655–668.
Ferguson W, Dvora S, Gallo J, Orth A, Boissinot S. 2008. Long-term balancing selection at the West Nile virus resistance gene, Oas1b, maintains trans-specific polymorphisms in the house mouse. *Mol Biol Evol.* 25:1609–1618.
Figueroa F, Gunther E, Klein J. 1988. MHC polymorphism pre-dating speciation. *Nature* 335:265–267.
Fischer A, Pollack J, Thalmann O, Nickel B, Paabo S. 2006. Demographic history and genetic differentiation in apes. *Curr Biol.* 16:1133–1138.
Fu Y-X, Li W-H. 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.
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.
Goodbourn S, Didcock L, Randall RE. 2000. Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures. *J Gen Virol.* 81:2341–2364.
Gribaudo G, Lembo D, Cavallo G, Landolfo S, Lengyel P. 1991. Interferon action: binding of viral RNA to the 40-kilodalton 2′-5′-oligoadenylate synthetase in interferon-treated HeLa cells infected with encephalomyocarditis virus. *J Virol.* 65:1748–1757.
Griffiths RC, Tavare S. 1994. Simulating probability distribution in the coalescent. *Theor Popul Biol.* 46:131–159.
Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.
Hamano E, Hijikata M, Itoyama S, et al. (15 co-authors). 2005. Polymorphisms of interferon-inducible genes OAS-1 and MxA associated with SARS in the Vietnamese population. *Biochem Biophys Res Commun.* 329:1234–1239.
Hartmann R, Justesen J, Sarkar SN, Sen GC, Yee VC. 2003. Crystal structure of the 2′-specific and double-stranded RNA-activated interferon-induced antiviral protein 2′-5′-oligoadenylate synthetase. *Mol Cell.* 5:1173–1185.
Hey J. 2010. The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol Biol Evol.* 27:921–933.
Hill A. 2006. Aspects of genetic susceptibility to human infectious diseases. *Annu Rev Genet.* 40:469–486.
Hodgkinson A, Eyre-Walker A. 2010. The genomic distribution and local context of coincident SNPs in human and chimpanzee. *Genome Biol Evol.* 2:547–557.
Hovanessian AG, Justesen J. 2007. The human 2′-5′oligoadenylate synthetase family: unique interferon-inducible enzymes catalyzing 2′-5′ instead of 3′-5′ phosphodiester bond formation. *Biochimie* 89:779–788.
Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164.
Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159.
Knapp S, Yee LJ, Frodsham AJ, et al. (12 co-authors). 2003. Polymorphisms in interferon-induced genes and the outcome of hepatitis C virus infection: roles of MxA, OAS-1 and PKR. *Genes Immun.* 6:411–419.
Kumar S, Mitnik C, Valente G, Floyd-Smith G. 2000. Expansion and molecular evolution of the interferon-induced 2′-5′ oligoadenylate synthetase gene family. *Mol Biol Evol.* 17: 738–750.
Lawlor DA, Ward FF, Ennis PD, Jackson AP, Parham P. 1988. HLA-A, -B polymorphisms predate the divergence of humans and chimpanzees. *Nature* 335:268–271.
Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
Lim J, Lisco A, McDermott D, et al. (14 co-authors). 2009. Genetic variation in OAS1 is a risk factor for initial infection with West Nile virus in man. *PLoS Pathog.* 5:e1000321.
Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR. 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86:367–377.
MacFie T, Nerrienet E, de Groot N, Bontrop R, Mundy N. 2009. Patterns of diversity in HIV-related loci among subspecies of chimpanzee: concordance at *CCR5* and differences at *CXCR4* and *CX_3CR1*. *Mol Biol Evol.* 26:719–727.
Malathi K, Paranjape J, Bulanova E, Shim M, Guenther-Johnson J, Faber P, Eling T, Williams B, Silverman R. 2005. A transcriptional signaling pathway in the IFN system mediated by 2′-5′-oligoadenylate activation of RNase L. *Proc Natl Acad Sci U S A.* 102:14533–14538.
Mashimo T, Lucas M, Simon-Chazottes D, Frenkiel MP, Montagutelli X, Ceccaldi PE, Deubel V, Guenet JL, Despres P. 2002. A nonsense mutation in the gene encoding 2′-5′-oligoadenylate synthetase/L1 isoform is associated with West Nile virus susceptibility in laboratory mice. *Proc Natl Acad Sci U S A.* 99:11311–11316.
Mayer WE, Jonker D, Klein D, Ivanyi P, van Seventer G, Klein J. 1988. Nucleotide sequence of chimpanzee MHC class I alleles: evidence for transspecies mode of evolution. *EMBO J.* 7:2765–2774.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351:652–654.
Morral N, Bertranpetit J, Estivill X, et al. (31 co-authors). 1994. The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nat Genet.* 7:169–175.
Perelygin AA, Scherbik SV, Zhulin IB, Stockman BM, Li Y, Brinton MA. 2002. Positional cloning of the murine flavivirus resistance gene. *Proc Natl Acad Sci U S A.* 99:9322–9327.
Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol Biol Evol.* 13:735–748.
Sawyer SA. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol.* 6:526–538.
Schroder H, Kelve M, Muller W. 1994. The 2-5A system and HIV infection. *Prog Mol Subcell Biol.* 14:176–197.
Steiper ME, Young NM. 2006. Primate molecular divergence dates. *Mol Phylogenet Evol.* 41:384–394.
Stephens HA, Klaythong R, Sirikong M, et al. (13 co-authors). 2002. HLA-A and -B allele associations with secondary dengue virus infections correlate with disease severity and the infecting viral serotype in ethnic Thais. *Tissue Antigens*. 60: 309–318.

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.

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

Verrelli BC, McDonald JH, Argyropoulos G, Destro-Bisol G, Froment A, Drousiotou A, Lefranc G, Helal AN, Loiselet J, Tishkoff SA. 2002. Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *Am J Hum Genet.* 71:1112–1128.

Wegmann D, Excoffier L. 2010. Bayesian inference of the demographic history of chimpanzees. *Mol Biol Evol.* 27:1425–1435.

Woolhouse ME, Webster JP, Domingo E, Charlesworth B, Levin BR. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet.* 32:569–577.

Worobey M, Bjork A, Wertheim J. 2007. Point, counterpoint: the evolution of pathogenic viruses and their human hosts. *Annu Rev Ecol Syst.* 38:515–540.

Yang Z. 2000. PAML (phylogenetic analysis by maximum likelihood) version 3.0. London: University College.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol.* 17:32–43.

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