

Evolutionary Analysis Identifies an MX2 Haplotype Associated with Natural Resistance to HIV-1 Infection

Manuela Sironi,^{*,†,1} Mara Biasin,^{†,2} Rachele Cagliani,¹ Federica Gnudi,² Irma Saulle,² Salomè Ibba,² Giulia Filippi,³ Sarah Yahyaei,² Claudia Tresoldi,¹ Stefania Riva,¹ Daria Trabattoni,² Luca De Gioia,³ Sergio Lo Caputo,⁴ Francesco Mazzotta,⁴ Diego Forni,¹ Chiara Pontremoli,¹ Juan Antonio Pineda,⁵ Uberto Pozzoli,¹ Antonio Rivero-Juarez,⁶ Antonio Caruz,⁷ and Mario Clerici^{8,9}

¹Scientific Institute IRCCS E. MEDEA, Bioinformatics, Bosisio Parini, Italy

²Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy

³Department of Biotechnology and Biosciences, University of Milan-Bicocca, Milan, Italy

⁴S. Maria Annunziata Hospital, Florence, Italy

⁵Infectious Diseases and Microbiology Clinical Unit, Valme Hospital, Seville, Spain

⁶Maimonides Institut for Biomedical Research (IMIBIC), Reina Sofia University Hospital, University of Cordoba, Cordoba, Spain

⁷Immunogenetics Unit, Department of Experimental Biology, University of Jaen, Jaen, Spain

⁸Department of Physiopathology and Transplantation, University of Milan, Milan, Italy

⁹Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

[†]These authors contributed equally to this work.

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

Associate editor: Michael Nachman

Abstract

The protein product of the myxovirus resistance 2 (MX2) gene restricts HIV-1 and simian retroviruses. We demonstrate that MX2 evolved adaptively in mammals with distinct sites representing selection targets in distinct branches; selection mainly involved residues in loop 4, previously shown to carry antiviral determinants. Modeling data indicated that positively selected sites form a continuous surface on loop 4, which folds into two antiparallel α -helices protruding from the stalk domain. A population genetics–phylogenetics approach indicated that the coding region of MX2 mainly evolved under negative selection in the human lineage. Nonetheless, population genetic analyses demonstrated that natural selection operated on MX2 during the recent history of human populations: distinct selective events drove the frequency increase of two haplotypes in the populations of Asian and European ancestry. The Asian haplotype carries a susceptibility allele for melanoma; the European haplotype is tagged by rs2074560, an intronic variant. Analyses performed on three independent European cohorts of HIV-1-exposed seronegative individuals with different geographic origin and distinct exposure route showed that the ancestral (G) allele of rs2074560 protects from HIV-1 infection with a recessive effect (combined $P = 1.55 \times 10^{-4}$). The same allele is associated with lower in vitro HIV-1 replication and increases MX2 expression levels in response to IFN- α . Data herein exploit evolutionary information to identify a novel host determinant of HIV-1 infection susceptibility.

Key words: MX2, adaptive evolution, positive selection, HIV-1 infection.

Introduction

The protein product of the human myxovirus resistance 2 (MX2, also known as MxB) gene was recently shown to act as an inhibitor of HIV-1 infection (Goujon et al. 2013; Kane et al. 2013; Liu, Pan, et al. 2013). MX2 blocks HIV-1 replication at a late postentry step by decreasing nuclear accumulation and chromosomal integration of nascent viral DNA (Goujon et al. 2013; Kane et al. 2013; Liu, Pan, et al. 2013). HIV-1 capsid mutations that alter the nuclear import pathway used by the virus relieve the inhibition operated by MX2 (Goujon et al. 2013; Kane et al. 2013; Liu, Pan, et al. 2013). Thus, MX2 is thought to target the capsid and to interact with other cellular proteins such as peptidylprolyl isomerase A

(cyclophilin A) to block HIV-1 nuclear import (Liu, Pan, et al. 2013).

MX2 and its paralog MX1 (or MxA) belong to the dynamin-like large GTPase superfamily and share a common structure consisting of an N-terminal GTPase domain and a C-terminal stalk connected by a bundle signaling element (fig. 1) (Haller and Kochs 2011). The protein products of the two genes display high identity but different cellular localization and diverse antiviral specificity. MX1 can restrict a wide variety of viruses including orthomyxoviruses (e.g., influenza and Thogoto), bunyaviruses (e.g., La Crosse and rift valley viruses), and hepatitis B virus (Haller and Kochs 2011). Conversely, MX2 is known to counteract retroviruses only (Goujon et al. 2013; Kane et al. 2013; Liu, Pan, et al. 2013),

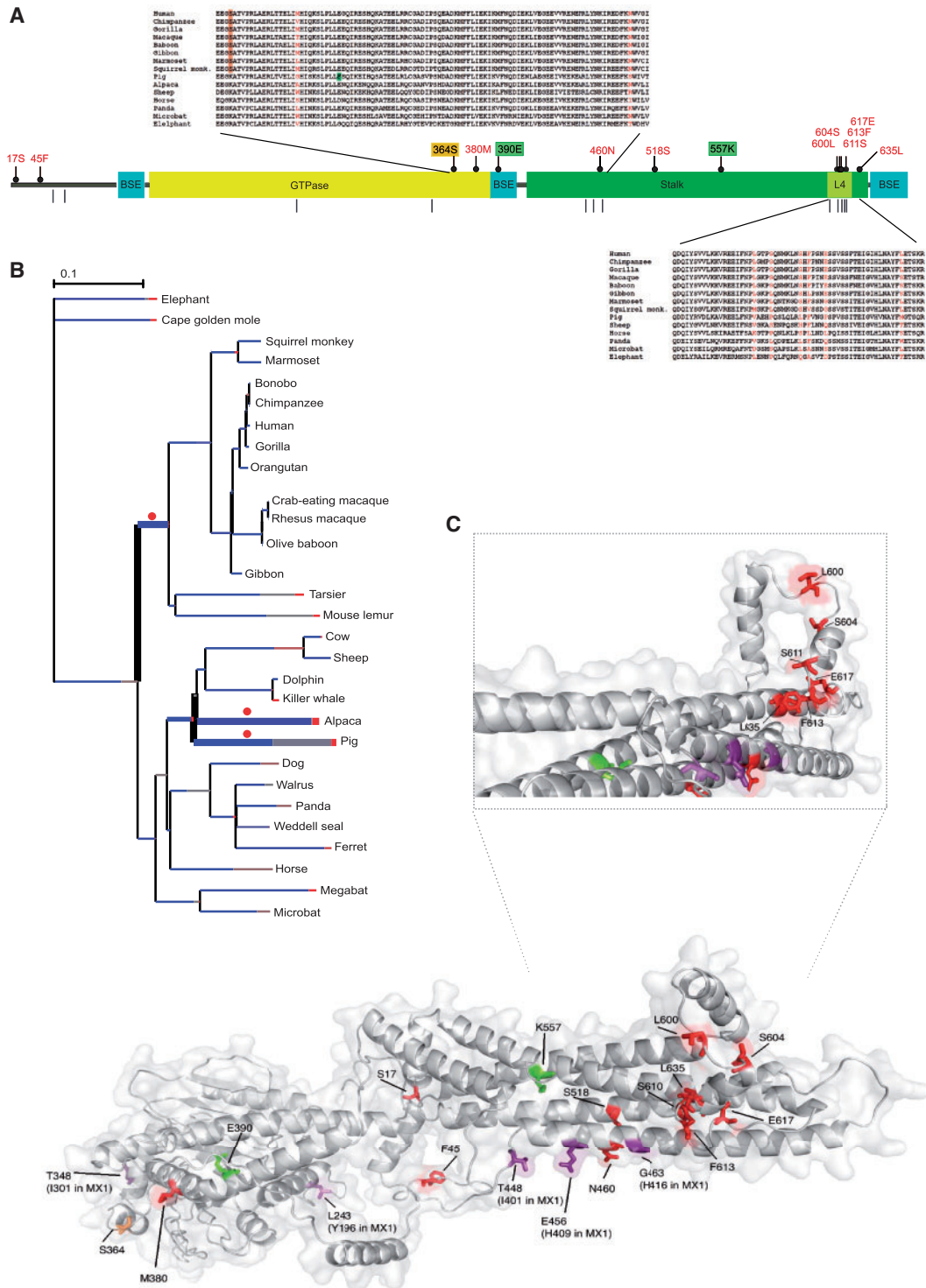


Fig. 1. Adaptive evolution of MX2 in mammals. (A) Schematic representation of the domain structure of MX2 (cyan, bundle signaling element regions; yellow, GTPase domain; dark green, stalk domain with L4 in light green). The position of positively selected MX sites is shown together with sequence alignments for a few representative mammals. Positively selected sites are shown in red (sites identified in the whole phylogeny); primate- and pig-specific sites are in orange and green, respectively. Gray vertical bars below the domain cartoon denote positively selected sites in MX1 (Mitchell et al. 2012). (B) Branch-site analysis of positive selection. Branch lengths are scaled to the expected number of substitutions per nucleotide, and branch colors indicate the strength of selection (ω). Red, positive selection ($\omega > 5$); blue, purifying selection ($\omega = 0$); and gray, neutral evolution ($\omega = 1$). The proportion of each color represents the fraction of the sequence undergoing the corresponding class of selection. Thick branches indicate statistical support for evolution under episodic diversifying selection as determined by BS-REL. Red dots denote branches that were also detected to be under positive selection using the PAML branch-site models (after Bonferroni correction for multiple tests). (C) Structural model of the full-length MX2 protein predicted by I-TASSER. Positively selected sites identified in the whole phylogeny are in red, primate- and pig-specific sites are in orange and green, respectively. Previously identified positively selected sites in the GTPase and stalk domains of MX1 are indicated in magenta. An enlargement of loop 4 and part of the stalk domain is also shown (with rotation).

and to exert a weak effect against vesicular stomatitis virus infection (Liu et al. 2012). Specifically, the human and macaque MX2 proteins efficiently restrict HIV-1 and other simian immunodeficiency viruses, but have a modest effect against retroviruses that infect nonprimate species such as murine leukemia virus and feline immunodeficiency virus (Goujon et al. 2013; Kane et al. 2013). This observation points to species-specific virus–host interactions that may result from evolutionary arms races (i.e., genetic conflicts between the host and the virus), a scenario previously described for MX1 (Mitchell et al. 2012).

Herein, we analyze the evolutionary history of MX2 at the inter- and intraspecific level and use this information to identify a haplotype that associates with natural resistance to HIV-1 infection in humans.

Results

MX2 Evolutionary History in Mammals

To analyze the evolutionary history of MX2 in placental mammals, we obtained coding sequence information for 29 species from public databases (supplementary table S1, Supplementary Material online). Rodent gene sequences were not included because of uncertain orthology (they likely arose by duplication from an ancestral MX1-like gene) (Verhelst et al. 2013). The multiple sequence alignment was screened for recombination using Genetic Algorithm Recombination Detection (GARD) (Kosakovsky Pond et al. 2006), which detected no breakpoint. The average nonsynonymous substitution/synonymous substitution rate (dN/dS , also known as ω) for MX2 amounted to 0.39 (95% confidence intervals [CIs]: 0.37–0.41), indicating a major role for purifying selection (which leads to dN/dS values <1). However, although constraints on protein function and structure often result in purifying selection being the primary evolutionary force acting on gene regions, diversifying selection ($dN/dS >1$) might involve specific sites or domains. Indeed, this has previously been shown to be the case for primate MX1 genes (Mitchell et al. 2012). To test this possibility, we applied maximum Likelihood Ratio Test (LRT) statistics implemented in the phylogenetic analysis by maximum likelihood (PAML) package (Yang 1997, 2007). Specifically, we used the codeml

program to compare models of gene evolution that allow (NSsite models M2a and M8, positive selection models) or disallow (NSsite models M1a and M7, null models) a class of codons to evolve with $dN/dS >1$. As reported in table 1, both null models were rejected in favor of the positive selection models; the same result was obtained using different codon frequency models (F3x4 and F61) (table 1). Thus, MX2 evolved adaptively in placental mammals.

To identify specific sites subject to positive selection, we applied the Bayes Empirical Bayes (BEB) method (with a cutoff of 0.90) (Anisimova et al. 2002; Yang et al. 2005). Because this approach may yield some false positives when a relatively large number of sequences is analyzed (Kosakovsky Pond and Frost 2005), we used the mixed-effects model of evolution (MEME) (with the default cutoff of 0.1) (Murrell et al. 2012) as a second criterion. To be conservative, only sites detected by both methods were considered to be positively selected, although this may result in an underestimation of the actual number of selected sites. These analyses allowed the identification of 11 positively selected sites in MX2 (fig. 1A).

In order to explore possible variations in selective pressure among different lineages, we first tested whether a model that allows dN/dS to vary along branches (model M1) had significant better fit to the data than a model that assumes one same dN/dS across the entire phylogeny (model M0) (Yang and Nielsen 1998). This was indeed the case (table 1), indicating that different mammals experienced variable levels of selective pressure. We thus used the branch-site random-effects likelihood (BS-REL) method (Kosakovsky Pond et al. 2011) to identify lineages on which a subset of sites have evolved under positive selection. BS-REL makes no a priori assumption about which lineages are more likely to represent selection targets; the method identified three branches (fig. 1B). These were cross-validated using the branch-site models implemented in PAML (Zhang et al. 2005), which apply a likelihood-ratio test to compare a model (MA) that allows positive selection on one or more lineages (foreground lineages) with a model (MA1) that does not allow such positive selection (table 2). As suggested (Anisimova and Yang 2007), a Bonferroni correction was applied to these P values, as multiple hypotheses (three in this case) are

Table 1. Likelihood Ratio Test (LRT) Statistics for Models of Variable Selective Pressure among Sites and among Branches.

LRT Model	Codon Frequency Model	Degree of Freedom	$-2\Delta\ln L^a$	P Value ^b	Percentage of Sites with $\omega >1$ (average dN/dS)
M1a versus M2a ^c	F3x4	2	67.83	1.86×10^{-15}	4.3% (2.79)
	F61	2	34.83	2.73×10^{-8}	3.4% (2.42)
M7 versus M8 ^d	F3x4	2	86.04	2.08×10^{-19}	7.9% (1.98)
	F61	2	48.22	3.40×10^{-11}	10% (1.63)
M0 versus M1 ^e	F3x4	56	110.56	1.91×10^{-5}	—
	F61	56	99.00	3.5×10^{-4}	—

^aTwice the difference of the natural logs of the maximum likelihood of the models being compared.

^b P value of rejecting the neutral models in favor of the positive selection models.

^cM1a is a nearly neutral model that assumes one ω class between 0 and 1, and one class with $\omega = 1$; M2a (positive selection model) is the same as M1a plus an extra class of $\omega >1$.

^dM7 is a null model that assumes that $0 < \omega < 1$ is beta distributed among sites; M8 (positive selection model) is the same as M7 but also includes an extra category of sites with $\omega >1$.

^eM0 and M1 are free-ratio models which assume all branches to have the same ω (M0) or allow each branch to have its own ω (M1).

Table 2. Likelihood Ratio Test (LRT) Statistics for Models of Positive Selection on Specific Branches.

Foreground Branch ^a	Codon Frequency Model	Degree of Freedom	$-2\Delta\ln L^b$	P Value ^c	Bonferroni Corrected P Value
Primates	F3x4	1	5.92	0.015	0.045
	F61	1	5.84	0.016	0.048
Alpaca	F3x4	1	31.52	1.97×10^{-8}	5.91×10^{-8}
	F61	1	30.90	2.7×10^{-8}	8.13×10^{-8}
Pig	F3x4	1	139.05	4.29×10^{-32}	1.29×10^{-31}
	F61	1	130.09	4.90×10^{-30}	1.20×10^{-27}

^aMA and MA1 are branch-site models that assume four classes of sites: The MA model allows a proportion of codons to have $\omega \geq 1$ on the foreground branches (those to be tested for selection), whereas the MA1 model does not.

^b $2\Delta\ln L$: Twice the difference of the natural logs of the maximum of the models being compared.

^cP value of rejecting the neutral model in favor of the positive selection model.

being tested on the same data set. PAML analysis confirmed the three branches identified by BS-REL (table 2, fig. 1B). The PAML branch-site models offer the possibility of identifying specific sites that evolved under positive selection in the foreground branches; this is achieved through implementation of a BEB analysis (Zhang et al. 2005). Notably, the simultaneous inference of both the site and the branch subject to diversifying selection is difficult (Zhang et al. 2005; Murrell et al. 2012); thus, BEB analysis is accurate but has low power (Zhang et al. 2005). Also, because identifying sites subject to selection is more difficult than testing whether such sites exist, the branch-site test may provide statistical support for positive selection of the foreground branch(es), but BEB may fail to detect sites with a posterior probability of >0.90 (Zhang et al. 2005). This is the case of the alpaca lineage: We found evidence of episodic selection (fig. 1B and table 2) but failed to identify the selected site(s). Conversely, BEB detected one site in the primate and two sites in the pig branches (fig. 1A). We reasoned that, because MEME was specifically developed to detect episodic positive selection (in addition to pervasive selection), lineage-specific BEB sites should have been identified by the MEME analysis we performed on the whole phylogeny. Indeed, the three sites identified by BEB were also detected by MEME (fig. 1A). These sites thus represent lineage-specific selection targets.

Structural Modeling and Analysis of MX2 Selected Sites

As indicated above, analyses allowed identification of 11 MX2 sites subject to diversifying selection in the whole phylogeny; five of these are located in an unstructured loop (loop 4), where several positively selected sites were also detected in primate MX1 genes (Mitchell et al. 2012) (fig. 1A). To gain further insight into the functional role of positively selected residues, we modeled the three-dimensional structure of MX2 using the known crystal structure of MX1 (63% identity at the protein level) as a template. The use of I-TASSER (Roy et al. 2010) also allowed a reliable ab initio prediction of loop 4 structure (supplementary fig. S1, Supplementary Material online), which includes two antiparallel alpha helices forming a bump protruding from the stalk domain. Positively selected sites in loop 4 form a continuous surface and are mainly located on unstructured loops (fig. 1). Analysis of nonloop 4 selected sites in MX2 and comparison with MX1 selection

targets (Mitchell et al. 2012) indicated that the same region of the stalk domain carries the MX2 460N residue and three selected sites in MX1; also, the S518 residue is spatially close to these residues (fig. 1C). Interestingly, the MX2 primate-specific selection target in the GTPase domain (364S) is in spatial proximity to a site subject to diversifying selection in primate MX1 genes.

Evolution of MX1 and MX2 in the Human Lineage

We next applied a recently developed population genetics–phylogenetics approach to study the evolution of MX1 and MX2 in the human lineage. Specifically, we ran the GammaMap program (Wilson et al. 2011) that jointly uses intraspecific variation and interspecific diversity to estimate the distribution of fitness effects (DFE) (i.e., selection coefficients, γ) along coding regions. To this aim, we exploited data from the 1000 Genomes Pilot project deriving from the low-coverage whole genome sequencing of 179 individuals with different ancestry: Europeans (CEU), Yoruba from Nigeria (YRI), and Chinese plus Japanese (CHBJPT) (1000 Genomes Project Consortium et al. 2010). The ancestral sequence was reconstructed by parsimony from the human, chimpanzee, orangutan, and macaque sequences. We first applied GammaMap to obtain the overall distribution of selection coefficients along MX1 and MX2. A general preponderance of codons evolving under negative selection ($\gamma < 0$) was observed for both genes, with MX1 showing stronger constraint (fig. 2). GammaMap allows to identify specific codons evolving under positive selection. We defined positively selected codons as those having a cumulative probability >0.80 of $\gamma \geq 1$. None was found in either MX1 or MX2. Thus, coding variants in the two genes did not represent positive selection targets during human evolutionary history.

Population Genetic Analysis of MX2 in Humans

The results described above indicate that MX2 has been a selection target throughout the evolutionary history of mammals. In the human lineage, purifying selection had a major role in shaping MX2 coding sequence diversity. Inspection of the 1000 Genomes Project data indicated that one single nonsynonymous variant in MX2 segregates in human populations at a frequency higher than 1% (rs56680307). This single nucleotide polymorphism (SNP) is monomorphic or very rare in non-African populations, whereas it has a minor allele

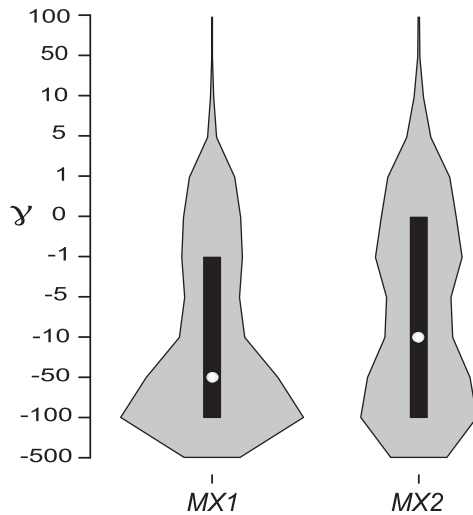


Fig. 2. Lineage-specific selection and DFE analysis. Violin plot of selection coefficients (DFE) for *MX1* and *MX2* (median, white dot; interquartile range, black bar). Selection coefficients (γ) are classified as strongly beneficial (100, 50), moderately beneficial (10, 5), weakly beneficial (1), neutral (0), weakly deleterious (−1), moderately deleterious (−5, −10), strongly deleterious (−50, −100), and inviable (−500).

frequency (MAF) around 20% in Africa. We thus investigated whether natural selection also affected genetic diversity at *MX2* in human populations acting either on rs56680307 or on noncoding variants (as recent data have indicated that regulatory polymorphisms represent the bulk of selection targets in human populations; Grossman et al. 2013; Forni et al. 2014). To this purpose, we initially exploited information from the Human Genome Diversity Panel (HGDP), which consists of approximately 650,000 SNPs genotyped in 52 populations distributed worldwide (Li et al. 2008). Using these data, we calculated F_{ST} among continental groups. F_{ST} is a measure of population genetic differentiation: Under selective neutrality, F_{ST} is mainly determined by demographic history and drift, but natural selection may drive allele frequencies to differ more than expected. Of 22 HGDP variants in *MX2*, three, namely rs45430 (A/G), rs379839 (A/G), and rs2074560 (A/G), displayed an F_{ST} value higher than the 95th percentile calculated from the distribution of HGDP variants in the same MAF class (see Materials and Methods) (fig. 3). Analysis of worldwide variation indicated that ancestral alleles for the three SNPs are almost fixed in African populations, whereas derived alleles reach high frequency outside Africa (fig. 4A). Thus, selective sweeps may have occurred during human colonization of Eurasia and the Americas.

To investigate this possibility further, we used the 1000 Genomes Pilot Project genotype data. Selected variants were searched for by integrating two approaches: One based on population genetic differentiation and the other on haplotype homozygosity. Thus, for all SNPs in *MX2*, we calculated YRI–CEU and YRI–CHBJPT F_{ST} , as well as the derived intra-allelic nucleotide diversity (DIND) test (Barreiro et al. 2009) in CEU and CHBJPT. The rationale behind the DIND test is that a derived allele under positive selection will rapidly increase in frequency in the population

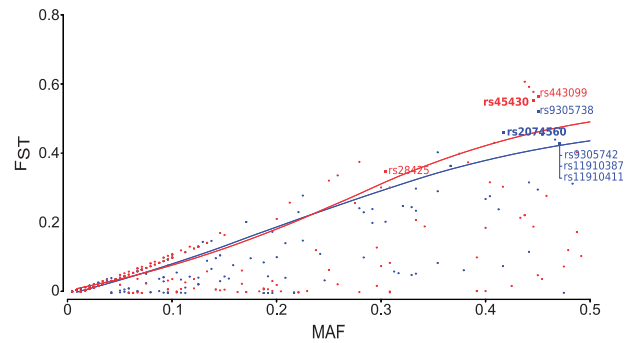


Fig. 3. Population genetic differentiation. F_{ST} analysis of *MX2* variants for YRI–CEU (blue) and YRI–CHBJPT (red) comparisons. Lines represent the 95th percentiles in each MAF class. Solid squares denote variants identified in the HGDP SNP analysis (bold) or SNPs with a significant DIND test.

and will consequently display lower levels of nucleotide diversity at linked sites than expected. Thus, the ratio of intra-allelic diversity associated with the ancestral and derived alleles ($i\pi A/i\pi D$) is analyzed against the frequency of the derived allele (DAF); given a DAF interval, a high value of $i\pi A/i\pi D$ indicates that the neutral diversity associated with the derived allele is limited, suggesting positive selection. DIND has higher power than similar tests in most DAF ranges and is well suited for low-coverage data (Barreiro et al. 2009; Fagny et al. 2014). The statistical significance of both tests was obtained by deriving empirical distributions of the same parameters calculated for a large number of randomly selected genes (see Materials and Methods).

In CEU 5, SNPs were outliers in the F_{ST} comparison (fig. 3) and had a significant DIND test (DAF: 0.75–0.85, DIND: 4.9–7.3). These variants include rs2074560 (derived allele: A) and, despite the high recombination rate in the region, define a homogeneous haplotype that carries derived alleles at several SNPs (fig. 4B). In CHBJPT, two variants had a significantly high F_{ST} value and a significant DIND test (rs28425 and rs443099, DAF: 0.57 and 0.82, DIND: 7.99 and 7.1, respectively) (fig. 3). The derived allele of rs443099 is carried by a group of homogeneous haplotypes, a fraction of which also displays the derived allele for rs28425 (fig. 4B). The rs45430 SNP, which we identified as an outlier in the initial F_{ST} screen, has been associated with melanoma susceptibility in a genome-wide association study (GWAS); this variant is in tight linkage disequilibrium (LD) ($r^2 = 0.95$, $D' = 1$) with rs443099, suggesting that the derived allele, associated with increased melanoma risk, has swept to high frequency outside Africa as a result of selection.

Overall, these results indicate that two distinct *MX2* haplotypes have been targeted by natural selection in Asian and European populations.

Interestingly, ENCODE annotations (ENCODE Project Consortium et al. 2012) deriving from experiments in white blood cells or lymphoblastoid lineages indicated that the CEU- and CHBJPT-selected variants are located in a large region extending downstream the *MX2* transcription start site where several regulatory motifs are located, as assessed by the analysis of DNase hypersensitive regions, transcription factor binding sites, and histone modifications (fig. 4C).

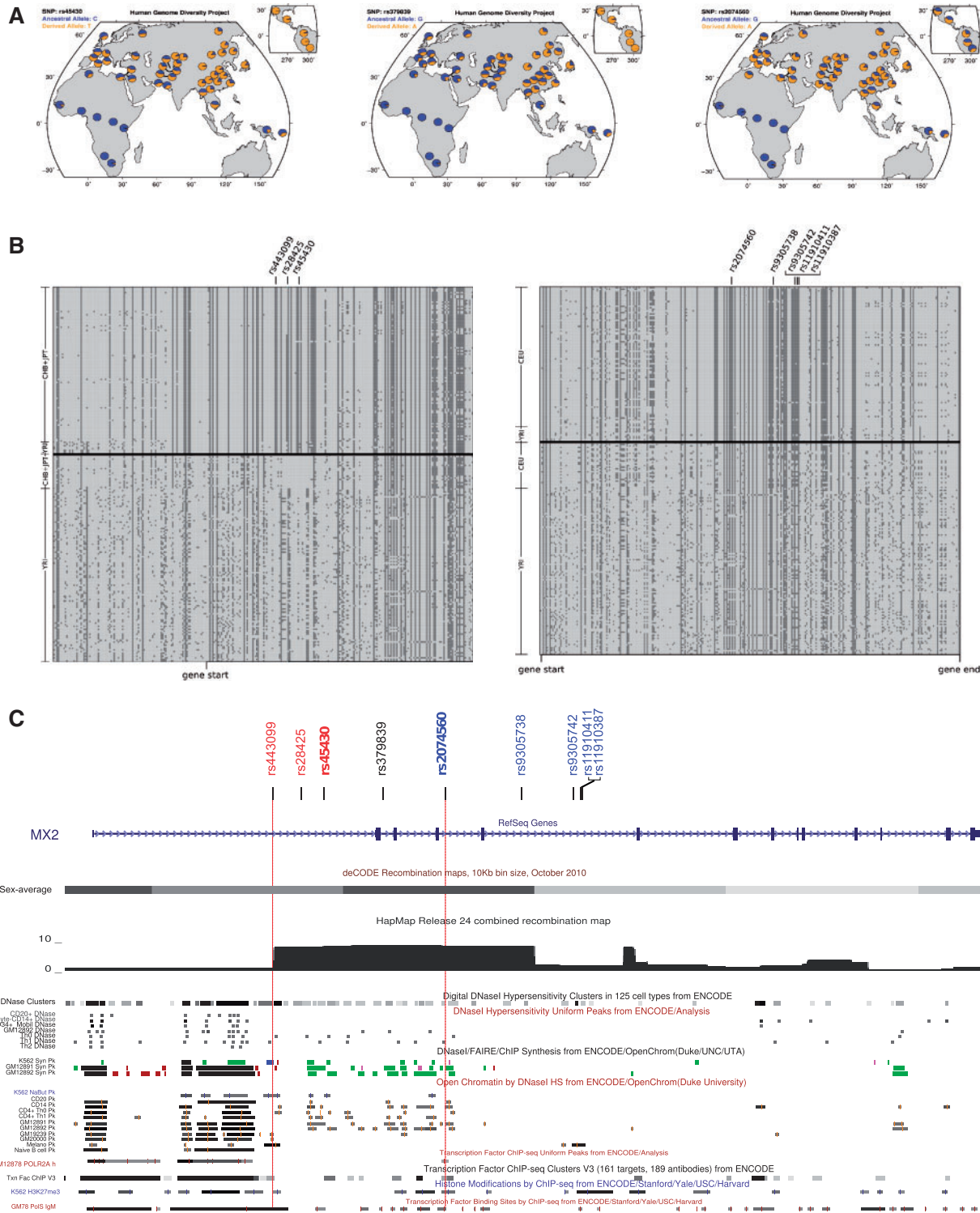


FIG. 4. Natural selection at MX2 in human populations. (A) Worldwide allele frequencies of the three HGDP variants showing outlier F_{ST} values. (B) Schematic representation of CHBJPT plus YRI (left) and CEU plus YRI (right) haplotypes in the regions surrounding rs443099 and rs2074560, respectively. Each line represents a haplotype, columns indicate polymorphic positions. Dark gray, derived alleles, light gray, ancestral alleles. The thick horizontal line separates haplotypes carrying the ancestral (bottom) and derived (top) allele at the selected variants (rs443099 and rs2074560 in the left and right panels, respectively). (C) Schematic representation of MX2 within the UCSC Genome Browser view. The location of the selection targets, recombination rate (Kong et al. 2002), and relevant ENCODE tracks are also shown. SNP color codes are as in figure 3.

Table 3. Association of rs2074560 with HIV-1 Infection Susceptibility.

Sample	Genotype Counts (GG/GA/AA)		$P_{\text{genotypic}}^a$	Genotype Counts (recessive)(GG/AG + AA)		$P_{\text{recessive}}^b$	OR (95% CI) ^c	MetaAnalysis $P_{\text{recessive}}$ and OR ^d
	HESN	CTR ^e		HESN	CTR			
IDU (Spain)	17/27/41	8/33/65	0.036	17/68	8/98	0.0143	3.06 (1.25–7.5)	1.55×10^{-4}
	HESN	HC		HESN	HC			
SexExp (Italy)	11/29/48	6/79/103	0.014	11/77	6/182	0.0052	4.33 (1.55–12.14)	OR: 3.12
SexExp (Spain)	4/15/18	10/78/92	0.5078	4/33	10/170	0.24	2.06 (0.61–6.97)	

^aLogistic regression P value for a genotypic model.

^bLogistic regression P value for a dominant model.

^cOdds ratio for a recessive model with 95% CIs.

^dRandom-effect metaanalysis P value (recessive model) and OR.

^eIDU HIV-1 positive (IDU-CTR in the text).

MX2 Haplotype Association with HIV-1 Infection Susceptibility

Given the recently described role of MX2 as a restriction factor for HIV-1, we investigated whether the positively selected European haplotype modulates susceptibility to HIV-1 infection. Most human subjects are susceptible to this virus, but a minority of individuals do not seroconvert despite multiple exposures. We thus genotyped rs2074560 (G/A) in a cohort of 191 Spanish individuals who were exposed to the virus through injection drug use (IDU). All of them were hepatitis C virus (HCV) positive, but 85 tested HIV-1 negative despite multiple exposures through needle sharing (HIV-1-exposed seronegative, IDU-HESN); the remaining subjects were HIV-1 positive (IDU-CTR). We observed a deviation from Hardy–Weinberg equilibrium (HWE) in IDU-HESN with an excess of homozygotes (HWE, $P = 0.0047$); IDU-CTR complied to HWE. A significant difference was observed in the genotypic distribution of rs2074560 in the two cohorts (table 3), and the frequency of individuals homozygous for the G allele was significantly higher in IDU-HESN (20.0%) compared with IDU-CTR (7.5%). The odds ratio (OR) for a recessive model with the GG genotype being protective against HIV-1 infection was 3.06 (95% CI: 1.25–7.5; logistic regression, $P = 0.014$). In order to replicate this finding, a second HESN population with different geographic origin and exposed to the virus through a different infection route was analyzed as well. Thus, rs2074560 was genotyped in a well-characterized cohort of 88 heterosexual Italian subjects who have a history of unprotected sex with their seropositive partners (sex-exposed HESN, SexExp-HESN). As a control, 188 Italian healthy controls (HCs) were genotyped; both samples complied to HWE. Again, a significant difference was observed in the genotype distribution of rs2074560 (logistic regression, $P = 0.014$) (table 3); similarly to what was observed in the Spanish sample, GG homozygotes were significantly overrepresented in HESN (12.5%) compared with controls (3.4%). Thus, a recessive model yielded a significant association of the GG genotype with HIV-1 protection (logistic regression, $P = 0.005$, OR: 4.33, 95% CI: 1.55–12.14). Further confirmation was sought in a third and smaller population of 37 SexExp-HESN from Spain. These subjects are women who exposed themselves repeatedly through unprotected intercourse with their HIV-1-infected partners. A sample of 180

Spanish HC was also analyzed. The genotype proportions of rs2074560 complied to HWE in both samples. Again, GG homozygotes were much more abundant among SexExp-HESN (10.8%) than among HC (5.5%), thus supporting results obtained in the two other case–control cohorts (table 3). As expected, the small sample size of the HESN population resulted in a nonsignificant P value due to lack of power. The results of the three association analyses were combined through a random-effect metaanalysis, which revealed no heterogeneity among samples (Cochrane's Q , $P = 0.66$, $I^2 = 0$) and yielded a P value of 1.55×10^{-4} (table 3). Overall, these results strongly suggest that the G allele of rs2074560 protects from HIV-1 with a recessive effect, irrespective of the infection route.

MX2 Haplotype Association with In Vitro Viral Infection

To verify whether rs2074560 affects HIV-1 replication in vitro, we performed an infection assay. Specifically, peripheral blood mononuclear cells (PBMCs) from 50 Italian HESN subjects were cultured and infected with HIV-1_{Ba-L}. Viral replication was assessed after 3 days by measuring viral p24 levels produced by the infected cells. Results indicated that the G allele is associated with significantly lower p24 antigen levels (Kruskal–Wallis rank-sum test, $P = 0.034$) (fig. 5).

MX2 Expression in Response to Interferon Treatment

Recent evidences have indicated that MX2 expression is increased in response to interferon alpha (IFN- α) and that intraindividual variability may exist in the level of induction (Goujon et al. 2013; Kane et al. 2013). Thus, we assessed whether allelic status at rs2074560 influences MX2 expression levels in response to IFN- α . To this aim, PBMCs from 45 healthy volunteers were stimulated with IFN- α and MX2 mRNA induction was evaluated by real-time polymerase chain reaction (PCR). Results indicated that the G allele of rs2074560 is associated with a significantly increased MX2 expression in response to IFN- α (Kruskal–Wallis rank-sum test, $P = 0.033$) (fig. 6).

Discussion

The human MX2 protein has long been thought to lack antiviral activity and to serve cellular functions such as

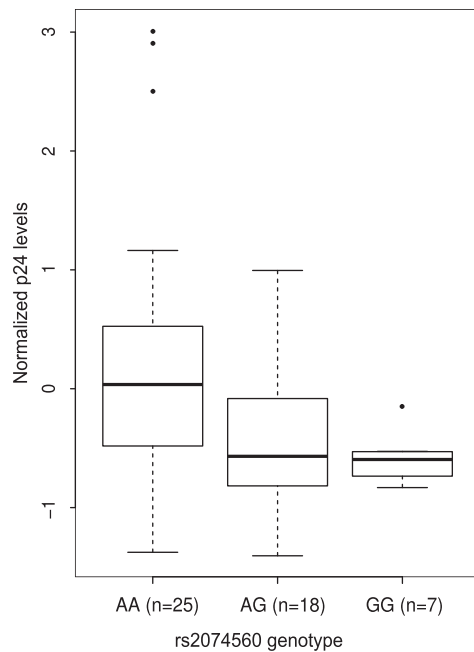


Fig. 5. Box-and-whisker plot of p24 antigen. PBMCs from 50 Italian HESN subjects were infected with HIV-1_{Ba-L}; the levels of viral p24 were measured after 3 days and normalized within experiments. Data are shown as a function of rs2074560 genotype in standard box-and-whisker plot representation (thick line, median; box, quartiles; whiskers, $1.5 \times$ interquartile range).

nucleocytoplasmic transport and cell-cycle progression (King et al. 2004). Recent evidences have challenged this view by showing that MX2 acts as a restriction factor against HIV-1 and other simian retroviruses (Goujon et al. 2013; Kane et al. 2013; Liu, Pan, et al. 2013). Herein, we follow-up on these observations and demonstrate that MX2 evolved adaptively in mammals and in human populations and that a polymorphism in the gene modulates both MX2 expression in response to IFN- α and HIV-1 infection susceptibility.

Antiviral effectors are known to represent common targets of natural selection (Daugherty and Malik 2012), which most likely ensues from host–pathogen genetic conflicts. Thus, determinants of antiviral activity are often accounted for by rapidly evolving residues. In primate MX1 genes, most sites targeted by positive selection are located in loop 4 and explain species-specific difference in antiviral activity against orthomyxoviruses, although additional determinants of antiviral specificity must be accounted for by residues outside loop 4 (Mitchell et al. 2012). The natural selection pattern we observed at MX2 closely reflects the one described for MX1, with loop 4 strongly targeted by natural selection and the same helix surface in the stalk domain carrying three positively selected sites in MX1 and the positively selected 460 residue in MX2 (Mitchell et al. 2012). Interestingly, one of the positively selected sites we identified in MX2, S518 in spatial proximity to the 460 residue, corresponds to a polymorphic position in the pig coding sequence; amino acid variation at this residue modulate the ability of the porcine MX2 protein to inhibit vesicular stomatitis virus replication (Sasaki et al.

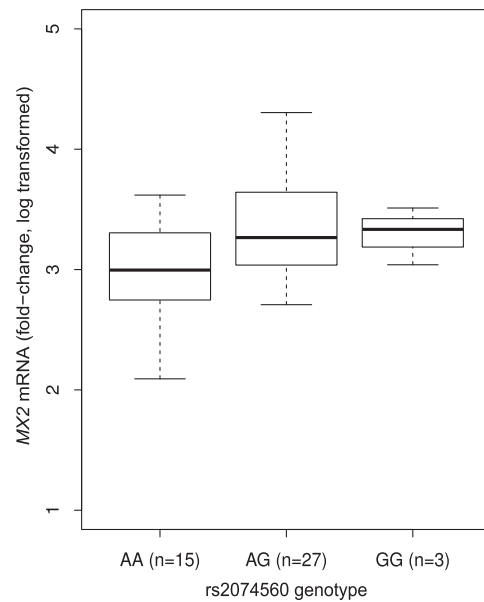


Fig. 6. Box-and-whisker plot of MX2 induction in response to IFN- α . Data derive from IFN- α stimulation of PBMCs deriving from 45 healthy volunteers. MX2 transcript levels are shown as log-transformed fold-change expression from the unstimulated sample. Data are shown in standard box-and-whisker plot representation, as in figure 5.

2014), supporting the hypothesis that positively selected sites determine antiviral activity.

Application of two different methods indicated that lineage-specific selective events shaped diversity at primate MX2 genes. Currently available tests have limited power to simultaneously detect the sites and the lineages subject to episodic positive selection (Zhang et al. 2005; Murrell et al. 2012). Consequently, our analysis cannot be regarded as comprehensive in this respect, and additional sites evolving under episodic selection may have gone undetected. Indeed, we were unable to identify significant sites in the alpaca lineage, and more sites may represent selection targets in primates and pigs. Nonetheless, we consider that the sites we did identify are supported by robust evidence, as they were validated by two distinct methods. Interestingly, one site evolving adaptively in primates is located in the MX2 GTPase domain, in spatial proximity to an MX1 site positively selected in primates (Mitchell et al. 2012). These observations strongly suggest that the GTPase domain might play a role as an antiviral determinant, although its precise function needs further experimental analysis, as MX2 mutants unable to bind or hydrolyze GTP retain their anti-HIV-1 activity (Goujon et al. 2013; Kane et al. 2013; Liu, Pan, et al. 2013). The role of loop 4 residues as antiviral determinants is more clearly established, as single amino acid replacements alter the ability of MX1 to restrict Thogoto virus and influenza A virus (Mitchell et al. 2012). A crystallographic study indicated that MX1 forms ring-shaped oligomeric structures, but failed to solve the structure of loop 4, which was however predicted to face the internal surface of the ring (Gao et al. 2011). Our modeling data indicate that MX2 loop 4 protrudes from the stalk backbone and, were MX2 to adopt the same conformation as

MX1 oligomers, might act as the foremost contact surface with the viral substrate (the capsid protein in the case of HIV-1; Goujon et al. 2013; Kane et al. 2013; Liu, Pan, et al. 2013).

Results herein also indicate a continuum in selective pressure acting in mammals and in human populations, and targeting MX2. Nonetheless, in line with a large-scale analysis of positive selection, which revealed most human selection targets to be located in noncoding regions (Grossman et al. 2013), positive selection acted on MX2 SNPs with a likely regulatory role, whereas the coding sequence was mainly subject to selective constraint in the human lineage. Population genetic analyses indicate that two different haplotypes swept to high frequency in the populations of non-African ancestry; in both cases the result is counter intuitive as selection favored the frequency increase of a melanoma susceptibility allele and of a polymorphism/haplotype associated with decreased IFN- α induction and augmented HIV-1 infection susceptibility. Previous analyses have indicated that risk alleles for human diseases may represent selection targets (Fumagalli et al. 2009; Zhernakova et al. 2010; Jostins et al. 2012; Raj et al. 2012; Forni et al. 2013) and this possibly reflects either changes in environmental pressures or pleiotropic effects of the selected alleles. Thus, the variant we analyzed may differently affect IFN- α response in cell types other than PBMCs. Indeed, rs2074560 is located in a region which likely modulates transcriptional activity in different cell types, as suggested by the presence of DNaseI hypersensitivity peaks and transcription factor binding sites. An alternative explanation is that increased levels of MX2 induction in response to IFN- α is deleterious to some cellular process unrelated to antiviral response, an hypothesis previously proposed to explain the loss of APOBEC3H activity in the populations of non-African ancestry (OhAinle et al. 2008). Conversely, our data indicate that the stronger response to IFN- α afforded by the rs2074560 ancestral allele confers protection in the context of HIV-1 infection. Indeed, we analyzed three populations of HESN subjects with different geographic origin and distinct routes of exposure to the virus. Results were consistent in showing that homozygosity for the G allele at rs2074560 protects from infection. Because association results were replicated in the three independent cohorts, we consider that evidences are strong enough to establish a role for rs2074560 as a determinant of HIV-1 infection susceptibility in humans, despite the relatively small sample of the HESN populations. Moreover, the association results are supported by the *in vitro* infection assay, which indicated that the G allele significantly decreases viral replication. Recently, Lane et al. (2013) performed a GWAS of resistance to HIV-1 infection in highly exposed uninfected hemophiliacs and detected no genome-wide significant association with the exclusion of the known CCR5 Δ 32 allele; nonetheless, given the average MAF of rs2074560 in Europeans (around 0.25), its recessive effect, and the OR estimated herein (OR: 3.12), their study was underpowered to reach genome-wide significance for this variant.

A recent study of HIV-1-infected and seronegative subjects confirmed that CCR5 Δ 32 homozygotes are protected from

infection (OR = 0.2), whereas no effect was detected for the HLA-B*57:01 and *27:05 alleles, both associated with slow progression to AIDS (McLaren et al. 2013). The CCR5 Δ 32 allele is almost exclusively found in Europe and West Asia, where its frequency follows a latitudinal cline (Novembre et al. 2005). In Italy and Spain, the estimated frequency of CCR5 Δ 32 homozygotes is around 0.25–0.5% (Novembre et al. 2005), indicating that this allele makes a minor contribution to HIV-1 susceptibility at the population level. Additional polymorphisms that modulate the susceptibility to HIV-1 infection have been identified in recent years; all of them map to genes involved directly (e.g., ERAP2, IRF1, and TLR3) or indirectly (FREM1 and CYP7B1) in immune response (Ball et al. 2007; Limou et al. 2012; Luo et al. 2012; Sironi et al. 2012; Biasin et al. 2013). Some of these loci showed a relatively strong effect, comparable to that estimated for the MX2 variants we describe herein (Ball et al. 2007; Limou et al. 2012; Luo et al. 2012; Biasin et al. 2013). Nonetheless, most of these analyses, including those reported in our study, were performed on relatively small population samples, because of the intrinsic difficulty in characterizing and recruiting HESN cohorts. Thus, calculation of the effect size on larger samples will be necessary to provide more reliable estimations.

Another point that will require further investigation is the reason(s) why rs2074560, which modulates expression levels, confers protection with a recessive effect. One possibility is that the cellular amount of MX2 protein is rate limiting for efficient HIV-1 restriction, so that a threshold needs to be reached to afford *in vivo* protection. This model might be appealing in light of the proposed model for MX1 antiviral activity, whereby oligomers enclose and sequester viral components (Reichelt et al. 2004; Gao et al. 2011). Another possibility is that, although in the association analysis infection susceptibility was treated as a dichotomous variable (i.e., HESN/non-HESN status), it may be a continuous trait. Thus, the HESN status possibly represents an extreme phenotype and selection of HESN cohorts may result in enrichment for the most protective genotype. This hypothesis is in line with the results from the *in vitro* infection assay, whereby rs2074560 genotype was analyzed against a continuous variable (p24 antigen levels); in this experiment, the effect of the G allele was also evident in heterozygous subjects, and the limited number of GG homozygotes makes it difficult to determine whether an additive model applies to these data.

The observation that the MX2 polymorphism that modulates susceptibility to HIV-1 infection also regulates MX2 expression in response to IFN- α is biologically interesting. MX2 is part of the interferon-regulated genes, a vast and heterogeneous family of genes whose expression is modulated by IFN- α . Some of these genes, including CH25H, were recently shown to be endowed with potent antiviral properties (Liu, Aliyari, et al. 2013). It is thus tempting to envision a scenario in which an initial infection with HIV-1 of target cells that are GG homozygous would trigger the upregulation of MX2; this in turn would block viral replication and impede nuclear accumulation and chromosomal integration of nascent viral DNA. The infection would be aborted and the immune response would be able to clear away the few cells

that had been initially infected, preventing the infection to spread in an uncontrolled fashion. This could explain the puzzling presence of HIV-specific T helper cells and CTL in the absence of any detectable ongoing infection: A phenomenon that is observed in HESN not only during, but, transiently, also after cessation of exposure (Miyazawa et al. 2009).

In summary, building on the recently described role of MX2 as HIV-1 restriction factors, we exploited evolutionary information to identify a variant that confers natural resistance to HIV-1 infection. Results herein establish the role of MX2 as a central element of antiviral response in mammalian species and a possible target for therapeutic intervention in HIV-1 treatment and prevention.

Materials and Methods

Evolutionary Analysis in Mammals

Mammalian sequences for MX2 were retrieved from the Ensembl and NCBI databases (<http://www.ensembl.org/index.html>; <http://www.ncbi.nlm.nih.gov/>, last accessed December 2013). The list of species is reported in [supplementary table S1, Supplementary Material](#) online. DNA alignments were performed using the RevTrans 2.0 utility (Wernersson and Pedersen 2003), which uses the protein sequence alignment as a scaffold to construct the corresponding DNA multiple alignment. This latter was checked and edited by hand to remove alignment uncertainties. For PAML analyses (Yang 2007), we used trees generated by maximum likelihood using the program PhyML (Guindon et al. 2009).

The site models implemented in PAML have been developed to detect positive selection affecting only a few amino acid residues in a protein. To detect selection, site models that allow (M2a, M8) or disallow (M1a, M7) a class of sites to evolve with $\omega > 1$ were fitted to the data using the F3x4 and the F61 codon frequency model. Positively selected sites were identified using the BEB analysis (with a cutoff of 0.90), which calculates the posterior probability that each codon is from the site class of positive selection (under model M8) (Anisimova et al. 2002). A second method, the MEME (with the default cutoff of 0.1) (Murrell et al. 2012), was applied to identify positively selected sites. MEME allows the distribution of ω to vary from site to site and from branch to branch at a site, therefore allowing the detection of both pervasive and episodic positive selection (Murrell et al. 2012).

To explore possible variations in selective pressure among different lineages, we applied the free-ratio models implemented in the PAML package: The M0 model assumes all branches to have the same ω , whereas M1 allows each branch to have its own ω (Yang and Nielsen 1998). The models are compared through likelihood-ratio tests (degree of freedom = total number of branches – 1). In order to identify specific branches with a proportion of sites evolving with $\omega > 1$, we used BS-REL (Kosakovsky Pond et al. 2011) with the PhyML-generated tree. Branches identified using this approach were cross-validated with the branch-site likelihood-ratio tests from PAML (the so-called modified model A and model MA1, “test 2”) (Zhang et al. 2005). A Bonferroni

correction was applied to account for multiple hypothesis testing (i.e., we corrected for the number of tested lineages), as suggested (Anisimova and Yang 2007). The advantage of this method is that it also implements a BEB analysis analogous to that described above to calculate the posterior probabilities that each site belongs to the site class of positive selection on the foreground lineages. Thus, BEB allows identification of specific sites that evolve under positive selection on specific lineages, although it has limited statistical power (Zhang et al. 2005).

GARD (Kosakovsky Pond et al. 2006), MEME (Murrell et al. 2012), and BS-REL analyses were performed through the DataMonkey server (Delpont et al. 2010) (<http://www.data-monkey.org>, last accessed December 2013).

Structural Model Prediction and Validation

Structural models of MX2 were initially predicted using three different methods: MODELLER (Eswar et al. 2006) with loop refinement, I-TASSER (Zhang 2008; Roy et al. 2010) with a defined template (MX1, PDB structure: 3SZR; Gao et al. 2011), or I-TASSER without any template. The quality of each model was assessed with VADAR (Willard et al. 2003). The overall quality was estimated with respect to its geometry and energy (packaging defects, free energy of folding, core hydrophobic and charged residues) ([supplementary table S2, Supplementary Material](#) online). Secondary structures found in loops were validated through the use of PSIPRED (McGuffin et al. 2000) server ([supplementary fig. S1, Supplementary Material](#) online). According to these criteria, the model produced by I-TASSER with MX1 as template was used for our analysis.

Population Genetics–Phylogenetics Analysis

Data from the Pilot 1 phase of the 1000 Genomes Project were retrieved from the dedicated Website (1000 Genomes Project Consortium et al. 2010). 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), the libsequence (Thornton 2003), and the mysqlpp libraries. Coding sequence information was obtained for MX1 and MX2. To analyze the DFE, we used GammaMap (Wilson et al. 2011). We assumed θ (neutral mutation rate per site), k (transitions/transversions ratio), and T (branch length) to vary among genes following log-normal distributions. For each gene, we set the neutral frequencies of non-STOP codons (1/61) and the probability that adjacent codons share the same selection coefficient ($P = 0.02$). For selection coefficients, we considered a uniform Dirichlet distribution with the same prior weight for each selection class. For each gene, we run 10,000 iterations with thinning interval of ten iterations.

Population Genetic Analysis in Humans

HGDP genotype data derive from a previous work (Li et al. 2008). Atypical or duplicated samples and pairs of close

relatives were removed (Rosenberg 2006). F_{ST} among continental groups was calculated for all HGDP SNPs. Because F_{ST} values are not independent from allele frequencies, variants were divided in 100 percentile classes based on MAF and F_{ST} distributions were calculated for each class; outliers were defined as variants with an F_{ST} higher than the 95th percentile in the distribution of SNPs in the same MAF class.

Data from the Pilot 1 phase of the 1000 Genomes Project were retrieved as described above. Genotype information was obtained for MX2 and for 1,200 randomly selected RefSeq genes. Orthologous regions in the chimpanzee, orangutan, or macaque genomes (outgroups) were retrieved using the liftOver tool (<http://hgdownload.cse.ucsc.edu/>, last accessed December 2013); genes showing less than 80% human–outgroup aligning bases were discarded. This originated a final set of 987 genes (control set). F_{ST} (Wright 1950) and the DIND test (Barreiro et al. 2009) were calculated for all SNPs mapping to MX2 and to the control gene sets. For F_{ST} variants were binned variants based on their MAF (100 classes) and the 95th percentile for each MAF class was calculated. As for the DIND test, it was originally developed for application to Sanger or high-coverage sequencing data (Barreiro et al. 2009), so that statistical significance can be inferred through coalescent simulations. This is not the case for the 1000 Genomes Project data; thus, we calculated statistical significance by obtaining an empirical distribution of DIND–DAF value pairs for variants located within control genes. Specifically, DIND values were calculated for all SNPs using a constant number of 40 flanking variants (20 up- and downstream). The distributions of DIND–DAF pairs for YRI, CEU, and CHBJPT were binned in DAF intervals (100 classes, bin = 0.01) and for each class the 95th percentile was calculated. As suggested (Barreiro et al. 2009), for values of $i\pi D = 0$, we set the DIND value to the maximum obtained over the whole data set plus 20.

Subject Cohorts, Genotyping, and Statistical Analysis

We recruited 191 males exposed to HIV-1 infection by IDU and enrolled in prospective cohort studies in Spain (Valme Hospital, Sevilla) who had shared needles for >3 months. Concurrent markers of HCV infection were present in 100% of IDU subjects. Eighty-five of these subjects were HIV-1 negative (IDU-HESN) and 106 were HIV-1 positive (IDU-CTR).

Thirty-eight Spanish HESN exposed to the virus through unprotected sexual intercourse (SexExp-HESN) were also recruited. These subjects are female partners of HIV-1 positive patients (without treatment and viremic, mean number of unprotected sexual intercourse per year: 110, mean number of years as sexual partners: 5, range 3–17). HCs ($n = 180$) that were anonymous blood donors from The City of Jaen Hospital were also recruited. All these individuals were seronegative for both HIV-1 and HCV. All subjects were Spanish of Caucasian origin. The study was designed and performed according to the Helsinki declaration and was approved by the Ethics Committee of the participating hospitals and the University of Jaen. All patients and healthy

blood donors provided written informed consent to participate in this study.

As for Italian SexExp-HESN, inclusion criteria were a history of multiple unprotected sexual episodes for more than 4 years at the time of enrolment, with at least three episodes of at-risk intercourse within 4 months prior to study entry and an average of 30 (range, 18 to >100) reported unprotected sexual contacts per year (Miyazawa et al. 2009). SexExp-HESN and 188 HCs were recruited at the S. M. Annunziata Hospital, Florence; all of them were Italian of Caucasian origin. The study was reviewed and approved by the institutional review board of the S. M. Annunziata Hospital, Florence. Written informed consent was obtained from all subjects.

rs2074560 was genotyped through PCR amplification and direct sequencing (primer sequences are available upon request). PCR products were treated with ExoSAP-IT (USB Corporation, Cleveland OH), directly sequenced on both strands with a Big Dye Terminator sequencing Kit version 3.1 (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. Genetic association analyses were performed by logistic regression and results from the three cohorts were combined using a random-effect metaanalysis; all analyses were performed using PLINK (Purcell et al. 2007).

HIV Infection Assay

PBMCs from 50 HESN subjects were separated on lymphocyte separation medium (Organon Teknica, Malvern, PA); 10×10^6 cells/ml were cultured for 2 days at 37°C and 5% CO₂ in RPMI 1640 containing FBS (20%), phytohemagglutinin (7.5 µg/ml), and interleukin-2 (IL-2) (15 ng/ml). After viability assessment, 2.5×10^6 cells were resuspended in medium containing 1 ng of HIV-1_{Ba-L} p24 viral input/ 10^6 PBMC and incubated for 3 h at 37°C. Cells were then washed and resuspended in 3 ml of complete medium with IL-2 (15 ng/ml). Cells were plated in 24-well tissue culture plates and incubated at 37°C and 5% CO₂. After 3 days supernatants were collected for p24 antigen enzyme-linked immunosorbent assay (ELISA) analyses. Absolute levels of p24 were measured using the Alliance HIV-1 p24 ELISA Kit (PerkinElmer). PBMCs from the 50 subjects were split and infected in three independent experiments. To account for minor differences in virus titer, p24 levels were normalized within experiment. HIV-1_{Ba-L} was provided through the EU programme EVA centre for AIDS Reagents NIBSC, UK.

IFN- α Stimulation and MX2 Transcript Quantification

Whole blood was collected from 45 HCs by venupuncture in Vacutainer tubes containing EDTA (Becton Dickinson, NJ), and PBMCs were separated on lymphocyte separation medium (Organon Teknica, Malvern, PA). Based on data derived from a kinetic study (data not shown), 5×10^5 freshly isolated PBMCs were incubated for 3 h with medium alone or 400 U/ml IFN- α (Sigma Aldrich).

RNA was extracted from cultured PBMC by using the acid guanidium thiocyanate–phenol–chloroform method. The RNA was dissolved in RNase-free water, and purified from genomic DNA with RNase-free DNase (RQ1 DNase; Promega, Madison, WI). One microgram of RNA was reverse transcribed into first-strand cDNA in a 20thinsp;µl final volume containing 1 µM random hexanucleotide primers, 1 µM oligo dT, and 200 U Moloney murine leukemia virus reverse transcriptase (Clontech, Palo Alto, CA). cDNA quantification for MX2 and GAPDH was performed by a real-time PCR strategy (DNA Engine Opticon 2; MJ Research, Ramsey, MN). Reactions were performed using a SYBR Green PCR mix (5 Prime, Gaithersburg, MD). Results were expressed as $\Delta\Delta Ct$ and presented as ratios between the target gene and the GAPDH housekeeping mRNA.

Supplementary Material

Supplementary figure S1 and tables S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The work was supported by the fellowships of the Doctorate School of Molecular Medicine, University of Milan to D.F. and C.P.

References

- 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.
- Anisimova M, Bielawski JP, Yang Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol Biol Evol.* 19:950–958.
- Anisimova M, Yang Z. 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Mol Biol Evol.* 24:1219–1228.
- Ball TB, Ji H, Kimani J, McLaren P, Marlin C, Hill AV, Plummer FA. 2007. Polymorphisms in IRF-1 associated with resistance to HIV-1 infection in highly exposed uninfected Kenyan sex workers. *AIDS* 21: 1091–1101.
- Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O, Gicquel B, et al. 2009. Evolutionary dynamics of human toll-like receptors and their different contributions to host defense. *PLoS Genet.* 5:e1000562.
- Biasin M, Sironi M, Saule I, de Luca M, la Rosa F, Cagliani R, Forni D, Agliardi C, Lo Caputo S, Mazzotta F, et al. 2013. Endoplasmic reticulum aminopeptidase 2 haplotypes play a role in modulating susceptibility to HIV infection. *AIDS* 27:1697–1706.
- 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.
- Daugherty MD, Malik HS. 2012. Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet.* 46:677–700.
- Delpont W, Poon AF, Frost SD, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457.
- ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. 2006. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics.* Chapter 5:Unit 5.6.
- Fagny M, Patin E, Enard D, Barreiro LB, Quintana-Murci L, Laval G. 2014. Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing datasets. *Mol Biol Evol.* 31(7):1850–1868.
- Forni D, Cagliani R, Pozzoli U, Colleoni M, Riva S, Biasin M, Filippi G, De Gioia L, Gnudi F, Comi GP, et al. 2013. A 175 million year history of T cell regulatory molecules reveals widespread selection, with adaptive evolution of disease alleles. *Immunity* 38:1129–1141.
- Forni D, Cagliani R, Tresoldi C, Pozzoli U, De Gioia L, Filippi G, Riva S, Menozzi G, Colleoni M, Biasin M, et al. 2014. An evolutionary analysis of antigen processing and presentation across different time-scales reveals pervasive selection. *PLoS Genet.* 10:e1004189.
- 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.
- Gao S, von der Malsburg A, Dick A, Faelber K, Schroder GF, Haller O, Kochs G, Daumke O. 2011. Structure of myxovirus resistance protein a reveals intra- and intermolecular domain interactions required for the antiviral function. *Immunity* 35:514–525.
- Goujon C, Moncorgé O, Bauby H, Doyle T, Ward CC, Schaller T, Hué S, Barclay WS, Schulz R, Malim MH. 2013. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* 502: 559–562.
- Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, Park DJ, Griesemer D, Karlsson EK, Wong SH, et al. 2013. Identifying recent adaptations in large-scale genomic data. *Cell* 152:703–713.
- Guindon S, Delsuc F, Dufayard JF, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol.* 537:113–137.
- Haller O, Kochs G. 2011. Human MxA protein: an interferon-induced dynamin-like GTPase with broad antiviral activity. *J Interferon Cytokine Res.* 31:79–87.
- Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA, et al. 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491:119–124.
- Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, Wilson SJ, Schoggins JW, Rice CM, Yamashita M, Hatzioannou T, et al. 2013. MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* 502:563–566.
- King MC, Raposo G, Lemmon MA. 2004. Inhibition of nuclear import and cell-cycle progression by mutated forms of the dynamin-like GTPase MxB. *Proc Natl Acad Sci U S A.* 101:8957–8962.
- Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, et al. 2002. A high-resolution recombination map of the human genome. *Nat Genet.* 31:241–247.
- Kosakovsky Pond SL, Frost SD. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 22:1208–1222.
- Kosakovsky Pond SL, Murrell B, Fourment M, Frost SD, Delpont W, Scheffler K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol.* 28:3033–3043.
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol.* 23:1891–1901.
- Lane J, McLaren PJ, Dorrell L, Shianna KV, Stemke A, Pelak K, Moore S, Oldenburg J, Alvarez-Roman MT, Angelillo-Scherer A, et al. 2013. A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia A. *Hum Mol Genet.* 22:1903–1910.
- Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319:1100–1104.
- Limou S, Delaneau O, van Manen D, An P, Sezgin E, Le Clerc S, Coulonges C, Troyer JL, Veldink JH, van den Berg LH, et al. 2012. Multicohort genomewide association study reveals a new signal of protection against HIV-1 acquisition. *J Infect Dis.* 205:1155–1162.
- Liu SY, Aliyari R, Chikere K, Li G, Marsden MD, Smith JK, Pernet O, Guo H, Nusbaum R, Zack JA, et al. 2013. Interferon-inducible cholesterol-

- 25-hydroxylase broadly inhibits viral entry by production of 25-hydroxycholesterol. *Immunity* 38:92–105.
- Liu SY, Sanchez DJ, Aliyari R, Lu S, Cheng G. 2012. Systematic identification of type I and type II interferon-induced antiviral factors. *Proc Natl Acad Sci U S A*. 109:4239–4244.
- Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, Cen S, Guo F, Liang C. 2013. The interferon-inducible MxB protein inhibits HIV-1 infection. *Cell Host Microbe*. 14:398–410.
- Luo M, Sainsbury J, Tuff J, Lacap PA, Yuan XY, Hirbod T, Kimani J, Wachihi C, Ramdahin S, Bielawny T, et al. 2012. A genetic polymorphism of FREM1 is associated with resistance against HIV infection in the Pumwani sex worker cohort. *J Virol*. 86:11899–11905.
- McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* 16:404–405.
- McLaren PJ, Coulonges C, Ripke S, van den Berg L, Buchbinder S, Carrington M, Cossarizza A, Dalmau J, Deeks SG, Delaneau O, et al. 2013. Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. *PLoS Pathog*. 9:e1003515.
- Mitchell PS, Patzina C, Emerman M, Haller O, Malik HS, Kochs G. 2012. Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA. *Cell Host Microbe* 12:598–604.
- Miyazawa M, Lopalco L, Mazzotta F, Lo Caputo S, Veas F, Clerici M, ESN Study Group. 2009. The “immunologic advantage” of HIV-exposed seronegative individuals. *AIDS* 23:161–175.
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 8:e1002764.
- Novembre J, Galvani AP, Slatkin M. 2005. The geographic spread of the CCR5 Delta32 HIV-resistance allele. *PLoS Biol*. 3:e339.
- OhAinle M, Kerns JA, Li MM, Malik HS, Emerman M. 2008. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* 4:249–259.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81:559–575.
- Raj T, Shulman JM, Keenan BT, Chibnik LB, Evans DA, Bennett DA, Stranger BE, De Jager PL. 2012. Alzheimer disease susceptibility loci: evidence for a protein network under natural selection. *Am J Hum Genet*. 90:720–726.
- Reichelt M, Stertz S, Krijnse-Locker J, Haller O, Kochs G. 2004. Missorting of LaCrosse virus nucleocapsid protein by the interferon-induced MxA GTPase involves smooth ER membranes. *Traffic* 5:772–784.
- Rosenberg NA. 2006. Standardized subsets of the HGDP-CEPH human genome diversity cell line panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet*. 70:841–847.
- Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 5: 725–738.
- Sasaki K, Tungrakoolsub P, Morozumi T, Uenishi H, Kawahara M, Watanabe T. 2014. A single nucleotide polymorphism of porcine MX2 gene provides antiviral activity against vesicular stomatitis virus. *Immunogenetics* 66:25–32.
- Sironi M, Biasin M, Cagliani R, Forni D, De Luca M, Saule I, Lo Caputo S, Mazzotta F, Macias J, Pineda JA, et al. 2012. A common polymorphism in TLR3 confers natural resistance to HIV-1 infection. *J Immunol*. 188:818–823.
- Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* 19:2325–2327.
- Verhelst J, Hulpiou P, Saelens X. 2013. Mx proteins: antiviral gatekeepers that restrain the uninvited. *Microbiol Mol Biol Rev*. 77: 551–566.
- Wernersson R, Pedersen AG. 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res*. 31:3537–3539.
- Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS. 2003. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res*. 31:3316–3319.
- Wilson DJ, Hernandez RD, Andolfatto P, Przeworski M. 2011. A population genetics-phylogenetics approach to inferring natural selection in coding sequences. *PLoS Genet*. 7:e1002395.
- Wright S. 1950. Genetical structure of populations. *Nature* 166: 247–249.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13:555–556.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
- Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*. 46:409–418.
- Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 22: 1107–1118.
- Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22:2472–2479.
- Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.
- Zhernakova A, Elbers CC, Ferwerda B, Romanos J, Trynka G, Dubois PC, de Kovel CG, Franke L, Oosting M, Barisani D, et al. 2010. Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. *Am J Hum Genet*. 86:970–977.