Genes and Immunity (2008) 9, 61–68

www.nature.com/gene

ORIGINAL ARTICLE

Prevalence and evolutionary origins of autoimmune susceptibility alleles in natural mouse populations

N Limaye[1], KA Belobrajdic[1], AE Wandstrat[1], F Bonhomme[2], SV Edwards[3] and EK Wakeland[1]
[1]*Department of Immunology, University of Texas Southwestern Medical Center, Dallas, TX, USA;* [2]*CNRS-UMR 5172, Universite MontpellierII, Montpellier, France and* [3]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA*

The evolutionary origin of genetic diversity in the SLAM/CD2 gene cluster, implicated in autoimmune lupus susceptibility in mice, was investigated by sequence analysis of exons from six members of the cluster in 48 wild mouse samples derived from the global mouse population. A total of 80 coding region SNPs were identified among the six genes analyzed, indicating that this gene cluster is highly polymorphic in natural mouse populations. Phylogenetic analyses of these allelic sequences revealed clustering of alleles derived from multiple Mus species and subspecies, indicating alleles at several SLAM/CD2 loci were present in ancestral Mus populations prior to speciation and have persisted as polymorphisms for more than 1 million years. Analyses of nonsynonymous/synonymous ratios using likelihood codon substitution models identified several segments in Cd229, Cd48 and Cd84 that were impacted by positive diversifying selective pressures. These findings support the interpretation that selection favoring the generation and retention of functional polymorphisms has played a role in the evolutionary origin of genetic polymorphisms that are predisposing to autoimmunity.
Genes and Immunity (2008) **9,** 61–68; doi:10.1038/sj.gene.6364446; published online 20 December 2007

Keywords: *SLAM/CD2; autoimmunity; evolution; selection; polymorphism*

Introduction

Many of the genes that regulate the immune system are extremely polymorphic, as exemplified by the genetic diversity of class I and class II genes of the major histocompatibility complex (MHC). MHC alleles differ by an excess of nonsynonymous variations, which tend to cluster in gene segments that encode peptide-binding sites.[1–4] Allelic lineages of MHC class I and class II genes also have extremely long coalescence times, often persisting in natural populations over time spans that encompass multiple speciation events.[5–7] Although the diversity found at MHC loci remains unparalleled in the mammalian genome, several other immunoregulatory gene complexes also exhibit significant levels of diversity[8,9] suggesting that selection may favor the development and retention of functional polymorphisms in immunoregulatory genes.

Variations in immunoregulation can also have deleterious effects by potentially increasing susceptibility to autoimmunity and/or autoaggressive immune responses. This is best illustrated by the strong association of MHC class II alleles with susceptibility to diabetes, rheumatoid arthritis and systemic lupus erythematosus (SLE).[10–12] Although genetic predisposition to these autoimmune diseases is multifactorial, involving interactions between multiple genes and environmental factors,[13,14] MHC polymorphisms often dominate the genetics of susceptibility. Intriguingly, many of the MHC class I and class II allelic lineages that are associated with susceptibility have been determined to have ancient evolutionary origins.[15,16]

The SLAM/CD2 gene cluster, another immunoregulatory family, contains seven adhesion/costimulatory molecules that influence a wide variety of immune system functions.[17,18] We recently identified extensive polymorphisms throughout the SLAM/CD2 gene cluster, which distinguish two stable SLAM/CD2 haplotypes in laboratory mouse strains. Interestingly, while autoimmunity was associated with the more common of these haplotypes, disease only occurred when this version of the SLAM/CD2 cluster was expressed in the context of the C57BL/6 (B6) genome.[9] This autoimmune-prone SLAM/CD2 haplotype is contained within the potent SLE susceptibility locus, *Sle1b*, and is responsible for the production of antinuclear antibodies (ANAs) in B6.*Sle1b* congenic mice.[9,19] More recently, we have demonstrated that polymorphisms of this gene cluster impair the induction of immunologic tolerance in the immature B cell compartment.[20]

Here we describe the evolutionary origins of the genetic diversity of the SLAM/CD2 gene cluster. Our findings indicate that the family is highly polymorphic in natural mouse populations and that diversifying selection has favored the generation of variability within

Correspondence: Professor EK Wakeland, Department of Immunology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, USA.
E-mail: edward.wakeland@utsouthwestern.edu
Received 5 July 2007; accepted 3 August 2007; published online 20 December 2007

the functional domains of specific family members. In addition, phylogenetic analyses revealed clustering of alleles derived from multiple *Mus* species and subspecies, indicating alleles at several SLAM/CD2 loci were present in ancestral *Mus* populations. These results support the interpretation that diversifying selection has favored the retention of functional polymorphisms in this gene cluster.

Results

The SLAM/CD2 gene cluster is highly diversified and forms haplotypes in wild mouse populations

Diversity of the SLAM/CD2 gene cluster was assessed by sequencing exons 2 and 3, which encode the extracellular ligand-binding Ig domains, from *Cd229, Cs1, Cd48, Cd150, Cd84* and *Ly108*. We first assessed the stability of the two haplotypes identified among standard laboratory mice by sequencing these alleles in 15 wild mouse-derived inbred strains with defined *Mus musculus* origins. As shown in Figure 1, two of the four *molossinus* strains share haplotype 1 (characteristic of B6), while five of the seven *musculus* strains contain the autoimmune-associated haplotype 2 (characteristic of NZM2410) (Figure 1; mouse strains listed in Supplementary Table 1). The remaining wild-derived strains share some SNPs with each, forming three additional recombinant versions of the cluster. This analysis revealed only five additional coding region SNPs to those identified in our earlier analysis of 34 standard inbred strains, indicating that this *Mus musculus* collection contains little additional SLAM/CD2 diversity from that observed in laboratory strains. The exception was *Mus pahari*, which is a distantly related *Mus* species belonging to a different subgenus (*Coelomys*). These results indicate that the two SLAM/CD2 haplotypes found in standard laboratory strains are both present in natural mouse populations, along with several recombinant versions.

This analysis was subsequently extended to 33 additional wild-derived mouse DNAs, thus giving us a panel of 48 samples representing *Mus musculus, Mus spretus, Mus macedonius, Mus spicilegus* and *Mus (C.) pahari*. The majority of the panel was *Mus musculus* samples, further divided into the following subspecies*:* 15 *Mus m musculus*, 5 *Mus m molossinus*, 5 *Mus m castaneus*, 11 *Mus m domesticus* and 4 *Mus m spp (spp* indicates that the subspecific status of populations from the center of the species' range is not yet fully settled). This panel of genomic DNAs, obtained from wild derived closed colonies and often heterozygous for SLAM/CD2 alleles, is a representative sampling of the global mouse population (detailed list of strains and geographical origins provided in Supplementary Table 2).

A 1.1 kb segment spanning nearly the entire mitochondrial control region, the most polymorphic segment of the mitochondrial genome,[21] was sequenced to provide an estimate of neutral gene genealogies of this mouse panel. Analysis of the phylogenetic trees for these mitochondrial sequences revealed the predicted clustering by subspecies and species (Figure 2), with some overlap between subspecies that are known to be parapatric and interbreed in hybridization zones on the borders of their ranges.[22] This distribution follows the accepted pattern of phylogenetic relationships of these *Mus* species obtained with mtDNA[23] and is consistent with data obtained from hybridization studies on single copy nuclear DNA and analyses of nuclear genes sequence.[22,24,25]

Having established this baseline, the sequences of exons 2 and 3 from members of the SLAM/CD2 gene cluster were analyzed. There are 80 SNPs within the exons encoding the Ig domains of six SLAM/CD2 genes in the wild mice, almost a threefold increase over the variation observed among wild-derived inbred and laboratory strains. The calculated relative recombination rate across these SNPs, defined as the number of

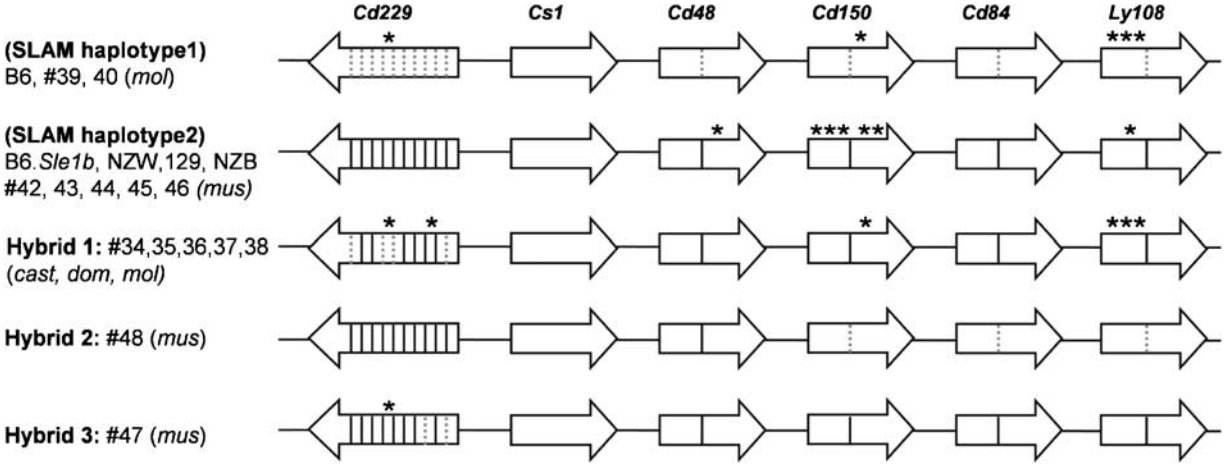

Figure 1 SLAM/CD2 haplotypes found in wild-derived inbred strains from distinct *Mus* taxa. The extracellular Ig domains of six members of the SLAM/CD2 family were sequenced in a panel of wild-derived inbred strains from five distinct *Mus musculus* subspecies. The two haplotypes originally found in common lab strains (Haplotype 1 and the autoimmune-associated Haplotype 2) are both present amongst the wild-inbred strains. Wild-derived inbred strains are designated by numbers (corresponding strain names with subspecies information can be found in Supplementary Table 1, published as supporting information on PNAS website). The subspecies that share each haplotype are also indicated (*mol*: *Mus m molossinus*; *mus*: *Mus m musculus*; *cast*: *Mus m castaneus*; *dom*: *Mus m domesticus*). The arrows indicate the order and transcriptional direction of the genes, going from centromere to telomere on mouse chr 1. The SNPs that distinguish the two major haplotypes are indicated by dashed (Haplotype 1 alleles) or solid (Haplotype 2 alleles) lines. Asterisks indicate additional SNPs found in only a subset of the strains in each group.

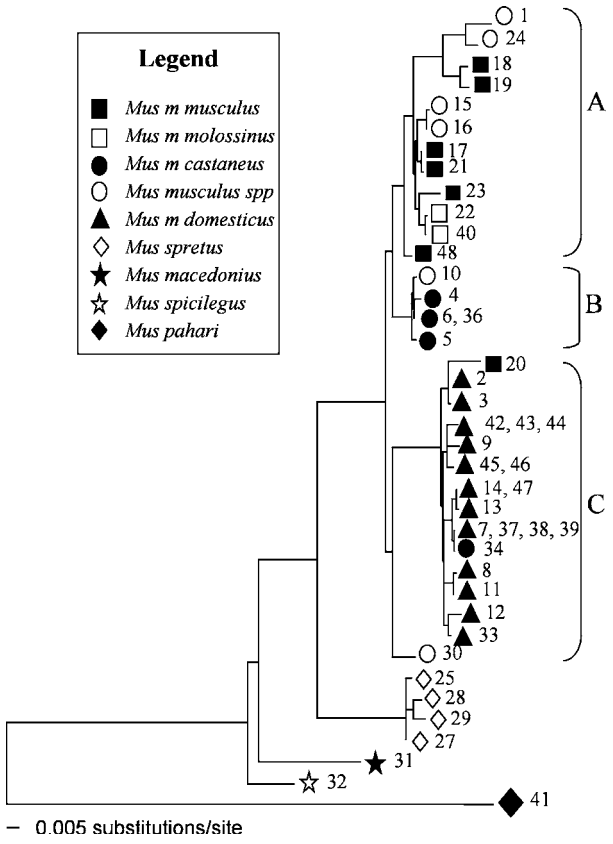

Figure 2 Phylogenetic analysis of Mitochondrial D-loop regulatory region sequences shows that mouse strains cluster by species. The tree generated by phylogenetic analysis of sequences generated from the noncoding D-loop region of mitochondria from 33 wild-derived strains, and 15 wild-derived inbred strains from different *Mus* taxa is shown. The legend lists the symbols used to denote the various *Mus* species. Along with species, the individual strains on each branch are also indicated by their assigned numbers. Corresponding strain names, along with substrain and geographic origin information can be found in Supplementary Tables 1 and 2. Horizontal branch lengths correspond to the number of substitutions per amino acid site as indicated, showing the extent of diversification from the previous node. The number of nodes separating one strain from another indicates how closely related their sequences are. Strains on the same branch share the same allele while strains separated by a single node have more closely related alleles than those separated by three nodes and so on. Thicker lines indicate bootstrap values of $>50\%$ for the clustering of strains from the node that immediately follows. Brackets A, B and C are used to highlight regions that are described in the text.

crossovers per nucleotide per generation, was found to be fairly uniform across the entire SLAM/CD2 region, with $r = 0.0008$, which is about twice the average of human autosomal genes (Supplementary Figure 1). This result indicates that no predominant SLAM/CD2 haplotypes were detectable in this sample of wild-derived strains.

The SLAM/CD2 family Ig *regions exhibit ancestral polymorphism*

In contrast with the results obtained when analyzing mitochondrial DNA, the phylogenetic trees generated from the SLAM/CD2 family genes exhibited extensive clustering between different *Mus* subspecies and species. This is illustrated with *Cd229* and *Cd48* in Figures 3a and b respectively, with the trees for the remaining SLAM/CD2 members presented in Supplementary Figures 2 and 3. The most striking feature of the *Cd229* Ig domain tree (Figure 3a) is the interspersed distribution of different *Mus* species. For example, one *Mus spretus* strain (no. 29) clusters more closely with *Mus macedonius* and *Mus spicilegus,* rather than with the rest of the *spretus* alleles (*Bracket A*). Similarly, the *spretus* cluster shares a node with a heterogenous group of *Mus musculus* subspecies (*Bracket B*). More branches of the *Cd229* phylogenetic tree carry multiple subspecies than are observed in the mitochondrial tree, including *castaneus* with *domesticus* and *musculus* with *domesticus*. Interestingly, the autoimmune Haplotype 2 allele is present in several *Mus m musculus* strains as well as *Mus m domesticus* (no. 7 and 33) while Haplotype 1 is found in *Mus m molossinus* (no. 22, 39 and 40). The occurrence of the autoimmune-associated allele in multiple subspecies suggests that it first arose in an ancestral mouse species and has persisted through several speciation events.

Similar to *Cd229*, the *Cd48* tree (Figure 3b) does not separate the major *Mus* species into distinct clades. The two alleles from the heterozygous *Mus spicilegus* strain (no. 32) are separated by three nodes, with 32B sharing a 'younger' node with the *Mus musculus* subspecies, *Mus macedonius* and *Mus spretus* (*Bracket* A). There are several instances where multiple *Mus musculus* subspecies cluster together on one branch; which is a more pronounced feature of *Cd48* than *Cd229*. The B6 Haplotype 1 allele is grouped on a single branch with *molossinus* (outbred no. 22 and wild-derived inbred no. 39 and 40), *musculus* (no. 19) and a *Mus m spp* strain (no. 15). Significantly, the Haplotype 2 allele from B6.*Sle1b* is present in multiple strains from all of the *Mus musculus* subspecies (*musculus*: no. 45, 46, 47, 48; *domesticus:* no. 2, 7, 9, 13, 14, 35; molossinus: no. 37, 38; *castaneus*: no. 4, 5, 34, 36; *Spp*: 1, 15, 30). *Mus pahari* remains a clear outlier in this distribution.

Overall, an impressive amount of ancestral polymorphism is observed in all of the trees generated from the SLAM/CD2 gene cluster (Supplementary Figures 2 and 3). The level of ancestral polymorphism observed in each of these trees far exceeds that observed in the neutral genealogy provided by the mitochondrial control analysis region (Figure 2), indicating that SLAM/CD2 polymorphisms that arose prior to evolutionary speciation have persisted in these members of the genus *Mus*, though we cannot exclude that introgression at secondary contacts or occasional interbreeding between sympatric species may have contributed to this evolutionary pattern.

The extended coalescence times for alleles in the SLAM/CD2 family suggest that balancing selection, possibly pathogen driven, has impacted the accumulation of diversity in the ligand-binding domains of these genes. Of particular interest is the fact that, in every tree, multiple subspecies cluster on the same branch as the autoimmune-associated alleles of this cluster. The persistence of these potentially deleterious alleles over evolutionary time spans may indicate that autoimmune diseases that predominantly impact older mice are not potent selective pressures in natural mouse populations, or that advantageous phenotypes mediated by these alleles may dominate over autoimmunity.

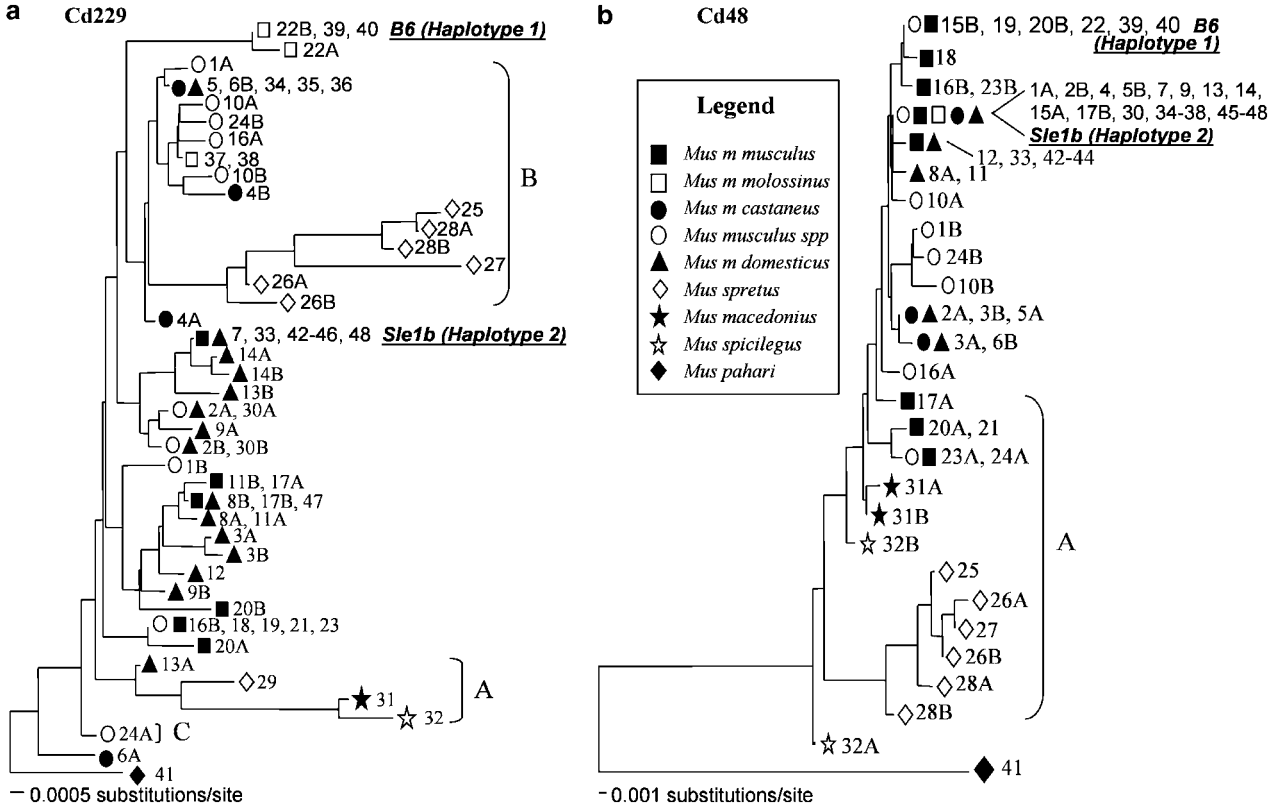

Figure 3 Phylogenetic analyses of the SLAM/CD2 regions show the presence of trans-species polymorphisms. Phylogenetic analyses were carried out on each of the SLAM/CD2 Ig region coding sequences (IgV + IgC; or IgV1 + IgC1 + IgV2 + IgC2 in the case of *Cd229*). (**a**) Shows the *Cd229* tree, and (**b**) shows the *Cd48* tree. Heterozygous strain alleles are denoted by the strain names, followed by (A or B) (for example, 22A and 22B). Representative strains from Haplotype 1 (B6) and Haplotype 2 (B6.*Sle1b*, NZW, 129 and NZB) were also included.

Table 1 d_N/d_S analysis of the SLAM/CD2 family

Locus	*No. of codons*	*Likelihood ratio test*[a]	*Models of selection*[b]				
			Neutral model (M1)		*Discrete model (M3)*		
			p0 (ω0 = 0)	*p1 (ω1 = 1)*	*p0 (ω0)*	*p1 (ω1)*	*p2 (ω2)*
Cd48	191	**6.62207**	0.69931	0.30069	0.82299 (0.00001)	0.11271 (2.78916)	0.06430 (6.75520)
Cd84	195	**7.392841**	0.42873	0.57127	0.61480 (0.00001)	0.38511 (3.94486)	0.00010 (3.94574)
Cs1	191	−0.295038	0.22489	0.77511	0.54498 (0.75884)	0.44886 (0.75886)	0.00616 (36.35146)
Ly108	198	4.053447	0.27926	0.72074	0.00000 (0.00001)	0.99488 (0.71488)	0.00512 (77.69282)
Cd150	210	0.037811	0.87935	0.12065	0.09056 (0.00001)	0.79798 (0.0001)	0.11146 (1.14935)
Cd229	395	**42.702684**	0.79409	0.20591	0.20272 (0.16628)	0.76018 (0.16630)	0.03710 (20.38612)

Codon-substitution analysis of SLAM/CD2 family members.
[a]Bold values achieve significance at $P = 0.05$ (value > 5.991, critical value for χ^2-test with two degree of freedom).
[b]See Materials and methods for a description of these models and explanation of analysis.

Polymorphic codons in Cd229, Cd48 *and* Cd84 *are impacted by diversifying selection*

The ratio of nonsynonymous (change in amino acid) to synonymous (silent) substitutions ($\omega = d_N/d_S$) is a sensitive measure of the selective pressure acting on molecular sequences. We assessed the types of selection operating on SLAM/CD2 coding regions using maximum likelihood codon-substitution models as described in Methods.[26] This approach allows for heterogeneous selection pressure at different amino acid sites, and is therefore more sensitive to selection operating on a subset of codons. Table 1 presents the results of this analysis for the SLAM/CD2 family members. The difference in likelihood between the Neutral Model (M1) and the Discrete Model (M3, which allows for positive selection), shown as the Likelihood Ratio Test value, is significant for *Cd229* and *Cd48* and *Cd84*, suggesting that polymorphisms in the Ig-domains of

these genes are subject to positive selective pressures. This analysis suggested that diversifying selective pressure might be operating on all of the other SLAM/CD2 genes (ω2 being >1 in each case); however, the number of codons involved may be too limited to reach significance.

Protein-structure alignments of the CD229 and CD84 Ig-domains with those of human and rat CD2, for which the extracellular domain structure is known, are shown in Figure 4, with the specific residues that are impacted by diversifying selection in each indicated. CD229 has eleven codons under selection, eight of which are within the IgV1 domain that aligns with the N-terminal IgV of rat-CD2. Of these, seven (including all four encoded by IgV1 region haplotype-SNPs) lie in β-strands known to be critical to ligand binding by the CD2 molecule and/or actually correspond to residues that have specifically been shown to participate in its ligand binding.[27,28] The C β-strand, which contains two such selected codons, is important in homophilic ligand binding of human CD229.[29] CD84 has eight IgV-region and four IgC-region codons under selection. Codons 27 (a haplotype distinguishing SNP) and 28 lie immediately proximal to the A β-sheet. Two other selected codons lie between β-strands, just proximal to C and C′, both important in CD2 ligand binding, while codon 73 lies within the C′ β-strand. These findings support the notion that diversifying selective mechanisms are favoring structural variations in SLAM/CD2 molecules in a manner that impacts their ligand-binding properties.

Ctla4 *polymorphism linked to autoimmunity is common in multiple* Mus *species and subspecies*

The notion that selection was favoring the retention of alleles of the SLAM/CD2 gene family that are associated with predisposition to autoimmunity raised the issue of whether this was a feature of other immunoregulatory genes with allelic associations with autoimmunity. A silent SNP at position 77 in exon 2 of *Ctla4* is associated with susceptibility to diabetes in the NOD murine model.[30] This SNP has been postulated to influence splicing of *Ctla4*, with the resistant B6 mice containing an 'A' at this position, while the susceptible NOD carries the 'G' allele. We sequenced exon 2 of *Ctla4*, along with 5′ and 3′ flanking intron sequences, in several laboratory strains and our entire panel of wild and wild-derived inbred strains. As shown in Table 2, the NOD-like allele is found in NZW, NZB and BALB/cJ. This autoimmune allele is also found in three of fifteen wild-derived inbred strains and in 31 of the 33 samples from the extended mouse global panel. In contrast to the SLAM/CD2 Ig regions, d_N/d_S analysis shows that exon 2 of *Ctla4* does not have any nonsynonymous mutations within exon 2, indicating this coding region is highly conserved in natural mouse populations. Thus, the autoimmune-associated SNP in *Ctla4* exon 2 is the most

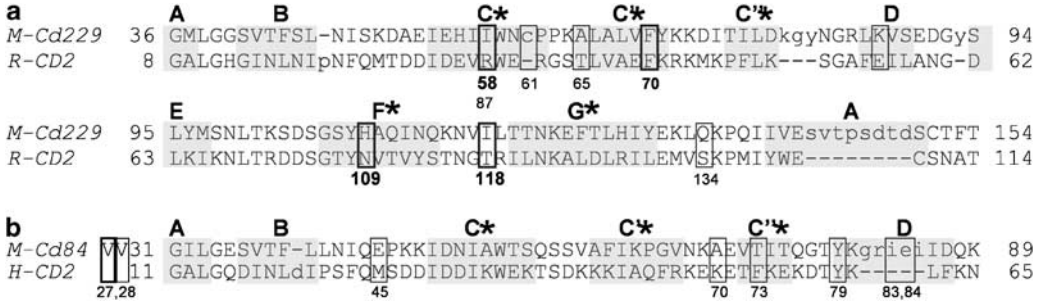

Figure 4 Protein sequence alignments of polymorphic members of the SLAM/CD2 gene cluster: (**a**) Mouse Cd229 with Rat CD2. (**b**) Mouse Cd84 with Human CD2. β-strand positions determined for CD2 are shown in shaded boxes with the strand name above. *Strands known to form the CD2-binding interface. Codons under selection in Cd229/ Cd84 are boxed, with codon-numbers below. Selected codons encoded by SNPs distinguishing the two major laboratory strain haplotypes are boxed in bold. Codon numbers in bold represent selected codons corresponding to critical CD2 ligand-binding residues.

Table 2 Distribution of *Ctla4* SNP associated with diabetes in NOD mouse

	Diabetes resistant: A at position 77	*Diabetes susceptible: G at position 77*	
Inbred lab strains	MRL/MpJ, 129/SvJ, B6, B10	NZW, NOD NZB/BlNJ, BALB/cJ	
Wild-derived inbred	*Mus m musculus* (no. 42, 43, 45, 46, 47) *Mus m domesticus* (no. 35) *Mus m castaneus* (no. 34, 36) *Mus m molossinus* (no. 37, 38, 39)	*Mus m musculus* (no. 48) *Mus m molossinus* (no. 40)	
Wild, outbred	None	*Mus m musculus* *Mus m domesticus* *Mus m castaneus* *Mus m molossinus* *Mus m Spp*	*Mus Spicilegus* *Mus Macedonius* *Mus Spretus*

Inbred lab strains, wild-derived inbred strains and wild-derived samples that share B6-like (Resistant, column 1) and NOD-like (Diabetes-susceptible, column 2) alleles of the *Ctla4* position 77 SNP are indicated. Strain names corresponding to numbers can be found in Supplementary Tables 1 and 2.

common allele in natural mouse populations, again indicating that any deleterious selective pressures mediated by the autoimmune-prone nature of this allele must be either muted, or counterbalanced effectively by other phenotypes that this allele expresses in natural mouse populations.

Discussion

Our previous analyses of the SLAM/CD2 gene cluster defined two highly divergent haplotypes among standard laboratory strains.[9] Alleles of individual genes in the SLAM/CD2 cluster differed extensively between these two haplotypes, exhibiting nonsynonymous sequence changes in functional domains, a gene expansion/contraction of *Cd224*, changes in splice isoform utilization in multiple members, and significant variations in expression levels on B and CD4 T lymphocytes.[9] The present study delineates extensive additional genetic diversification of the SLAM/CD2 family in the global wild mouse population and demonstrates that the development of SLAM/CD2 family polymorphisms has been favored by diversifying selective pressures and that allelic lineages have persisted in natural populations over long evolutionary time spans. These findings suggest that selective pressures are favoring the diversification of the functional properties of the SLAM/CD2 molecules among individuals in natural mouse populations.

Although the specific selective mechanisms responsible for the diversification of the SLAM/CD2 cluster are unknown, it is reasonable to predict that pathogen-driven selection may be at least partially responsible for these results. Studies in multiple mammalian species have defined connections between the SLAM/CD2 family and infectious disease: In humans and other mammals, members of the *morbillivirus* family directly bind and use molecules belonging to the SLAM/CD2 family as receptors for infection. In particular, the measles virus, the canine distemper virus and rinderpest virus all bind *Cd150*.[31] Targeted deletions of *Ly108* and *Cd150* have each been shown to impact infection with *Leishmania* in mice,[32,33] consistent with the notion that functional polymorphisms of these molecules may modulate susceptibility to infectious diseases.

The potential impact of polymorphisms in SLAM/CD2 family molecules on the ability and/or avidity of homotypic or heterotypic binding of family members may be a key element in the selective pressures that favor their diversification in natural mouse populations. Such variations in binding could affect immune responses in a variety of contexts including pathogen resistance, perhaps through modulating the class of downstream signaling molecules that are recruited upon receptor engagement.[34,35] Several investigations have associated changes in cytokine secretion with disruptions of genes in the SLAM/CD2 family.[32] In addition, SAP (SLAM-associated protein or SH2D1A), which acts directly downstream of this family, plays a role in Th2 regulation and cytokine production.[36,37] Mutations in SAP have been linked to dysregulations in viral immunity, including X-linked lymphoproliferative syndrome, which is caused by an inappropriate response to Epstein–Barr Virus (EBV) infection.[31] Interestingly, EBV infection has also been implicated as an environmental trigger for SLE,[38,39] making an intriguing connection between these genes, viral immunity and autoimmunity.

One of the most striking findings of the current study is that SLAM/CD2 family polymorphisms that are associated with autoimmune disease are highly prevalent in wild mice and are maintained by selective pressures. This may not be as surprising as it initially seems, for several reasons. First, although *Sle1b* is a potent locus, mediating ANA production and some degree of splenomegaly in B6.*Sle1b* congenic mice[9,19] it only has an effect on mortality when it is combined with other susceptibility factors in the genome.[40] Thus, selection against these alleles may only occur in specific deleterious *combinations* of alleles at several loci, which should occur rarely in natural populations of mice. We would hypothesize that the potential deleterious effects of autoimmune disease associated with these alleles is more than counterbalanced in natural mouse populations by the value of diversification in immune responsiveness as an evolutionary adaptation to the extensive heterogeneity and rapid evolution of microbial pathogens. Also, because wild mice probably only develop autoimmunity late in life after they have reproduced, it is unlikely that the selection against autoimmune alleles is very strong.

The presence of ancestral polymorphisms, as evidenced by phylogenetic tree analyses on the extracellular Ig regions of the SLAM/CD2 gene cluster and *Ctla4*, indicate that some form of balancing selection is operating to maintain the presence of multiple alleles that potentiate autoimmunity. Balancing selection is generally thought to result from selective pressures that vary over time or space, which is consistent with pathogen-driven mechanisms. We postulate that these autoimmune susceptibility alleles modulate immune responsiveness in ways that are differentially advantageous, possibly in the context of infectious disease, with autoimmunity representing a by-product of the selected variations. Thus, selective mechanisms that favor the diversification of immune responsiveness, coupled with epistatic interactions among immunoregulatory genes, may be a root cause of genetic predisposition to autoimmunity.

Materials and methods

Mouse strains and genomic DNA

DNA from the inbred laboratory strains and wild-derived inbred strains was obtained from strains maintained under specific pathogen free (SPF) conditions in the Animal Resources Center at the University of Texas Southwestern Medical Center at Dallas or from Jackson Laboratory (Bar Harbor, ME, USA) as described previously.[9] A detailed list of the strains of mice used, their source, subspecies and geographical distribution are listed in Supplementary Tables 1 and 2, along with the numbers assigned and used to refer to them throughout this study. Additional information regarding the wild-derived strains held in the Montpellier genetic repository is available at http://www.univ-montp2.fr/~genetix/souris.htm.

PCR and sequencing of gene products

The two extracellular Ig domains of the SLAM/CD2 family of genes were amplified using primers within the introns flanking exons 2(IgV) and 3 (IgC) of each of the following genes: *Cd229, Cs1, Cd48, Cd150, Cd84 and Ly108*. The coding region sequences from the two exons of each gene were concatenated together for the phylogenetic and d_N/d_S analyses. In the case of *Ctla4*, exon 2 along with 325 base pairs of flanking intron sequence was amplified. As a control, 1.1 kilobases of the mitochondrial regulatory D-loop region were amplified in two overlapping products of approximately 700 base pairs and 550 base pairs. For all gene products, 100 µl PCR reactions were performed on 250–500 ng genomic DNA. Primers and cycling conditions used for each gene and the resulting product sizes are listed in Supplementary Table 3. Amplified products were purified using the High Pure PCR Product Purification Kit (Roche, Indianapolis, IN, USA) or the QIAquick 96 PCR Purification Kit (Qiagen, Valencia, CA, USA). Products were sequenced on the Beckman CEQ2000XL capillary sequencer (Beckman Coulter, Fullerton, CA, USA) or the ABI 3730XL (Applied Biosystems, Foster City, CA, USA), according to manufacturer instructions. The resulting sequence was edited and assembled using Sequencher software (Gene Codes Corporation, Ann Arbor, MI, USA).

Phylogenetic analysis of sequence data

Phylogenetic trees were made with a Bayesian analysis using MrBayes version 3.03b.[41] We used default priors and ran the analysis for 2 million cycles, employing a HKY85 model of nucleotide substitution with gamma distributed rates among sites. We also used the neighbor-joining method, with a Tamura-Nei distance method and a gamma-rate parameter of 0.5.

Likelihood analysis of selection

d_N/d_S analysis was carried out using codon substitution models,[26] a phylogeny-based likelihood ratio test that tests for positive selection among different codon positions in a gene. The log-likelihood value of a Neutral Model (Model 1), in which a proportion $p0$ of codons is conserved (those with $\omega 0 = d_N/d_S = 0$), and a proportion p1 is neutral ($\omega 1 = d_N/d_S = 1$), is compared to that of a Discrete Model (Model 3) in which d_N/d_S ratios are estimated for three classes of codons ($\omega 0$, $\omega 1$ and $\omega 2$) with corresponding proportions $p0$, $p1$ and $p2$. When $\omega 2 > 1$, this indicates that the codons in this class are under positive selection. If the difference in likelihood between Model 3 and Model 1 is greater than the critical value for a χ^2 test with two degrees of freedom (5.991), then model 1 can be rejected, thus providing evidence for selection.

Acknowledgements

Support for these studies was provided by grants from the National Institute of Allergy and Infectious Disease to EKW. NL and KB were supported by NIH training grant no. NIH T32 AI 005284. FB was supported by ISEM 2007-084.

References

1 Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 1988; **335**: 167–170.
2 Hughes AL, Nei M. Nucleotide substitution at class II MHC loci: evidence for overdominant selection. *Proc Natl Acad Sci USA* 1989; **86**: 958–962.
3 She JX, Boehme SA, Wang TW, Bonhomme F, Wakeland EK. Amplification of major histocompatibility complex class II gene diversity by intra-exonic recombination. *Proc Natl Acad Sci USA* 1991; **88**: 453–457.
4 Klein J, Satta Y, O'HUigin C, Takahata N. The molecular descent of the major histocompatibility complex. *Annu Rev Immunol* 1993; **11**: 269.
5 McConnell TJ, Talbot WS, McIndoe RA, Wakeland EK. The origin of MHC class II gene polymorphism within the genus Mus. *Nature* 1988; **332**: 651–654.
6 Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P. HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* 1988; **335**: 268–271.
7 Figueroa F, Gunther E, Klein J. MHC polymorphism predating speciation. *Nature* 1988; **335**: 265–267.
8 Parham P. MHC class I molecules and kirs in human history, health and survival. *Nat Rev Immunol* 2005; **5**: 201–214.
9 Wandstrat AE, Nguyen C, Limaye N, Chan AY, Subramanian S, Tian XH *et al*. Association of extensive polymorphisms in the SLAM/CD2 gene cluster with murine lupus. *Immunity* 2004; **21**: 769–780.
10 Graham RR, Ortmann WA, Langefeld CD, Jawaheer D, Selby SA, Rodine PR *et al*. Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am J Hum Genet* 2002; **71**: 543–553.
11 Todd JA, Bell JI, McDevitt HO. HLA-DQ gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 1987; **329**: 599–604.
12 Nepom GT, Kwok WW. Molecular basis for HLA-DQ associations with IDDM. *Diabetes* 1998; **47**: 1177–1184.
13 Raman K, Mohan C. Genetic underpinnings of autoimmunity—lessons from studies in arthritis, diabetes, lupus and multiple sclerosis. *Curr Opin Immunol* 2003; **15**: 651–659.
14 Wandstrat A, Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nat Immunol* 2001; **2**: 802–809.
15 Lundberg AS, McDevitt HO. Evolution of major histocompatibility complex class II allelic diversity: direct descent in mice and humans. *Proc Natl Acad Sci USA* 1992; **89**: 6545–6549.
16 Hughes AL, Hughes MK, Howell CY, Nei M. Natural selection at the class II major histocompatibility complex loci of mammals. *Philos Trans R Soc Lond B Biol Sci* 1994; **346**: 359–366.
17 Veillette A, Latour S. The SLAM family of immune-cell receptors. *Curr Opin Immunol* 2003; **15**: 277–285.
18 Engel P, Eck MJ, Terhorst C. The SAP and SLAM families in immune responses and X-linked lymphoproliferative disease. *Nat Rev Immunol* 2003; **3**: 813–821.
19 Morel L, Blenman KR, Croker BP, Wakeland EK. The major murine systemic lupus erythematosus susceptibility locus, Sle1, is a cluster of functionally related genes. *Proc Natl Acad Sci USA* 2001; **98**: 1787–1792.
20 Kumar KR, Li L, Yan M, Bhaskarabhatla M, Mobley AB, Nguyen C *et al*. Regulation of B cell tolerance by the lupus susceptibility gene Ly108. *Science* 2006; **312**: 1665–1669.
21 Nachman MW, Boyer SN, Searle JB, Aquadro CF. Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. *Genetics* 1994; **136**: 1105–1120.
22 Guenet JL, Bonhomme F. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet* 2003; **19**: 24–31.
23 Boissinot S, Boursot P. Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the

house mouse: selection on the Y chromosome? *Genetics* 1997; **146**: 1019–1034.
24 Jansa SA, Weksler M. Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. *Mol Phylogenet Evol* 2004; **31**: 256–276.
25 Lundrigan BL, Jansa SA, Tucker PK. Phylogenetic relationships in the genus mus, based on paternally, maternally, and biparentally inherited characters. *Syst Biol* 2002; **51**: 410–431.
26 Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000; **155**: 431–449.
27 Somoza C, Driscoll PC, Cyster JG, Williams AF. Mutational analysis of the CD2/CD58 interaction: the binding site for CD58 lies on one face of the first domain of human CD2. *J Exp Med* 1993; **178**: 549–558.
28 Jining L, Makagiansar I, Yusuf-Makagiansar H, Chow VT, Siahaan TJ, Jois SD. Design, structure and biological activity of beta-turn peptides of CD2 protein for inhibition of T-cell adhesion. *Eur J Biochem* 2004; **271**: 2873–2886.
29 Romero X, Zapater N, Calvo M, Kalko SG, de la Fuente MA, Tovar V *et al*. CD229 (Ly9) lymphocyte cell surface receptor interacts homophilically through its N-terminal domain and relocalizes to the immunological synapse. *J Immunol* 2005; **174**: 7033–7042.
30 Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G *et al*. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 2003; **423**: 506–511.
31 Sidorenko SP, Clark EA. The dual-function CD150 receptor subfamily: the viral attraction. *Nat Immunol* 2003; **4**: 19–24.
32 Howie D, Laroux FS, Morra M, Satoskar AR, Rosas LE, Faubion WA *et al*. Cutting edge: the SLAM family receptor Ly108 controls T cell and neutrophil functions. *J Immunol* 2005; **174**: 5931–5935.
33 Wang N, Satoskar A, Faubion W, Howie D, Okamoto S, Feske S *et al*. The cell surface receptor SLAM controls T cell and macrophage functions. *J Exp Med* 2004; **199**: 1255–1264.
34 Shlapatska LM, Mikhalap SV, Berdova AG, Zelensky OM, Yun TJ, Nichols KE *et al*. CD150 association with either the SH2-containing inositol phosphatase or the SH2-containing protein tyrosine phosphatase is regulated by the adaptor protein SH2D1A. *J Immunol* 2001; **166**: 5480–5487.
35 Veillette A. SLAM family receptors regulate immunity with and without SAP-related adaptors. *J Exp Med* 2004; **199**: 1175–1178.
36 Davidson D, Shi X, Zhang S, Wang H, Nemer M, Ono N *et al*. Genetic evidence linking SAP, the X-linked lymphoproliferative gene product, to Src-related kinase FynT in T(H)2 cytokine regulation. *Immunity* 2004; **21**: 707–717.
37 Cannons JL, Yu LJ, Hill B, Mijares LA, Dombroski D, Nichols KE *et al*. SAP regulates T(H)2 differentiation and PKC-theta-mediated activation of NF-kappaB1. *Immunity* 2004; **21**: 693–706.
38 Moon UY, Park SJ, Oh ST, Kim WU, Park SH, Lee SH *et al*. Patients with systemic lupus erythematosus have abnormally elevated Epstein–Barr virus load in blood. *Arthritis Res Ther* 2004; **6**: R295–R302.
39 Kang I, Quan T, Nolasco H, Park SH, Hong MS, Crouch J *et al*. Defective control of latent Epstein–Barr virus infection in systemic lupus erythematosus. *J Immunol* 2004; **172**: 1287–1294.
40 Croker BP, Gilkeson G, Morel L. Genetic interactions between susceptibility loci reveal epistatic pathogenic networks in murine lupus. *Genes Immun* 2003; **4**: 575–585.
41 Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 2001; **294**: 2310–2314.

Supplementary Information accompanies the paper on Genes and Immunity website (http://www.nature.com/gene)