The Evolution of the Anti-viral OAS1 Gene in Mice and Primates

by

William G. Ferguson

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy,

The City University of New York

2008

This manuscript has been read and accepted for the Graduate Faculty in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date	Chair of Examining Committee Dr. Stephane Boissinot
Date	Executive Officer Dr. Laurel Eckhardt

Dr. Zahra Zakeri, Queens College

Dr. Cathy Savage-Dunn, Queens College

Dr. Adolfo Garcia-Sastre, Mt. Sinai School of Medicine

Dr. Weigang Qiu, Hunter College

Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

The Evolution of the Anti-Viral OAS1 Gene in Mice and Primates

by

William Ferguson

Advisor: Dr. Stephane Boissinot

The interaction between host defense genes and virulent pathogens is of great concern for both medical and evolutionary research. I analyzed the variation in the oligoadenylate synthetase (OAS) gene family in both mice and primates. The OAS family of genes is part of the innate cellular response to viral infection and is found in all mammals, as well as some lower invertebrates. In both populations of the house mouse (Mus musculus) and chimpanzees (Pan troglodytes) I identified divergent alleles of the OAS family evolving under balancing selection. I analyzed three different mouse Oas1 genes and found different levels of variation for each gene within populations of M. musculus. Using molecular clocks I identified the age of alleles for Oas1b (> 3 million years old), Oas1g (between 1.5 and 2.8 million years old), and Oas1e (< 0.5 million years old). In chimp populations we identified two divergent alleles of Oas1 which were calculated to be over 9 million years old. Further functional analysis of the chimp alleles determined that the proteins had varying abilities to bind to viral RNA. This identifies viral RNA as the selective pressure within chimp populations and is a likely candidate for retaining alleles in mouse populations as well.

ACKNOWLEDGEMENTS

I would like to thank my mentor Dr. Stephane Boissinot for his support, advice, and all the opportunities he provided to me for the last five years. I could have not asked for a more helpful and compassionate advisor. In addition to the entire faculty in the Queens College biology program, I want to extend my thanks to Dr. Zahra Zakeri and Dr. Cathy Savage-Dunn, who helped guide me through this academic endeavor. This guidance was also enhanced by Dr. Adolfo Garcia-Sastre and Dr. Weigang Qiu who provided me with excellent comments and advice on the progress of my research projects.

I would also like to thank three of my former mentors who encouraged me to entire a doctoral program – Dr. Erich Grotewold, Dr. Dave Kaplan, and Dr. Scott Campbell.

Table of Contents

List of Tables

List of Figures

Introduction

Overview

A number of parameters can affect the outcome of viral infection, including age, sex, pre-existing infection and genetic susceptibility. It has been shown for a number of diseases that some individuals possess an innate resistance to viral infections. One of the best understood cases is the resistance to HIV conferred by a variant of the *Ccr5* gene (Samson et al., 1996). Studying the mechanism and evolution of innate resistance is of considerable interest because it can lead to a better understanding of viral replication, infectivity, and possibly lead to new therapeutic strategies. One of the gene families that affects innate resistance is the oligoadenylate synthetase family (Part I of introduction), which protects against a number of viruses, including various strains of flavivirus (Part II of introduction).

Recent studies have determined that a single gene, the *Oas1b* gene, confers resistance to flavivirus (West Nile virus, Yellow Fever virus, and Dengue Fever virus) infection in the house mouse. The *Oas1b* gene constitutes one of the best model available to study the evolution of innate resistance for at least six reasons: (1) the resistance conferred by the *Oas1b* locus is flavivirus-specific; (2) Resistance is conferred by a single locus, making it easier to study; (3) there is a direct correlation between genotype and phenotype at the resistance locus; (4) the molecular basis of the resistance / susceptibility polymorphism is clearly established; (5) the genetic diversity of the house mouse is well characterized; (6) mice have been used extensively as a model in flaviviral research. My first goal was to assess the diversity of the *Oas1b* gene in the house mouse and to study the evolutionary history of this gene in the genus *Mus*.

This was accomplished by sequencing the *Oas1b* gene in a large number of wild derived strains of the house mouse. This analysis constitutes the first chapter of this thesis and has been published in *Molecular Biology and Evolution*.

However, the close proximity of duplicate *Oas1* genes surrounding *Oas1b* complicates the evolutionary analysis of this gene, because selection acting at a locus can affect the evolution of neighboring genes. Thus, my second goal was to extend the evolutionary analysis to other members of the mouse *Oas1* family. I compared the evolution of *Oas1b* with *Oas1g* and *Oas1e*. This work constituted the second chapter of this thesis and has been submitted for publication in the *Journal of Molecular Evolution*.

In order to determine whether the evolution of *Oas1* is unique in mice, we decided to analyze the *Oas1* gene in primates. Primates have a single copy of OAS1, which has been shown to be active in humans during infection with flaviviruses and other RNA based viruses, such as the human immunodeficiency virus (HIV-1), hepatitis C, influenza, ebola, and simian virus 40 (SV-40). I examined the evolution of OAS1 in hominoid primates with a focus on chimpanzees. This evolutionary data was combined with functional analysis of OAS1 in chimpanzees, which is presented in the third chapter of this thesis. The work is under revision from *PLOS genetics.*

Part I. The *Oas* gene family

Upon infection with viruses, mammalian hosts respond by releasing interferon-alpha and –beta. The release of interferon triggers the upregulation and subsequent activation of a plethora of innate host defense genes. The *Oas* gene family has been

shown to help protect primates and rodents against infection with various viruses and is one of the major players in anti-viral defense.

Evolution

Oas genes can be found in humans, horses, pigs, cows, rodents and marine sponges (Justesen J et al., 2000; Kumar S et al., 2000; Sarkar and Sen, 1998; Sarkar SN, 1998; Wiens M et al., 1999) indicating that this is a very ancient gene family. There are four human OAS genes, denoted OAS1, OAS2, OAS3 and OASL (*Oas-Like*). OAS1 appears to be the ancestral gene which underwent a segmental duplication resulting in OAS2, which contains two duplicate OAS1 domains (Kumar S et al., 2000). OAS2 then duplicated forming OAS3, which underwent an internal domain duplication resulting in a three domain gene (Kumar S et al., 2000). OAS1 has four established isoforms (p42/p46/p48/p52) (Figure 1), OAS2 has two isoforms (p69/p71), and OAS3 has one (p100). The fourth locus, OASL, encodes a 59 kDa protein (p59) which is enzymatically inactive. The four isoforms of OAS1 are identical over their first 346 amino acids and differ in the carboxy-termini.

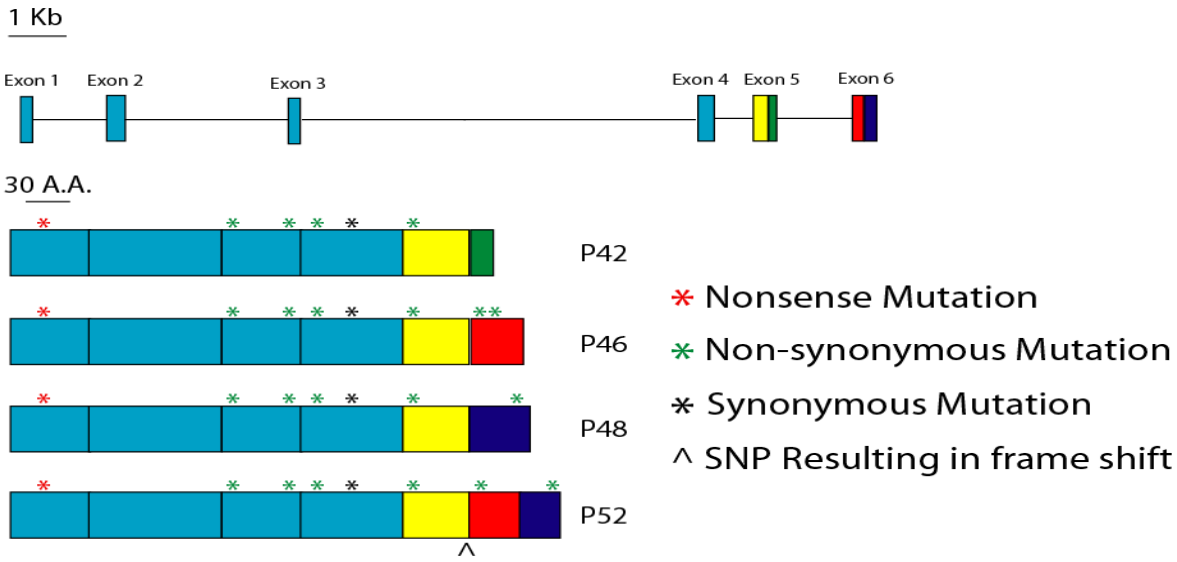

Figure 1. The human *Oas1* gene as it appears on chromosome 12. Below are the four established isoforms as well as identified SNP's within human populations.

Mice have homologs of the human OAS1, OAS2, and OAS3 genes (Hovanessian, 1991; Hovanessian AG, 1991; Rebouillat D and Hovanessian AG, 1999). The main difference between human and mouse *Oas1* is that in mice *Oas1* underwent multiple rounds of duplication resulting in eight *Oas1* genes, *Oas1a-h* (Eskildsen S et al., 2002; Kumar S et al., 2000). All of the *Oas1* genes lie adjacent to each other on chromosome 5 in the mouse genome (Figure 2). The fact that multiple copies have been retained, suggests that it might be advantageous for mice to have several *Oas1* genes (Mashimo et al., 2003). Although the sequences of the eight Oas1 genes are very similar, the promoter regions are divergent (Mashimo et al., 2003). Differences in promoter regions of the *Oas* family indicates that the genes are differentially regulated and this has been demonstrated experimentally (Mashimo et al., 2003).

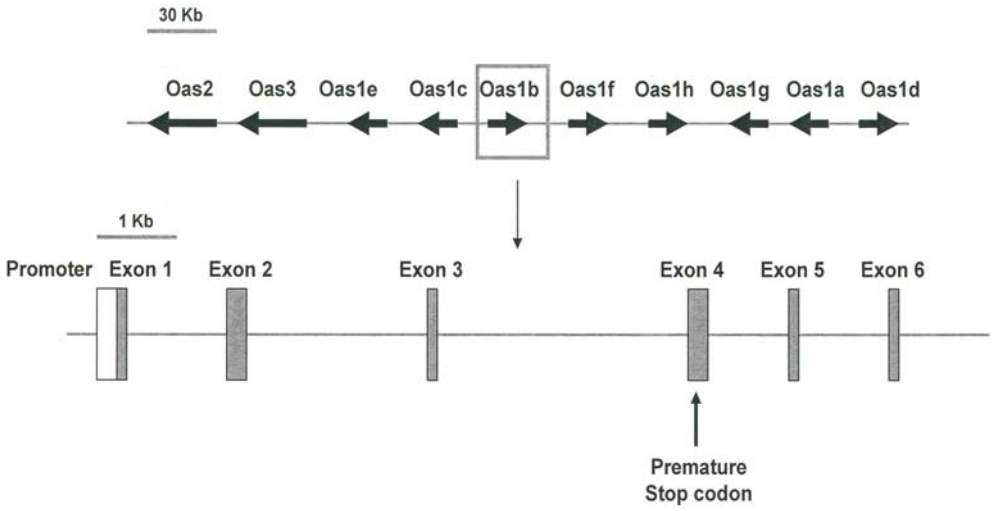

Figure 2. Arrangement and direction of *Oas* genes on chromosome 5 of the mouse. The lower portion of the figure shows the promoter, exons, and introns of *Oas1b* and the position of the nonsense mutation found in the susceptible allele.

Function of OAS proteins

The 2'-5'-oligoadenylate synthetase pathway has been shown to impede viral infections, to play a role in antiproliferation of cancer cells, and to participate in apoptosis (Desai SY et al., 1995; Ghosh SK et al., 1991; Hassel B et al., 1993; Hovanessian AG, 1991; Kerr IM and Brown RE, 1978). The *Oas1* gene is upregulated by interferon-alpha and its product participates in an anti-viral pathway that targets viral infections. OAS1 proteins require tetramerization and the presence of double stranded RNA for activation (Brinton MA and Perelygin AA, 2003; Desai SY et al., 1995; Ghosh A et al., 1997b; Hovanessian AG, 1991). The activated OAS1 protein synthesizes 2'-5' oligoadenylate acid (2-5A) from ATP molecules (Ghosh SK et al., 1991). Oligoadenylates cause Rnase L to dimerize thereby activating it (Figure 3) (Hassel B et al., 1993). Rnase L then degrades viral and cellular RNA, thus stopping the infection (Zhou et al., 1997). Human OAS1 is activated only after binding to non-specific segments of double-stranded RNA (Desai SY et al., 1995; Hovanessian AG, 1991). However different aptamers of single-stranded RNA, with different secondary structure, significantly affect the enzymatic activity of OAS proteins (Hartmann R et al., 1998). The length of the double stranded RNA (dsRNA) also affects OAS activity, with longer RNA molecules increasing the enzymatic activity (Hartmann R et al., 1998; Maitra RK et al., 1994; Sarkar SN, 1999).

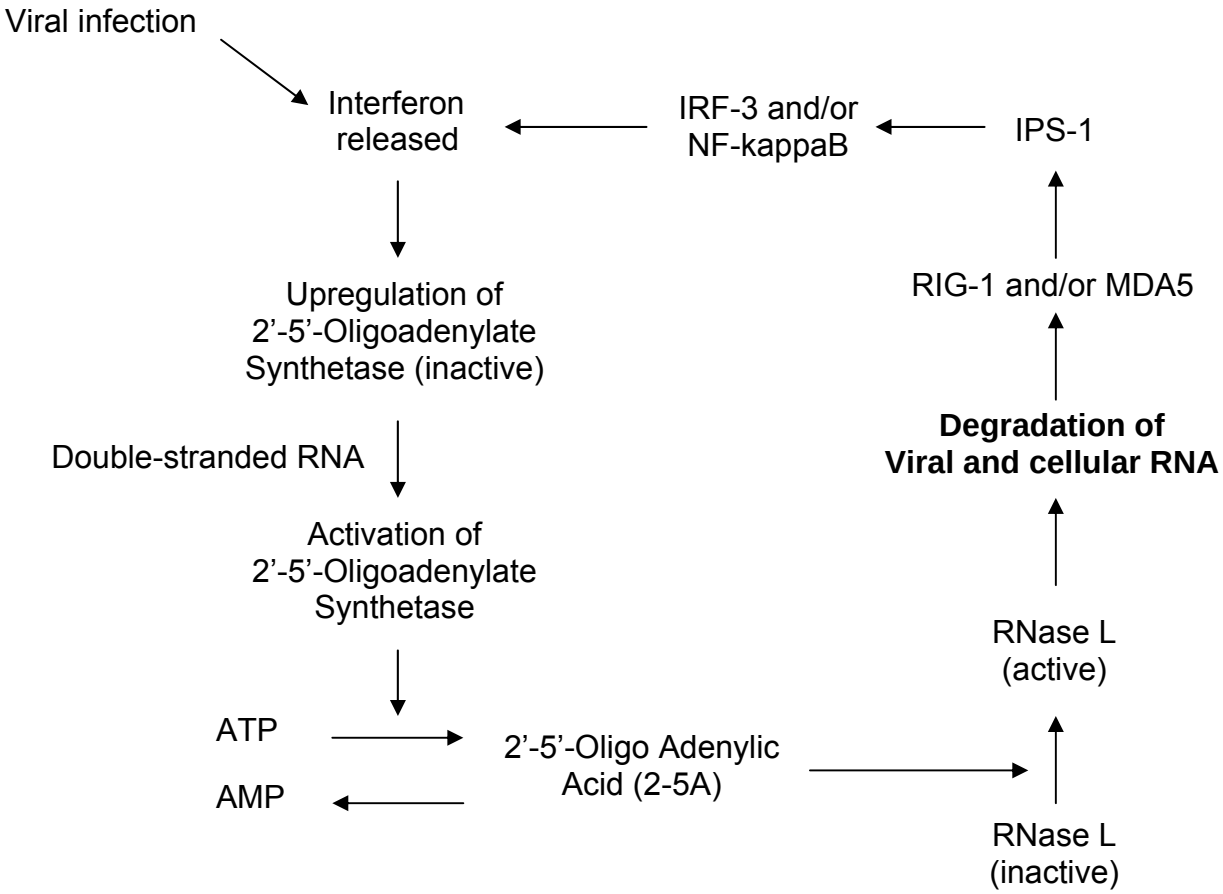

Figure 3: The OAS pathway. This figure details the OAS pathway from interferon induction to the final activation of Rnase L and subsequent RNA degradation. This figure also shows the possibility of OAS leading to apoptosis in the cell (Urosevic, 2003).

The function of Oas1 in the house mouse

Of the eight copies of *Oas1* in the mouse genome, the *Oas1b* gene has been the most studied, since it was found to specifically inhibit flavivirus infection (Mashimo et al., 2002; Perelygin AA et al., 2002). The cDNA of *Oas1b* is 1131 basepairs in length with six exons and translates into 376 amino acids (Mashimo et al., 2002; Perelygin AA et al., 2002). There is a susceptible allele of *Oas1b* which was found to have a nonsense mutation in exon 4, resulting in a truncated protein (Mashimo et al., 2002; Perelygin AA et al., 2002).

Oas1a and *Oas1g* have the ability to produce 2-5A's and are broadly expressed in many tissue types (Kakuta S et al., 2002; Mashimo et al., 2003; Mashimo et al., 2002). *Oas1b* has been shown to specifically prevent infection against flaviviruses, such as dengue fever virus, yellow fever virus and West Nile virus (Kakuta S et al., 2002; Mashimo et al., 2003; Mashimo et al., 2002). Yet this protein is enzymatically inactive and operates in a Rnase-L independent pathway which has yet to be characterized (Scherbik SV et al., 2006). It is unknown whether *Oas1b* interacts with other members of the *Oas1* family, however both *Oas1b* and *Oas1g* are highly upregulated during infection with West Nile virus (Scherbik SV et al., 2007)

Oas1e and *Oas1d* are very similar at the amino acid level (~84%) and are expressed during embryonic development as well as in some germ cells (Yan et al., 2005). Recent work has found that *Oas1d* can inhibit the function of *Oas1a in vivo* by blocking 2-5A production (Yan et al., 2005). The same study found that mice with *Oas1d* knocked-out showed reduced fertility, with defects in ovulation and preimplantation. It is hypothesized that both *Oas1d* and *Oas1e* may play a role in the

survival of embryos during viral infection. In this scenario both *Oas1d* and *Oas1e* may work against either the activation of Rnase L or another unknown pathway. The functions of *Oas1c, Oas1f,* and *Oas1h* have yet to be clarified.

The function of OAS1 i*n primates*

Interactions between OAS1 and various strains of viruses have been extensively analyzed for humans, chimpanzees, and other primates. Table 1 compiles data from eight infectious viruses which have been correlated with OAS1. All of the viruses listed are able to infect humans; however non-human primate infections vary (columns 5 through 7). All of the viruses, other than SARS, have been shown to upregulate OAS1 and three of the viruses have been found to directly interact with OAS1. The role of human OAS1 in the host-defense system has been used to develop anti-viral therapies. Therapeutic treatments with interferon have been used to trigger the OAS1 pathway, leading to greater viral protection. Interferon treatment has been shown to limit infection of all eight viruses in humans and in some non-human primates. Some viruses have been able to avoid the potent action of OAS1 through either RNA or protein interactions. This is clear during HIV-1 infection, where the *tat* protein from HIV-1 is able to inhibit the function of human OAS1 by competing for the TAR RNA (Maitra et al., 1994).

Differences in response to interferon treatments led to genetic analyses which have shown that indeed the human OAS1 gene sequence varied between individuals and that some of the differences lead to functional variants of OAS1. For instance, an association between a Single Nucleotide Polymorphism (SNP) in OAS1 and susceptibility to hepatitis C infection has been documented (Knapp S et al., 2003). A

more recent study found that a SNP affecting the structure of the OAS1 protein, affected individuals' susceptibility to SARS (Hamano et al., 2005). Another study found that a SNP in a regulatory region of OAS1 was responsible for variable levels of enzymatic activity (Bonnevie-Nielsen et al., 2005). These functional differences may result from the nature of the interactions between the OAS protein and the viral pathogens. However, the specificity of the interaction between OAS1 and viruses remains unclear.

Members of the OAS gene family have also been detected in other non-human primates. The OAS1, OAS2, and OAS3 loci can be found in the chimpanzee genome on chromosome 12 and in rhesus monkeys on chromosome 11 (UCSC website). The UCSC website also detected 2 copies of OASL on chimp chromosomes 12 and 2b, however no discernable OASL loci were found in the rhesus genome. Studies concerning non-human primate OAS genes have been limited, however recent work demonstrated that OAS1 is expressed during viral infection in chimps (Lindenbach et al., 2006).

Disease	Type of virus	Infection Rates in Humans	Death Rate in Humans	Found in wild-born chimps	Occurs in other primates	Prevelance in *P. t. verus*	Leads to expression of OAS	RNA or protein interation with OAS	Response to treatment with interferon
HIV-1	Retroviral	3.6 to 6.6 million cases/year worldwide	3.1 million deaths/year worldwide	No	Yes (humans	None	Yes (8)	Yes - Tar RNA (14)	Yes (8)
Influenza	Negative strand RNA virus	3 to 5 million cases/year worldwide	250,000 - 500,000 deaths/year worldwide	Unknown	Yes (humans)	Unknown	Yes (25)	Unknown	Yes (26)
SARS	Positive strand RNA virus	Unknown	Unknown	No	Yes (rhesus) (6)	None	Unknown	Unknown	Yes (17)
Dengue Fever	Positive strand RNA virus	50 to 100 million cases/year worldwide	21,000 deaths/year worldwide	Unknown - primate host (22)	Yes (humans)	Unknown	Yes (23)	Unknown	Yes (24)
Yellow Fever	Positive strand RNA virus	200,000 cases/year worldwide	30,000 deaths/year worldwide	Yes (*verus*) (1)	Yes (humans, gorilla) (1)	Unknown	Correlation between YFV vaccine and OAS activation (9)	Unknown	Yes (9)
Hepatitis B	Retroviral	4 to 6 million cases/year worldwide	500,000 to 1.2 million deaths/year worldwide	Yes (*verus, schweinfurthii, troglodytes, vellerosus*) (5)	Yes (human, orangutans, gorillas, and gibbons) (5)	Very Common in West Africa (5)	Yes (in chimps and human) (11)	Unknown	Yes (20)
Hepatitis C	Positive strand RNA virus	3 to 4 million cases/year worldwide	52,000 deaths/year worldwide	Unknown	Yes (humans) (7)	Unknown	Yes (in chimps and human) (12)	Yes - NS5 protein (15)	Yes (12)
Ebola	Negative strand RNA virus	Unknown	Unknown	Yes (*verus, troglodytes*) (1)	Yes (human, gorilla, rhesus) (1)	Outbreaks in verus from 1994-1996 (1)	Inhibits interferon expression (13)	Competes for dsRNA with other antiviral proteins (16)	Yes (21)

Table 1. A list of viruses that have primate hosts and their relation to OAS or interferon. 1.(Formenty et al., 1999), 2. (Santiago et al., 2003) and (Keele et al., 2006), 3. (Leendertz et al., 2004) and (Nerrienet et al., 2004), 4. (Switzer et al., 2005) and (Calattini et al., 2006), 5. (MacDonald et al., 2000) and (Starkman et al., 2003), 6. (Rowe et al., 2004), 7. (Simmonds, 2001), 8. (Kim et al., 2004), 9. (Bonnevie-Nielsen et al., 1989), 10. (Abel et al., 2002), 11. (Wieland et al., 2004), 12. (Lindenbach et al., 2006) and (Takeda et al., 1993), 13. (Jahrling et al., 1999) and (Basler et al., 2000), 14. (Maitra et al., 1994), 15. (Taguchi T et al., 2004), 16. (Cardenas et al., 2006) and (Feng et al., 2007), 17. (Enserink, 2004), 18. (Abel et al., 2002) and (Taylor et al., 1998), 19. (Rhodes-Feuillette et al., 1990) and (Falcone et al., 1999), 20. (Waked et al., 1988), 21. (Jahrling et al., 1999), 22. (Gubler, 1998b), 23. (Warke et al., 2003), 24. (Ajariyakhajorn et al., 2005), 25. (Kim et al., 2002), 26. (Gamian et al., 1991).

Structure of the OAS1 protein

The OAS1 protein therefore has two rather distinct domains that appear to have separate functions. Multiple protein motifs have been determined for the OAS1 protein (Figure 4). The helix-loop-helix linker region connects the C-terminus and N-terminus (Hartmann R et al., 2003). The alpha-beta-beta-beta-alpha region is commonly found in RNA binding molecules such as PKR, Staufen, and Rnase III (Carlson CB et al., 2003; Ryter JM and Schultz SC, 1998; Spanggord and Beal, 2000; Spanggord RJ, 2000; Stefl R et al., 2005). The first alpha helix of this motif is the least conserved secondary structure element, although it is crucial for shape recognition of dsRNA (Stefl R et al., 2005). Further evaluation of the secondary structure showed an alpha-beta-beta-alpha-beta-beta-beta domain. This motif is found in DNA polymerase beta and encompasses the catalytic domain of the protein (Sarkar SN et al., 1999). This structure similarity led to OAS being classified as a member of the Class I nucleotidyl transferase superfamily (Yue D et al., 1996). This superfamily includes molecules such as DNA polymerase β (polβ), terminal deoxynucleotidyl transferase, and poly-(A) polymerase. The main differences in function is that OAS is template-independent, can use mononucleotides as primers, and requires dsRNA for activation (Sarkar SN et al., 1999).

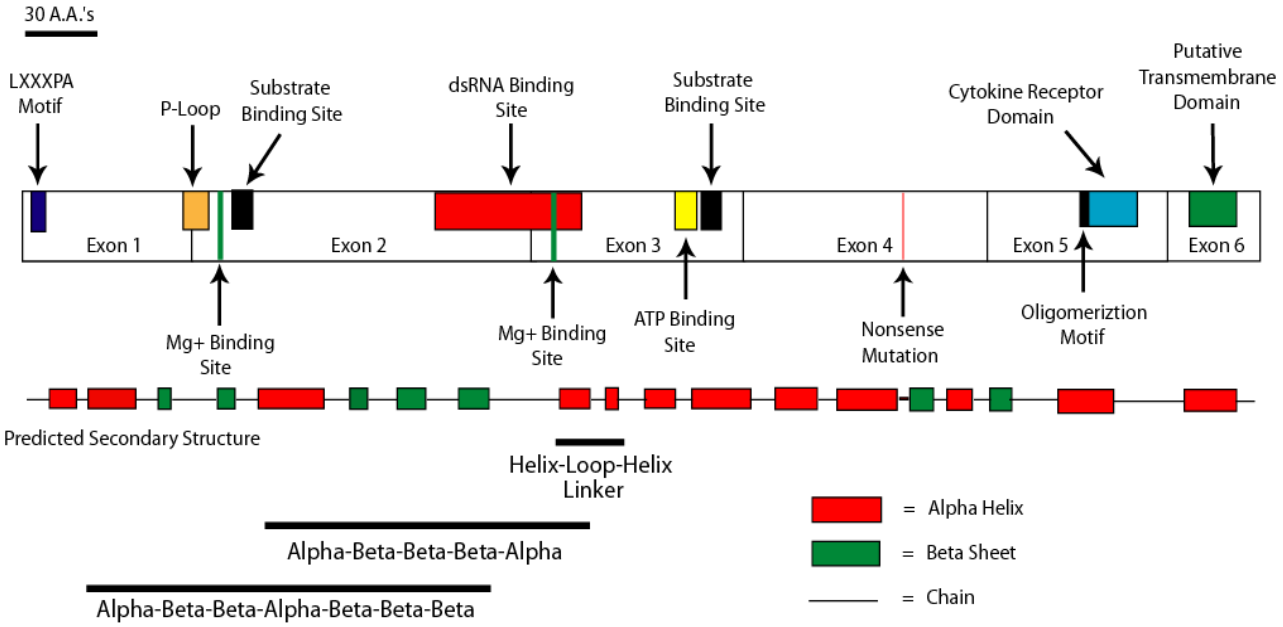

Figure 4. The upper portion of the diagram points out all known domains of OAS1, as well as the putative domains of OAS1b. Different colored boxes are used to separate the individual domains from each other. The predicted secondary structure is shown directly below the protein. The red blocks are predicted alpha helixes, the green boxes are beta sheets, and the lines between them are undetermined amino acid chains. Common secondary motifs are pointed out on the bottom. The black lines indicate the lengths of the motifs and where the motifs are located according to the predicted secondary structure. The entire diagram is drawn to scale based on the amino acid length.

A number of functional domains of human OAS1 protein have been determined experimentally. Three aspartic acids found at amino acid sites 71, 73, and 144 bind to either magnesium or manganese, which act as a catalyst (Hartmann R et al., 2001; Marie I, 1999; Marie et al., 1999; Sarkar SN et al., 1999). The LxxxPA motif is found at the N-terminus and is required for enzymatic activity in OAS1, though not required for dsRNA binding (Ghosh A et al., 1997a). Deletion experiments have shown that a region of amino acid residues between 104 and 158 in OAS1 was necessary for dsRNA binding (Ghosh SK et al., 1991). Other studies demonstrated that the C-terminus of OAS1 may also contain RNA binding sites, which correlates to the regions throughout exon 4 and 5 (Hartmann R et al., 2003). This N-terminal extension forms a hook at the end of the alpha-helix, which helps anchor the C-terminus domain (Hartmann R et al., 2003).

The substrate binding domains of OAS were determined using radiolabelled oligoadenylate dimers which where then cross-linked to OAS. After trypsin digestion and purification, mass spectrometry was used to determine the binding sites of OAS (Sarkar et al., 2002). Two such homologous domains are located from amino acid 83 to 91 and 201 to 209 of OAS1b. Both sites appear to bind independently and therefore do not constitute two parts of the same site (Sarkar et al., 2002). It was also determined that the region homologous to site 83 to 91 binds to the growing oligo chain and accepts new ATP's from the region homologous to site 201 to 209 (Sarkar et al., 2002). OAS1 functions as a multimer and the region necessary for oligomerization is found from amino acid 328 to 330 (Ghosh A et al., 1997b).

A few other domains have been predicted based upon sequence similarity. The p-loop is found between amino acid 58 to 65 of the OAS1b protein and has been hypothesized to interact with the double-stranded RNA of the flavivirus (Perelygin AA et al., 2002). It may also be possible that the p-loop is interacting with the substrate molecule. It has not yet been determined what effect this deletion has on the activity or function of OAS1b. Another interesting section of OAS1b found at position 328 to 345 is extremely similar to a catalytic region of the Cytokine Receptor Common Beta Chain (Blast by MetaServer Structure Predictor). This section of the Beta Chain has been shown to form a hydrophobic core which interacts with other hydrophobic regions of the protein (Bagley CJ et al., 1997). An extremely hydrophobic span of amino acids between sites 352 to 371 is followed by a short tail, which is unique to *Oas1b*. Protein prediction models have determined a significant possibility that this is a transmembrane domain (Figure 5).

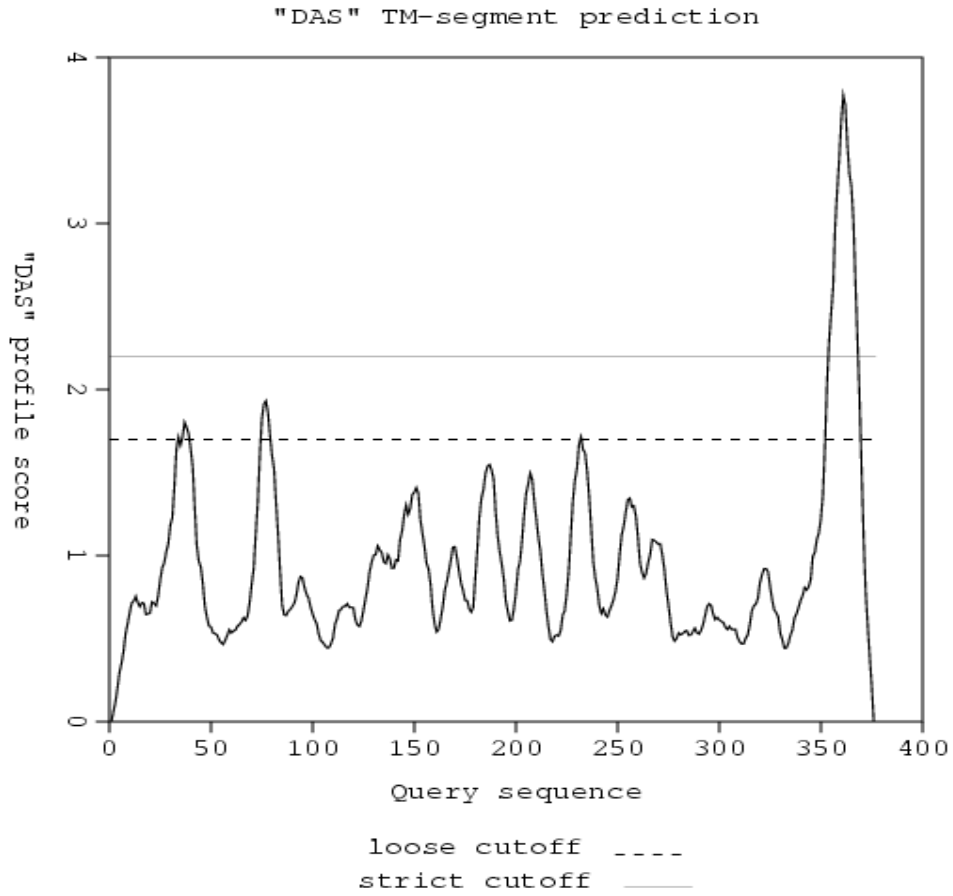

Figure 5. A display of the hydrophobicity profile scores for OAS1b. Note that the C-terminus of OAS1b well exceeds the strict cutoff for a transmembrane domain.

Part II. The structure, life cycle, and evolution of Flaviviruses

About 73 viruses belong to the genus *Flavivirus*. Flaviviruses are positive-sense, single-stranded, enveloped RNA viruses, most of which are arthropod-borne. Flaviviruses can be transmitted by ticks (17 species out of 73) or mosquitoes (34 species) but for some of them there are no known vectors (22 species) (Westaway EG et al., 1985). In most flavivirus species, the principal vertebrate hosts are rodents or birds; humans and domestic animals are generally dead-end hosts. However, at least 40 species of flaviviruses are associated with human disease (Monath TP, 1996). Some flavivirus-associated diseases have been a serious public health concern in recent years. There are emerging and re-emerging viruses that can cause epidemic outbreaks in human like dengue fever (50 to 100 million cases/year worldwide), Japanese encephalitis (30-50,000 cases) and Yellow fever (200,000 cases) (Bhamarapravati N, 1997; Monath TP and Heinz FX, 1994; Vainio J and Cutts F, 1998). In the United States, a West Nile virus outbreak that originated in New York City in 1999 has since spread across the entire country (for a review, see (Brinton MA, 2002), causing 284 deaths in 2002 (4156 cases).

Though similar in structure, different strains of flavivirus can cause very different symptoms and immune responses in the host. There are two main categories of diseases caused by flaviviruses:

- Encephalitis (caused by the Japanese encephalitis virus (JEV), the West Nile virus (WNV), the dengue fever virus (DFV) and tick-borne encephalitis viruses (TBEV)). The direct cause of encephalitis in flavivirus infection remains unknown, though apoptosis of neurons, synthesis of cytokines, nitrogen intermediates synthesized by

inflammatory cells, and neuronal injury by neutrophils have been implicated (Andrews DM et al., 1999; Matthews V, 2000; McMinn PC, 1997; McMinn PC, 1996; Parquet MC et al., 2001).

- Hemorrhagic fever (caused by Yellow fever virus (YFV), and DFV). Hemorrhagic fever leads to increased vascular permeability which causes plasma leakage throughout the body (WHO, 1997).

Much of what is known about the structure, life cycle, and evolution of flaviviruses has been discovered through studies using DFV, YFV, WNV, and JEV. All strains belonging to the flavivirus family have conserved protein motifs, gene structure, and replicative cycles. The mosaic of information presented is therefore assumed to be true for all recognized strains.

Structure

The flavivirus genome is approximately 11 kb in length, with an open reading frame (ORF) encoding a single polyprotein which is cleaved into ten separate proteins by viral and cellular proteases (Castle E et al., 1986; Hurrelbrink R and McMinn C, 2003; Mason PW et al., 1987). The RNA genome encodes three structural proteins and seven nonstructural proteins (Marin MS, 1995; Zanotto PM, 1996). The capsid protein (C), the membrane precursor protein (prM), and the envelope protein (E) are the structural proteins encoded by the viral genome. The seven nonstructural proteins are denoted NS1, NS2a, NS2b, NS3, NS4a, NS4b, and NS5 (Speight G et al., 1988; Trent DW et al., 1987). The 5' untranslated region (UTR) is capped and is conserved among flaviviruses (Brinton MA and Dispoto JH, 1988). The 3' UTR is conserved in sequence

and structure only in mosquito-borne flaviviruses and plays a crucial role in RNA replication of the YFV, DFV, and WNV (Lo MK, 2003; Zeng L et al., 1998).

The mature virion consists of the three structural proteins (C, E, and M) and the RNA genome (Bell JR et al., 1985; Boege U et al., 1983). Electron micrographs of dengue virions discovered a relatively smooth surface and an electron-dense core surrounded by a lipid bilayer (Kuhn RJ et al., 2002). Multiple copies of the capsid protein surrounds the genomic RNA forming the nucleocapsid core of the mature dengue virion (Zhang et al., 2003). The lipid bilayer surrounding the nucleocapsid is derived from the endoplasmic reticulum of the host cell (Zhang et al., 2003). The membrane envelope outside the lipid bilayer of dengue viruses is composed of 180 copies of the E glycoprotein and 180 copies of the M protein (Kuhn RJ et al., 2002). The E and M proteins are anchored to each other by their C-terminal domains (Lindenbach BD and Rice CM, 2001).

The NS1 protein of YFV has been shown to play a crucial role in viral RNA replication, either prior to or at minus strand synthesis (Lindenbach BD and Rice CM, 1997). The NS2a protein is a hydrophobic protein with several putative transmembrane domains (Henchal EA and Putnak JR, 1990). Studies of the Kunjin virus revealed that NS2a is essential for packaging RNA into virions or virus-like particles (Liu et al., 2003). Recent studies of the Kunjin NS2a protein have demonstrated its ability to inhibit interferon (Liu et al., 2004). NS2b is also a hydrophobic protein and a cofactor of NS3, which acts as a viral protease (Niyomrattanakit P et al., 2004). The Dengue NS3 protein is a serine protease and a helicase which is essential for viral replication (Arias CF et al., 1993; Niyomrattanakit P et al., 2004). Little is known about the roles of NS4a and

NS4b, but both proteins have interferon inhibiting properties, with NS4b showing higher levels of inhibition (Munoz-Jordan et al., 2003). The NS5 protein is an RNA-dependent RNA polymerase which has been shown to interact with the NS3 protein to form a replication complex in the Japanese Encephalitis virus (Chen CJ et al., 1997; Henchal EA and Putnak JR, 1990).

Life Cycle

Transmission

There are three main classes of flaviviruses which can be grouped by their (or lack of) vector; mosquito-borne, tick-borne, and no known vector (Gaunt MW et al., 2001; Kuno G et al., 1998; Leyssen P et al., 2001). Using the NS5 sequence (Kuno G et al., 1998) and the E protein sequence (Gaunt MW et al., 2001), phylogenetic analysis yielded trees showing the deepest branch between the no known vector viruses and the common ancestor of the mosquito- and tick-borne viruses (Figure 6). This indicates that the mosquito- and tick-borne clades arose from the no known vector flaviviruses, suggesting the use of vectors is a derived characteristic. As Figure 6 depicts, the mosquito-borne branch of flaviviruses can be further separated based on the species of mosquito that commonly transports the disease. Mosquito-borne flavivirus transmission is dominated by the *Aedes* and *Culex* genera of mosquito, which are among the small number of genera that are globally dispersed (Gaunt MW et al., 2001). By examining the blood meals of these mosquitoes, it was found that *Aedes* typically use mammals as hosts (Christensen HA et al., 1996; Savage HM, 1993). *Culex* mosquitoes primarily feed on both birds and mammals (Apperson CS et al., 2002; Arunachalam N et al., 2005). These feeding behaviors explains the association of *Aedes* vectored

flaviviruses predominance in mammals and *Culex* flaviviruses in both birds and mammals (Gaunt MW et al., 2001).

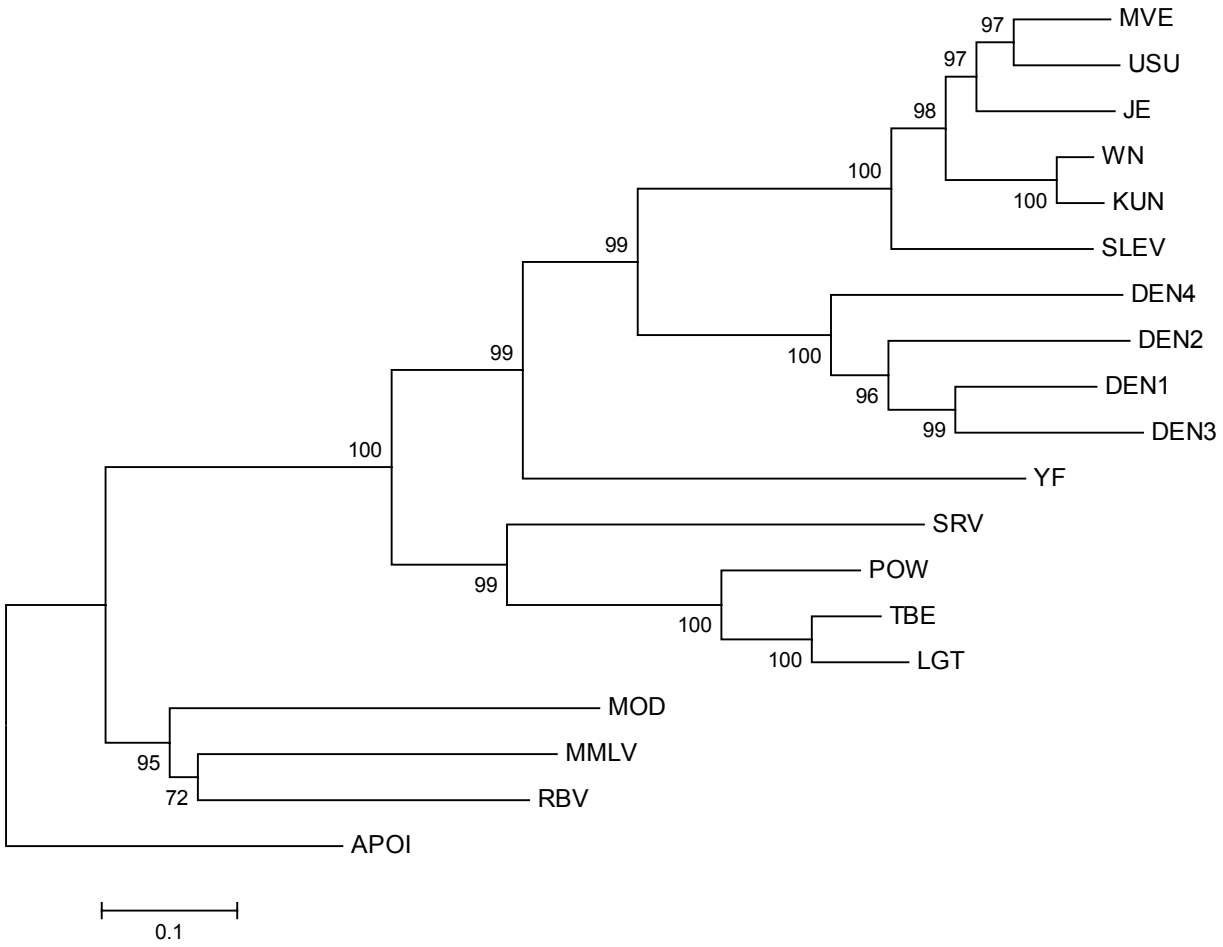

Figure 6. A phylogenetic tree based on the E gene amino acid sequence with the specific clades indicated at the right hand side. Dengue virus (DENV-1 through 4), St. Louis encephalitis virus (SLEV), West Nile virus (WNV), Kunjin virus (KUNV), Murray Valley encephalitis virus (MVEV), Japanese encephalitis virus (JEV), Powassan virus (POWV), Langat virus (LGT), Tick-borne encephalitis virus (TBEV), Apoi virus (APOIV), Modoc virus (MODV), Montana myotis leukoencephailitis virus (MMLV), and the Rio Bravo virus (RBV) are shown (Gaunt MW et al., 2001).

Entry

Dengue fever virus has been shown to initially interact with heparan sulfate on the outside of the host cells, thus accumulating on the surface of the cell (Chen Y et al., 1996; Chen Y, 1996; Leyssen P, 2000). The fusion protein of both the DFV and Tick-borne encephalitis virus is the E protein, which dimerizes upon acidification of its environment (Modis et al., 2004). These E protein dimers then trimerize forming clusters on the liposome surface (Modis et al., 2004). It has been suggested that this clustering of E trimers may induce curvature and promote fusion (Modis et al., 2004). The conformational changes in the E protein assembly has been implicated as the source of fusion and channel formation for infection by Dengue viruses (Kuhn RJ et al., 2002; Modis Y, 2004). Upon channel formation the nucleocapsid of the Dengue particle is expelled into the cytosol where the RNA genome is uncoated (Muñoz-Jordán JL et al., 2003) (See Figure 7).

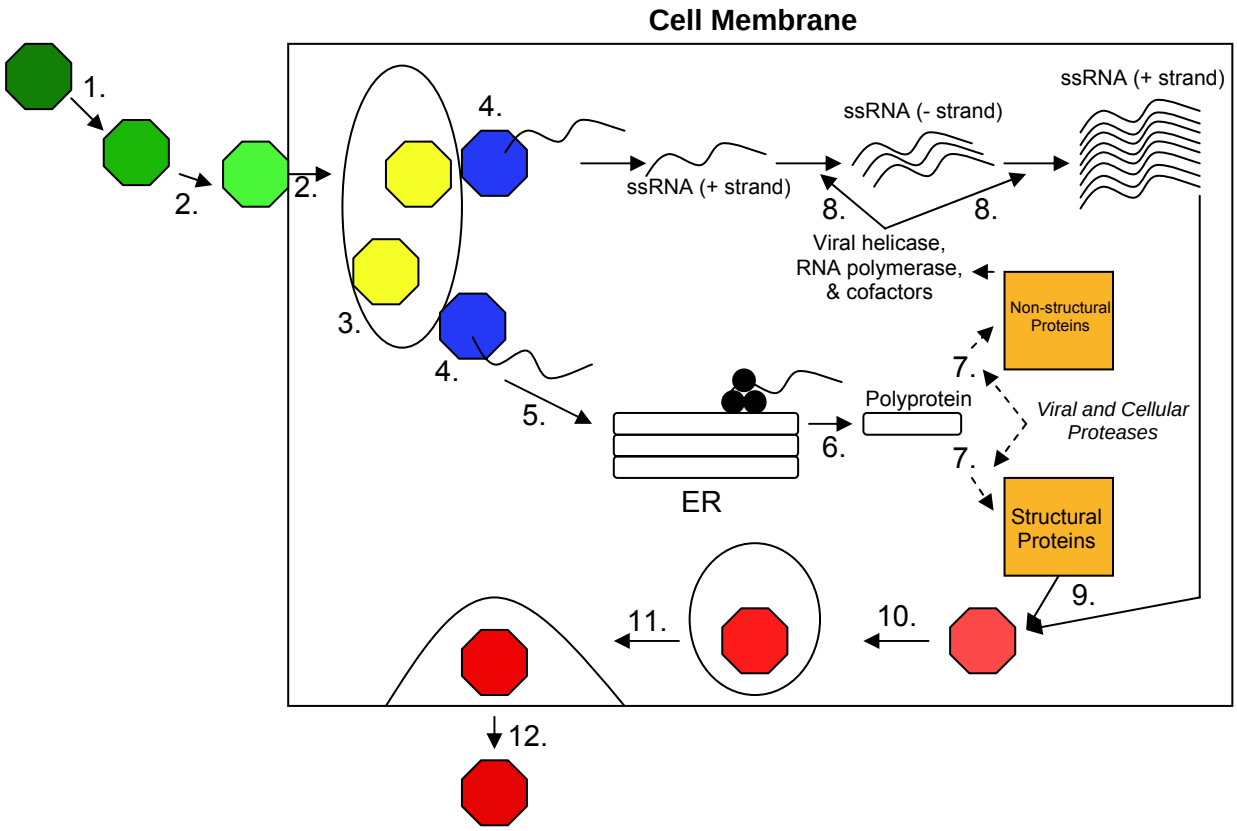

Figure 7. The replicative cycle of the flavivirus. 1, adsorption; 2, receptor-mediated endocytosis; 3, low-pH fusion in lysosomes; 4, uncoating; 5, cap-mediated initiation of translation; 6, translation of the viral RNA into viral precursor polyprotein; 7, co- and posttranslational proteolytic processing of the viral polyprotein; by cellular and viral proteases; 8, membrane-associated synthesis of template minus-strand RNA and progeny plus-strand RNA; 9, assembly of the nucleocapside; 10, budding of virions in the endoplasmatic reticulum; 11, transport and maturation of virions in the endoplasmatic reticulum and the Golgi complex; 12, vesicle fusion and release of mature virions. ss, single stranded (Leyssen et al., 2000).

Translation and replication

The 5' UTR of the viral genome directs the RNA to the ribosomes for translation of the polyprotein (Wengler G and Wengler G, 1993). The 5' UTR of flaviviruses contain a type I $m^7GpppN_1mpN_2$ cap structure, which may aid in translation initiation (Wengler G and Wengler G, 1993). Once translated, the polyprotein is cleaved by cellular proteases such as furin (Stadler K et al., 1997; Zhang Y, 2003), signalases (Pryor MJ et al., 1998), and viral proteases (Leyssen P et al., 2000). RNA replication is achieved by RNA dependent RNA polymerase (Bartholomeusz AI and Wright PJ, 1993). The NS5 protein of the Dengue virus has been shown to have RNA dependent RNA polymerase activity (Tan BH et al., 1996). The NS3 protein of Dengue virus was shown to have helicase properties in the c-terminal region, which unwinds the positive and negative strands of RNA after replication (Arias CF et al., 1993). The flavivirus nucleocapsid is assembled from the structural C protein which engulfs the viral RNA genome (Henchal EA and Putnak JR, 1990). The assembly and presence of immature Japanese encephalitis and Kunjin viral particles can be found in the rough endoplasmic reticulum of the cell (Oyanagi S et al., 1969), Budding is assumed to occur through the golgi followed by exocytosis of the mature virus particles (Henchal EA and Putnak JR, 1990; Mackenzie JM, 2001).

Flavivirus infection in the house mouse

It has been known for a long time that some laboratory strains (namely Det, BSVR, BRVR and PRI) of the house mouse are resistant to flavivirus infection (Sabin AB, 1952; Webster LT, 1937). Resistant mice are susceptible to infections with other

types of viruses, but are resistant to all flaviviruses (Shellam GR, 1998), with the exception of a few viral strains (Brinton MA and Fernandez AV, 1983; Sabin AB, 1952; Sangster MY et al., 1998). Resistant mice can be infected by flaviviruses but the virus titers in their tissues are 1,000-10,000 times lower than those of susceptible mice, and the spread of the infection is much slower (Goodman GT and Koprowski H, 1962; Sabin AB, 1952). Comparisons of susceptible and resistant mice indicate that survival of animals infected by the West Nile virus correlates with restriction of viral replication to the central nervous system (Mashimo et al. 2002).

Resistance is encoded by a single autosomal locus located on chromosome 5 in the mouse genome (Sangster MY et al., 1993). The responsible locus (*Oas1b*) was originally assigned the name *Flv* for its ability to prevent flavivirus infection and subsequent illness in mice (Sangster MY et al., 1994). Susceptible mice, such as C3H/eJ, C57BL/6J, and BALB/c, are denoted Flv^s, since they are susceptible to flavivirus infection. CASA/Rk, CAST/Ei, and C3H.M.domesticus strains were found the have the resistant allele Flv^r (Sangster MY et al., 1993; Urosevic N et al., 1999). The MOLD/Rk strain was found to have a different allele that was designated Flv^{mr}, (minor resistance) (Sangster MY et al., 1993). The Flv^{mr} allele has been shown to protect mice from YFV, but not from Murray Valley encephalitis (MVE) (Sangster MY et al., 1993), suggesting that different alleles are more effective against certain strains of flavivirus than others. The minor resistant allele and the resistant allele were shown to be dominant and the susceptible allele recessive (Sangster MY et al., 1993; Urosevic N et al., 1999).

However all of the strains of flavivirus used in these studies do not typically utilize the rodent as a host. One of the more notable mouse based strains of flavivirus is the Modoc virus, whose principal host is the deer mouse (*Peromyscus maniculatus*). The Modoc virus has the same genomic organization and conserved motifs as other flaviviruses, such as YFV, DFV and WNV. There is no known vector for the Modoc virus and there is evidence of horizontal transmission between mice in captivity (Davis JW and Hardy JL, 1973; Davis JW and Hardy JL, 1974; Davis JW et al., 1974; Leyssen P et al., 2001). Because Modoc virions can be found in the urine of infected animals, it is believed that the Modoc virus may be passed from mouse to mouse through urine. Since mice inhale each others urine, either directly from another mouse or from urine markings, this would make a sensible transmission path and plausibly reflect what is happening in nature.

Interactions between primates and flaviviruses

The majority of flaviviruses use either birds or small mammals as their natural host; however primates serve as either the main host or the only known host for a handful of strains. One of the most severe human based strains is dengue fever virus (DFV), which causes 50 to 100 million cases per year and leads to about 21,000 deaths per year worldwide. Most outbreaks of the disease occur in equatorial regions of the world where the *Aedes aegypti* mosquito flourishes (Figure 8). There are four known serotypes of DFV which are designated DEN-1, DEN-2, DEN-3, and DEN-4. The outcome of infection can range from mild illness to severe or fatal hemorrhagic disease (Gubler, 1998a).

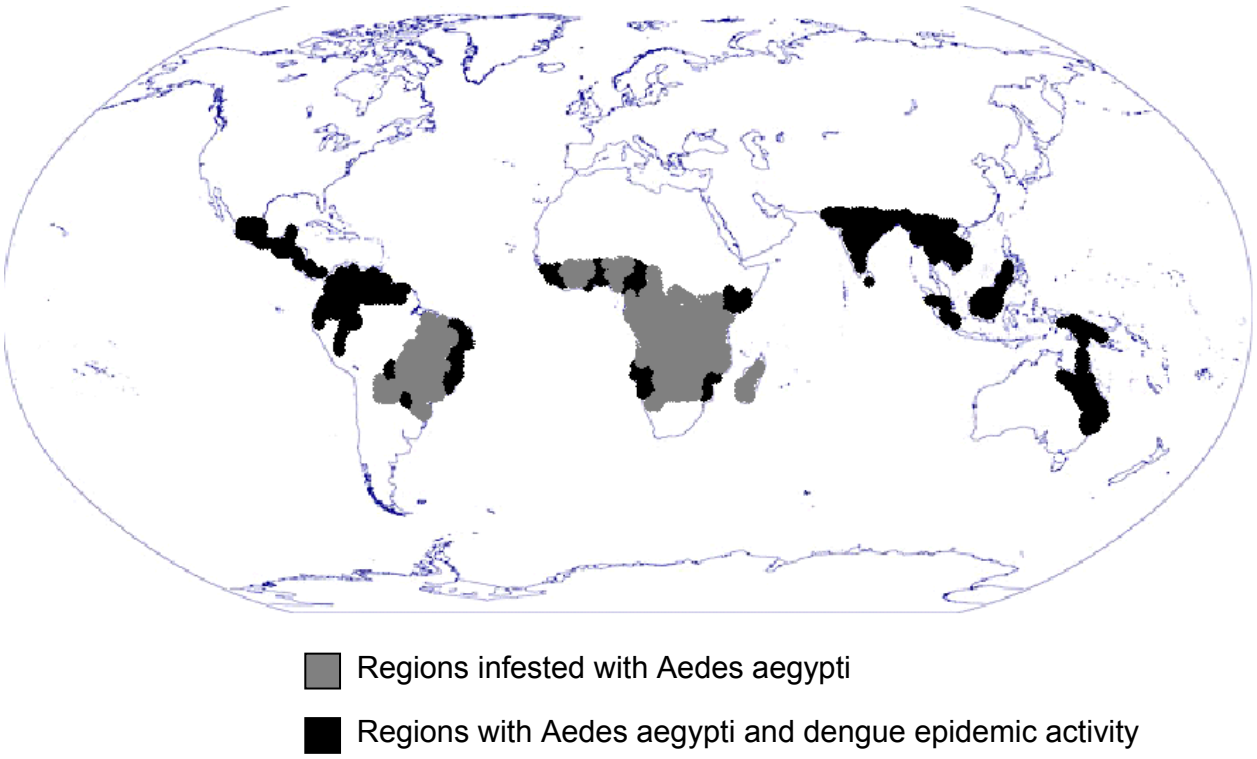

Figure 8. A map of the world which shows areas infested with the *Aedes aegypti* mosquito and regions that are prone to dengue epidemics (Gubler, 1998a).

Other primate based strains of flavivirus include yellow fever virus (200,000 cases per year and 30,000 deaths per year worldwide) and hepatitis C (3 to 4 million cases per year and 52,000 deaths per year worldwide). Many bird related strains of flaviviruses (WNV, JE, KV, and MVE) also have the ability to infect humans, thus leading to illness. Such infections are usually not deadly unless the infected individual is very young or immunodeficient.

Chapter 1

LONG-TERM BALANCING SELECTION AT THE WEST NILE VIRUS RESISTANCE GENE, *Oas1b,* MAINTAINS TRANSSPECIFIC POLYMORPHISMS IN THE HOUSE MOUSE

William Ferguson [*, †], Shira Dvora [*], Juliana Gallo[*], Annie Orth[‡], and Stéphane Boissinot [*, †]

[*] Department of Biology, Queens College, the City University of New York, 65-30 Kissena Boulevard, Flushing, NY 11367, USA.

[†] Graduate School and University Center, the City University of New York, 365 Fifth Avenue, New York, NY 10016, USA.

[‡]Biologie Intégrative, Institut des Sciences de l'Evolution, Université Montpellier 2—Centre National de la Recherche Scientifique, Case courrier 63, Place Eugène Bataillon, Montpellier, France

Corresponding author:
Stéphane Boissinot
Department of Biology
Queens College, CUNY
65-30 Kissena Boulevard
Flushing, NY 11367-1597
Tel: 718 997 3437
Fax: 718 997 3445
Email: stephane.boissinot@qc.cuny.edu

Published: Molecular Biology and Evolution 2008 25(8):1609-1618

ABSTRACT

Oligoadenylate Synthetases (OAS) are interferon-inducible enzymes that participate in the first line of defense against a wide range of viral infection. Recent studies have determined that *Oas1b*, a member of the OAS gene family in the house mouse (*Mus musculus*), provides specific protection against flavivirus infection (e.g. West Nile virus, Dengue Fever virus and Yellow Fever virus). We characterized the nucleotide sequence variation in coding and noncoding regions of the *Oas1b* gene for a large number of wild-derived strains of *M. musculus*, and related species. Our sequence analyses determined that this gene is one of the most polymorphic genes ever described in any mammal. The level of variation in non-coding regions of *Oas1b* is an order of magnitude higher than the level reported for other regions of the mouse genome and is significantly different from the level of intra-specific variation expected under neutrality. Furthermore, a phylogenetic analysis of intronic sequences demonstrated that *Oas1b* alleles are ancient and that their divergence pre-dates several speciation events, resulting in trans-specific polymorphisms. The amino-acid sequence of *Oas1b* is also extremely variable, with 1 out of 7 amino-acid positions being polymorphic within *M. musculus*. *Oas1b* alleles are comparatively more divergent at synonymous positions than most autosomal genes and the ratio of non-synonymous to synonymous substitution is remarkably high, suggesting that positive selection has been acting on *Oas1b*. The ancestry of *Oas1b* polymorphisms and the high level of amino-acid polymorphisms strongly suggest that the allelic variation at *Oas1b* has been maintained in mouse populations by long-term balancing selection. This pattern of evolution is extremely unusual and had only been demonstrated before for genes of the Major Histo-Compatibility Complex.

INTRODUCTION

Organisms and their pathogens are engaged in a co-evolutionary process resulting from their conflicting interests: hosts must retain their ability to resist and eventually clear infection, and pathogens must retain their infectivity and ability to evade host defense. Host defense genes are in the forefront of this process and bear the signature of the interactions between the host and its pathogens. Many defense genes have been shown to evolve rapidly at the protein level because of the action of positive selection. This pattern of evolution is caused either by a succession of selective sweeps of new resistance alleles (the “arms-race” model of evolution) or by the persistence of dynamic polymorphisms due to selective fluctuations in allelic frequency (the “Red Queen” model) (for a review see (Woolhouse et al., 2002)). The selective mechanism responsible for the maintenance of polymorphism in populations is generally referred to as balancing selection and can be caused either by heterozygote advantage, frequency-dependent selection (Kojima 1971) or selection that varies in time and space (Hedrick, Ginevan, and Ewing 1976). Although most balanced polymorphisms tend to be transient, in some very rare cases, balancing selection has been able to maintain alleles for very long periods of evolutionary time. At the Major Histo-Compatibility Complex (MHC) in mammals (Hedrick and Thomson 1983; Figueroa, Gunther, and Klein 1988), and at the plant resistance (*R*) genes (Bergelson et al. 2001), a remarkably large number of alleles have been maintained by balancing selection, and some of these polymorphisms are so ancient that they predate several speciation events, resulting in trans-specific polymorphisms (Figueroa, Gunther, and Klein 1988).

2’,5’-oligo-adenylate synthetases (OAS) are members of the interferon pathway, which plays an important anti-viral role (for a review see (Hovanessian and Justesen 2007)). Interferon upregulates the transcription of OAS genes and OAS proteins are converted to an enzymatically active form by a double-strand RNA-dependent process. From ATP molecules, activated OAS proteins synthesize ppp(A2’p)nA oligoadenylates (2-5A) which bind to the latent endoribonuclease Rnase L leading to the dimerization and activation of RNase L, followed by the degradation of cellular and viral RNA. The

OAS gene family has four members in humans (OAS1, OAS2, OAS3, and OASL) and ten members in the house mouse, *Mus musculus* (*Oas1a* to *Oas1h*, *Oas2* and *Oas3*) (Eskildsen et al. 2002; Mashimo et al. 2003). Some of the mouse *Oas1* copies, such as *Oas1b* and *Oas1c*, have lost their synthetase activity (Rogozin, Aravind, and Koonin 2003). As these copies remain otherwise relatively conserved, it is likely that they have additional functions.

It was recently found that the *Oas1b* gene confers resistance against infection with West Nile virus and other flaviviruses (e.g. Dengue Fever virus, Yellow Fever virus, and Japanese Encephalitis virus) (Mashimo et al. 2002; Perelygin et al. 2002). Resistant mice can be infected by flaviviruses but the virus titers in their tissues are 1,000 to 10,000 times lower than those of susceptible mice, and the spread of infection is much slower (Sabin 1952; Goodman and Koprowski 1962). The vast majority of laboratory strains are susceptible to flaviviral infection and develop severe encephalitis due to a nonsense mutation in exon 4 of *Oas1b* resulting in a truncated protein (Mashimo et al. 2002; Perelygin et al. 2002). In contrast, most wild mice are resistant (Sangster, Mackenzie, and Shellam 1998) and at least two functionally different resistance alleles exist in nature: the “major resistance” allele that confers resistance against the vast majority of flaviviruses and a “minor resistance” allele that protects efficiently against Yellow Fever virus yet has little effect on the outcome of infection with other viruses (e.g. Murray Valley Encephalitis virus (Sangster et al. 1993)). It was found that these two resistance alleles differ at 14 amino-acid positions (i.e. 4.2%) (Perelygin et al. 2002). This level of amino-acid divergence is extremely high and is comparable to the divergence usually observed between distantly related species (e.g. mouse and rat proteins differ on average by 5% at the amino-acid level; (Gibbs et al. 2004)). Although these data suggest that balancing selection could be acting on this gene, this pattern could also result from the complex evolutionary history of the house mouse, a species in which hybridization between sub-species (Ferris et al. 1983; Orth et al. 1998; Bonhomme et al. 2007) and with closely related species is known to occur (Greene-Till, Zhao, and Hardies 2000; Orth et al. 2002).

We assessed the variation of the *Oas1b* gene and we analyzed its molecular evolution in the house mouse and related species. Coding and non-coding regions of

Oas1b were sequenced in a sample of wild-derived strains representative of the entire genetic diversity of this species. The *Oas1b* gene is extremely polymorphic within *Mus musculus* and even within its subspecies, *M. m. musculus* and *M. m. domesticus*. Polymorphisms at *Oas1b* are very ancient and predate several speciation events. Our data demonstrates that the *Oas1b* gene has evolved under balancing selection for more than 2.8 Million years and constitutes one of the very few cases of old trans-specific polymorphisms maintained by long-term balancing selection in a mammal.

MATERIALS AND METHODS

Sampling – The house mouse *Mus musculus* is a complex species that originated in India less than 1 MYA (Boursot et al. 1996). From Northern India, this species gradually colonized the periphery of the Euro-Asiatic continent where three well-defined subspecies are found (*M. m. musculus* from Eastern Europe to Northern China, *M. m. domesticus* in Western Europe, North Africa, and the Middle East, and *M. m. castaneus* in South East Asia). These 3 subspecies still hybridize in natural conditions so that a clear-cut hybrid zone exists in Europe, while the subspecies seem to intermix to a greater extent in Asia. Mice from Japan are hybrids between *M. m. musculus* and *M. m. castaneus* and are sometimes referred to as "*M. m. molossinus*" (Yonekawa et al. 1986). We obtained DNA samples from 34 wild-derived strains of house mice, 23 from the "Conservatoire de la souris sauvage" (Université Montpellier II, Montpellier, France) and 11 from the Jackson laboratory (Bar Harbor, Maine, USA). This sample includes strains from each of the three major subspecies, *M. m. musculus* (11 strains), *M. m. domesticus* (12 strains), and *M. m. castaneus* (4 strains), strains with undefined sub-specific status from Iran, India and Pakistan (4 strains) and hybrid strains from Japan (5 strains). In addition, we obtained DNA from other species within the genus *Mus* to determine the phylogenetic context of *Oas1b* evolution. These species include the four European species of mice, *M. spretus* (3 strains), *M. macedonicus* (3 strains), *M. spicilegus* (4 strains) and *M. cypriacus* (1 strain), which collectively diverged from the house mouse ~1.4 MYA, the Indian species *Mus famulus* (2.8 MYA) and the South East

Asian species *Mus caroli* (3.2 MYA). As an outgroup we used the species *Coelomys pahari* which diverged from the genus *Mus* 6.5 MYA.

Molecular analysis – Each of the six exons of *Oas1b* was amplified by Polymerase Chain Reaction. We also amplified partial fragments from introns 2, 3, 4 and 5, from a fragment containing the 3' Untranslated Region (3'UTR) and from a non-coding region located 10Kb downstream of the *Oas1b* gene. The products of amplification were purified and sent for sequencing to the company Macrogen (Seoul, Korea). To assess the variation of the *Oas1b* gene, we compared it with the level of variation in two neutral regions of the mouse genome. We selected two autosomal segments located at least 1Mb from any known gene, one on chromosome 12 (from nucleotides 64,811,904 to 64,813,152 on the February 2006 assembly of the mouse genome at http://genome.ucsc.edu) and one on chromosome 19 (from nucleotides 19,726,706 to 19,727,769). These fragments were amplified by PCR in each wild-derived strain and sequenced. The list of primers used for amplification is available on request from the corresponding author. DNA sequences have been deposited at the EMBL database under accession numbers AM887890 to AM887932.

Data analysis – Sequences were aligned and manipulated using the BioEdit platform (Hall 1999). Distances, including synonymous (*ds*) and non-synonymous (*dn*) distances, were calculated using the MEGA3 software package (Kumar, Tamura, and Nei 2004). Phylogenetic analyses were performed using the neighbor joining method (as implemented in MEGA3) and the maximum likelihood method (using PHYML (Guindon et al. 2005)). The nucleotide diversity was estimated using the parameter π and Watterson's θ. We tested for selection using the Hudson-Kreitman-Aguade test (known as the HKA test (Hudson, Kreitman, and Aguade 1987)). The HKA test is based on the assumption that, in the absence of selection, silent-site polymorphisms and divergence are expected to be the same across all loci. Basically, the test compares the ratio of polymorphism to divergence between a gene of interest (*Oas1b* in our case) and a neutral region of the genome (the neutral regions on chromosome 12 and 19, see the "Molecular analysis" section). If the difference between the two ratios is statistically significant from the null hypothesis using a goodness-of-fit test, we can then reject the hypothesis of neutrality. We also used the McDonald and Kreitman test (known as the

MK test (McDonald and Kreitman 1991)): if variation at a locus is neutral, then the rate of substitution between species and the amount of variation within species are a function of the mutation rate. Therefore, under neutrality, the ratio of non-synonymous to synonymous fixed differences between species should be the same as the ratio of non-synonymous to synonymous polymorphisms within species. If the two ratios differ significantly, we can reject the hypothesis of neutrality. The DnaSP program (Rozas et al. 2003) was used to perform the HKA and MK tests. We tested the possibility that gene conversion between alleles and/or between paralogous copies of *Oas1b* affected the sequence of *Oas1b* alleles using the method of Sawyer (Sawyer 1989) implemented in the GENECONV program.

RESULTS

The coding sequence of *Oas1b* was obtained for 34 wild-derived strains of the house mouse (Figure 9). Fifty five amino-acid variants (at 51 positions) were detected within the species *Mus musculus*. This number is extremely large considering that this gene is only 377 amino acid long, i.e, about 1 out of 7 amino acid positions is variable. The premature stop codon responsible for the susceptible phenotype of the laboratory strains was found in 3 wild-derived strains (SKIVE, SF and DOT) and another stop codon, also in exon 4, was found in strain MBS. Single base pair deletions resulting in frame-shifts were detected in exon 1 and 4 of strains CTP and 22MO, respectively, and an in-frame deletion of 39 nucleotides resulting in a 13 amino acids deletion was found in exon 5 of the BZO strain. Out of the 55 variable amino acids, 24 polymorphisms are shared between at least two subspecies. In contrast, a single amino acid difference is fixed in one of the *M. musculus* subspecies (at amino-acid position 45, in *M. m. musculus*, including "*M. m. molossinus"*). This pattern of variation extends to inter-specific comparisons: there is not a single fixed amino acid that differentiates *M. musculus* from other species. In contrast, five polymorphisms are shared between *M. musculus* and either *M. spicilegus* or *M. spretus* and this number is presumed to be an underestimate due to the limited availability of strains of species other than *M. musculus*. The lack of fixed differences between taxa and the abundance of shared polymorphisms suggests that many polymorphisms are older than the origin of *M.*

musculus and at least 5 of them are older than the split between *M. musculus* and other European species (*M. spretus*, *M. spicilegus*, *M. macedonicus*, *M. cypriacus*), which occurred ~1.4 MYA. To confirm the ancestry of *Oas1b* alleles, the synonymous divergence (*ds*) between *M. m. musculus* and *M. m. domesticus* alleles was compared with the divergence of 183 autosomal genes for which a *M. m. domesticus* (from the

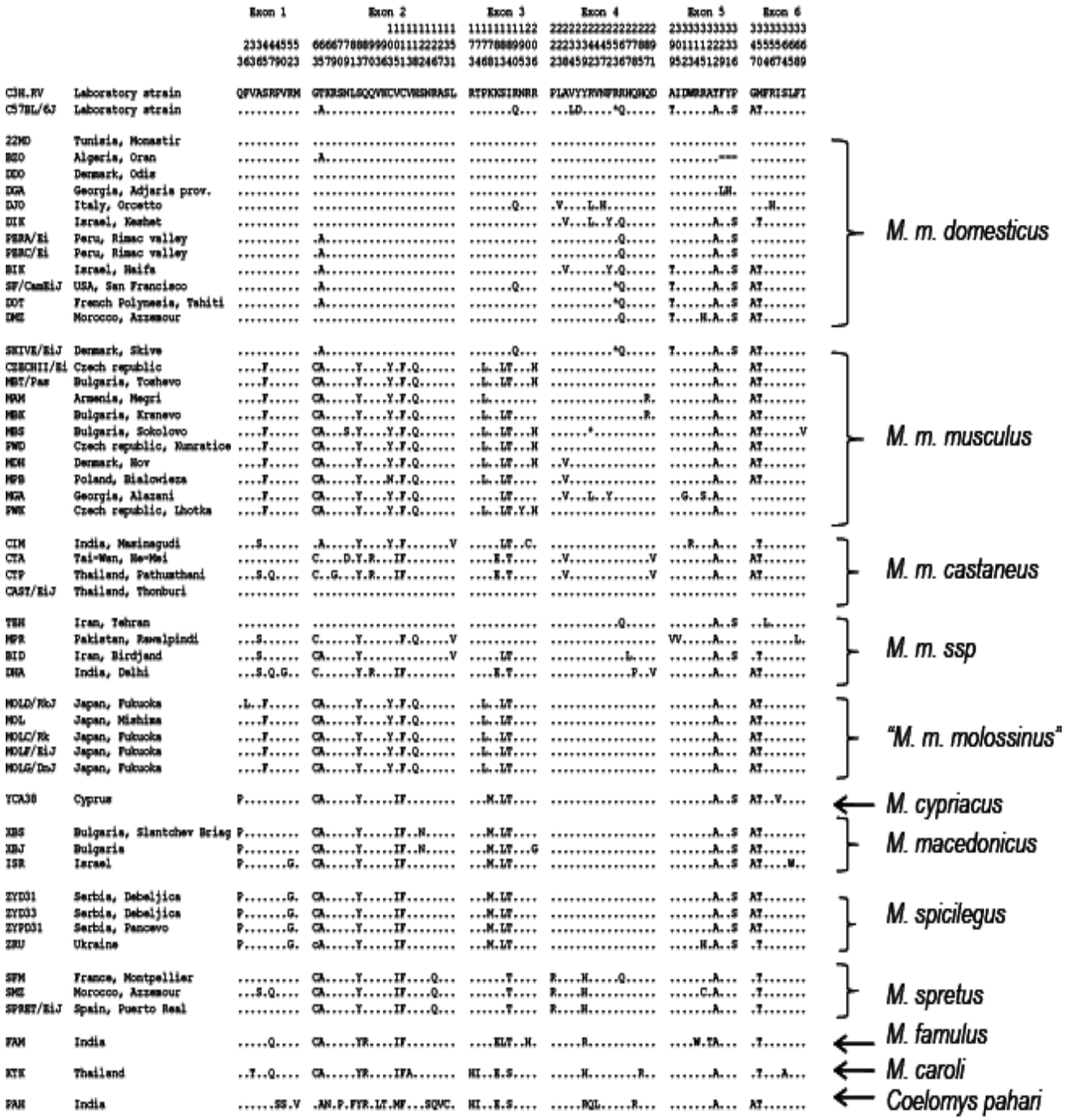

Figure 9: Variable amino acids at the Oas1b gene. The phenotypes associated with each allele are designated as follows: (s) susceptible strains carrying the premature stop codon in exon 4 (Mashimo et al. 2002; Perelygin et al. 2002), (ps) strains predicted to be susceptible based on their amino acid sequences, (r) resistant strains similar to the functionally characterized resistant allele (Sabin 1952; Goodman and Koprowski 1962; Mashimo et al. 2002; Perelygin et al. 2002), (pr) strains predicted to be resistant based on their amino acid sequence and (mr) strains carrying an allele similar to the minor resistance allele.

C57BL/6J strain) and a *M. m. musculus* (from the CZECHII/EiJ strain) sequence are available in databases (Figure 10a). On average, *Oas1b* alleles are much more divergent at synonymous sites than most autosomal genes and the *ds* between some *Oas1b* alleles can be as high as 4%. Only a handful of genes show such high *ds*, including several MHC genes. Figure 10a also shows that the distribution of inter-subspecific divergences is almost indistinguishable from the distribution of inter-specific divergences. On average, *M. m. musculus* and *M. m. domesticus Oas1b* alleles differ by 2.85% at synonymous sites, while they differ on average by 2.22% from European species. This clearly implies that allelic variation at the *Oas1b* gene is as old or older than the split between *M. musculus* and its European relatives.

The persistence of ancestral polymorphisms at *Oas1b* was further examined by performing a phylogenetic analysis on intronic sequences. Figure 11 shows a phylogenetic tree built using intron 5 sequences. This tree differs drastically from the species tree and shows strong evidence for trans-specific polymorphisms. Two clades supported by high bootstrap values and by several indels are apparent on the tree. Each of these clades contains *M. m. domesticus* sequences and sequences from other species. *M. m. domesticus* sequences of clade 1 are more closely related to *M. macedonicus*, *M. spicilegus* and *M. spretus* sequences than they are to *M. m. domesticus* sequences belonging to clade 2. Similarly, *M. m. domesticus* alleles in clade 2 are more closely related to *M. spicilegus* strain ZRU and to *M. famulus* than to clade 1. This clearly demonstrates that *M. m. domesticus* contains two deeply divergent allelic lineages at the *Oas1b* locus. These two lineages have separated before the split between *M. musculus* and *M. famulus* which took place ~2.8 MYA. These lineages have also been maintained in *M. spicilegus* because sequences of this species are found in both clades. For both *M. m. domesticus* and *M. m. musculus,* the nucleotidic diversity in introns 4 and 5 and in the 3'UTR is 5 to 10 times higher than the nucleotidic diversity in introns 2 and 3 and in two putatively neutral regions of the mouse genome (Table 2). This level of polymorphism is among the highest reported for any genomic region of the house mouse and is clearly outside the range reported by Zhang et al. for 44 genomic segments (~7 single nucleotide polymorphisms /Kb within *M. m. domesticus* (Zhang et al. 2005)). We tested if the excess of polymorphisms at the 3' end of *Oas1b* was

significantly different from the level of intraspecific variation expected under neutrality. To this end, an HKA test was performed using the two non-coding fragments on chromosome 12 and 19 as neutral regions of reference (see materials and methods). We were able to reject the null hypothesis of neutrality for introns 4 and 5 and the 3'UTR (in bold on Table 2), although not all comparisons yielded significant deviations from neutrality. In contrast, we can not reject the neutrality hypothesis for introns 2 and 3, and for the region located 10 Kb downstream of *Oas1b*. We also examined the possibility that gene conversion (Perelygin et al. 2006) could have affected the divergence of *Oas1b* alleles. Using GENECONV, we were unable to detect a single significant instance of gene conversion between *Oas1b* and another member of the *Oas1* gene family. Therefore, the large divergences between *Oas1b* intronic sequences result from the persistence of alleles over a long period of evolutionary time and not from the effect of gene conversion between paralogous copies. This analysis clearly indicates that *Oas1b* polymorphisms are ancient and strongly suggests that balancing selection has acted on *Oas1b* polymorphisms. It further suggests that the sites under balancing selection are located at the C-terminus of the *Oas1b* protein because the strongest evidence for trans-specific polymorphism is found for introns 4 and 5, and for the 3'UTR.

The high level of divergence of *Oas1b* alleles is also apparent at non-synonymous sites (Figure 10b). The average divergence at non-synonymous sites (*dn*) between *M. m. musculus* and *M. m. domesticus* is 2.00%. Very few genes show such a high *dn* and in our sample of 183 genes, only five genes (all involved in host-pathogens interactions) had a *dn* greater than 2.00%. This high level of *dn* is consistent with the exceptionally large number of amino-acid polymorphisms at *Oas1b*. We then calculated the *dn/ds* ratio for all genes in our data set. A high value for the ratio *dn/ds* indicates that selection favors non-synonymous substitutions, i.e. amino-acid changes, and is one of the predictions of balancing selection (Garrigan and Hedrick 2003). The *dn/ds* ratio is much higher for the *Oas1b* gene than it is for the vast majority of genes (Figure 10c). In general, the *dn/ds* ratio is very low (<0.1 and (Gibbs et al. 2004)) because most genes are evolving under strong purifying selection. In contrast, the *dn/ds* ratio is always higher than 0.4 for *Oas1b* and a number of pairwise comparisons are higher

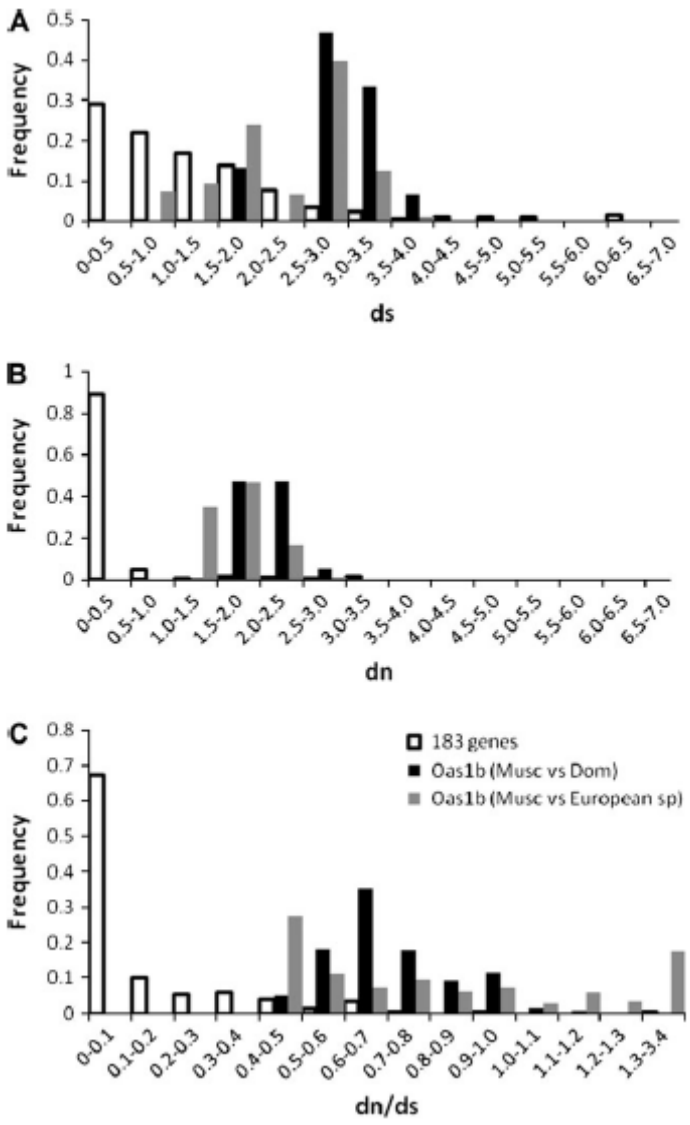

Figure 10: Comparison of the synonymous divergence (ds, panel A), the nonsynonymous divergence (dn, panel B), and their ratio (dn/ds, panel C) between Oas1b alleles and between 183 autosomal genes for which a Mus musculus musculus and Mus musculus domesticus alleles are available in the database. The dn/ds distribution is based on 153 genes because the ds value for 30 genes is 0, making dn/ds infinite.

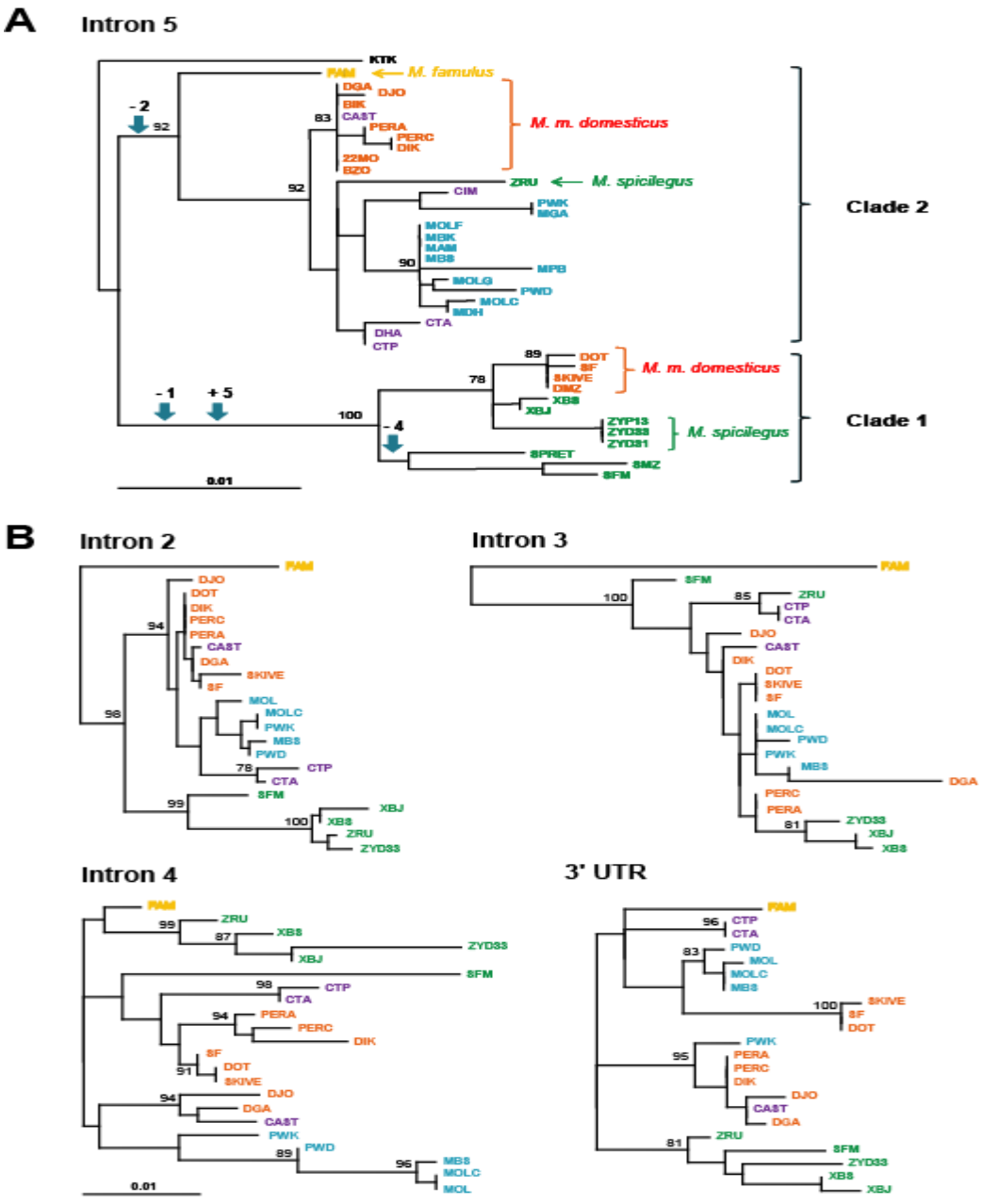

Figure 11: (A) Maximum likelihood phylogeny of Oas1b alleles based on intron 5 sequences. Arrows represent phylogenetically informative insertions (þ) and deletions (_). For instance, the arrow labeled þ5 on the branch leading to clade 1 indicates that all sequences in this clade share a 5-bp insertion. Bootstrap values higher than 75% are indicated. The tree was built using the HKY model of substitution but other models and methods produced nearly identical trees. Mus musculus musculus and Mus musculus molossinus strains are in blue, Mus musculus domesticus in orange, Mus musculus castaneus in purple, and European species (Mus spretus, Mus spicilegus, Mus macedonicus, and Mus cypriacus) in green. KTK is Mus caroli, and FAM is Mus famulus. (B) Maximum likelihood phylogeny of Oas1b alleles based on introns 2, 3, and 4 and on the 3# UTR.

than 1. The *Oas1b dn/ds* ratio was also calculated between *M. musculus* and the other species in our sample (Figure 10) and values of *dn/ds* as high as 3.0 were obtained for some comparisons. Although it is plausible that an undetermined fraction of non-synonymous changes resulted from a recent relaxation of selection on *Oas1b* (suggested by the presence of inactivating mutations in several strains) and caused an overestimation of *dn*, values of *dn/ds* higher than 1 are not consistent with neutral evolution. Instead, the high values of *dn/ds* suggest that selection in favor of non-synonymous mutations is indeed acting on *Oas1b* and confirm the action of balancing selection on *Oas1b* polymorphisms.

The balancing selection hypothesis further predicts an excess of non-synonymous polymorphisms within species. In *M. m. domesticus* and *M. m. musculus*, 65% and 82% of all polymorphisms are non-synonymous, respectively (Table 3). In contrast, the number of fixed non-synonymous and synonymous differences between *M. m. domesticus* and either *M. famulus* or *M. caroli* are basically identical. For *M. m. musculus* the difference is even more striking as the number of fixed synonymous differences is higher than the number of fixed nonsynonymous differences, for both comparisons. The MK test shows that the ratio of nonsynonymous to synonymous fixed difference differs significantly from the ratio of nonsynonymous to synonymous polymorphisms in *M. m. musculus* but not in *M. m. domesticus*, most likely because of the conservative nature of the MK test (Table 3). Although the excess of non-synonymous polymorphisms in both *M. m. musculus* and *M. m. domesticus* is consistent with the action of balancing selection at *Oas1b*, the result of the MK test needs to be interpreted with caution because some non-synonymous changes could reflect a recent relaxation of selection on *Oas1b*.

Summary Statistics for Different Noncoding Regions of *Oas1b* and for 2 Neutral Regions of the Mouse Genome (on chromosomes 12 and 19; see text)

							HKA Chromosome 12[g]		HKA Chromosome 19[h]	
Regions	L[a]	N[b]	S[c]	π (%)[d]	θ (%)[e]	D[f]	χ^2	P Value	χ^2	P Value
Intron 2	1525									
Mus musculus musculus		13	23	0.309	0.486	−1.05	0.248	0.618	0.593	0.441
Mus musculus domesticus		12	11	0.144	0.238	−1.78	1.808	0.179	0.134	0.715
Intron 3	725									
M. m. musculus		11	6	0.152	0.285	−1.61	0.176	0.675	0.919	0.338
M. m. domesticus		10	17	0.625	0.837	−0.31	1.672	0.196	0.088	0.767
Intron 4	710									
M. m. musculus		13	33	1.356	1.649	−0.31	2.656	0.103	4.568	**0.033**
M. m. domesticus		11	27	1.603	1.436	−0.59	7.360	**0.007**	5.777	**0.016**
Intron 5	690									
M. m. musculus		13	41	1.305	1.978	−0.79	1.851	0.174	3.461	0.063
M. m. domesticus		11	26	1.556	1.319	1.62	5.271	**0.022**	2.817	0.093
3′ UTR	667									
M. m. musculus		8	23	1.395	1.438	1.14	2.147	0.143	4.920	**0.027**
M. m. domesticus		11	21	1.444	1.162	1.37	6.025	**0.014**	2.963	0.085
3′ fragment	645									
M. m. musculus		9	7	0.267	0.399	−1.03	0.023	0.879	0.264	0.608
M. m. domesticus		5	4	0.372	0.298	1.70	1.121	0.289	0.204	0.651
Chromosome 12	817									
M. m. musculus		10	13	0.435	0.277	−0.99			0.109	0.742
M. m. domesticus		9	1	0.032	0.053	−1.14			0.518	0.475
Chromosome 19	1093									
M. m. musculus		11	8	0.150	0.257	−1.46	0.109	0.742		
M. m. domesticus		10	8	0.176	0.266	−1.46	2.833	0.092		

NOTE.—P values <0.05 are indicated in bold.
[a] Length of the fragment in base pairs.
[b] Number of strain sequenced.
[c] Number of segregating sites.
[d] Nucleotidic diversity.
[e] Watterson's estimator of nucleotidic diversity.
[f] Tajima's D.
[g] Results of the HKA test using the region on chromosome 12 as neutral region of reference.
[h] Results of the HKA test using the region on chromosome 19 as neutral region of reference.

Table 2: Values for pi, theta, Ka, Ks and Tajima D calculated using DnaSP 4.50.2 and Mega3. HKA values were determined using the putatively neutral regions of chromosome 12 and 19 in the mouse genome. Significant values are shown in bold.

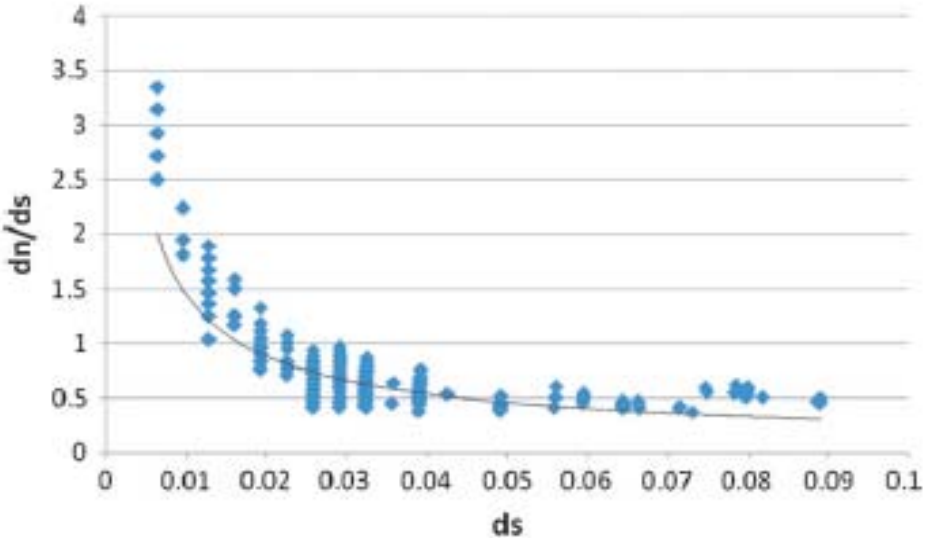

Figure 12: Relationship between synonymous divergence (ds) and the ratio dn/ds for Oas1b alleles in the genus Mus. The plot is best fitted to a power regression with a strong R2 value (0.67).

MK Neutrality Tests

	Mus famulus			*Mus caroli*		
	Synonymous[a]	Nonsynonymous[b]	*P* Value	Synonymous	Nonsynonymous	*P* Value
Mus musculus musculus						
Polymorphic changes	4	18	0.0095	4	18	0.0059
Fixed differences	15	12		21	16	
Mus musculus domesticus						
Polymorphic changes	14	26	0.5529	14	26	0.3065
Fixed differences	8	9		13	13	

NOTE.—Tests were performed with either *M. famulus* or *M. caroli* as outgroups.

[a] Synonymous (silent) mutations.

[b] Nonsynonymous (replacement) mutations.

Table 3: Results of the MK test for *Oas1b* using *M. famulus* and *M. caroli* for fixed difference analyses.

DISCUSSION

Our analysis revealed an extremely high level of variation at the 3' end of the *Oas1b* gene and provides strong evidence for trans-specific polymorphism in the genus *Mus*. The nucleotidic diversity in several introns of *Oas1b* is an order of magnitude higher than the diversity in putatively neutral regions of the mouse genome (the regions on chromosome 12 and 19 and the 44 regions analyzed in (Zhang et al. 2005)). The phylogenetic analysis of *Oas1b* introns further indicates that the allelic lineages found in *M. musculus* are older than the split between *M. musculus* and *M. famulus*, which occurred 2.8 MY ago. Thus, this polymorphism is the second oldest reported in mice, after the genes of the MHC (Edwards et al. 1997). Although the presence of such divergent allelic lineages within a species could result from incomplete lineage sorting, this is very unlikely because the maintenance of alleles over 2.8 MY would require an unrealistically large long-term effective population size. Assuming a conservative generation time of two generations/year, a long-term population size of more than one million individuals would be necessary to maintain the level of variation observed at the 3' end of *Oas1b* (using $\theta = 4N_e\mu$, where θ is the nucleotidic diversity, μ the mutation rate per site per generation and Ne the effective population size). This value of Ne is three to ten times higher than previous estimates (Eyre-Walker et al. 2002) and it is extremely unlikely that such a large effective population size was maintained for more than 2.8 MY, considering that house mice went through several episodes of speciation that are typically associated with population bottlenecks. It is also unlikely that a recent hybridization event could account for the presence of divergent lineages in *M. musculus*. Although *M. m. domesticus* occasionally interbreed with *M. spretus* under natural conditions (Greene-Till, Zhao, and Hardies 2000; Orth et al. 2002), the molecular divergence between these two species is much smaller than the divergence between *Oas1b* allelic lineages and experimental crosses between *M. musculus* and its Asian relatives fail to produce viable birth. Instead, our data strongly suggest that *Oas1b* polymorphisms have been maintained in populations for the last ~3 MY by the action of balancing selection. We provide several lines of evidence that strongly support this hypothesis, including an extremely high nucleotidic diversity in introns 4, and 5, and in

the 3'UTR, a significant deviation from neutrality as revealed by the HKA test, a high *dn/ds* ratio and an excess of non-synonymous polymorphisms. Balancing selection has been shown to act on a small number of loci in mammals (Hedrick and Thomson 1983; Figueroa, Gunther, and Klein 1988; Shyue et al. 1995; Bamshad et al. 2002; Verrelli et al. 2002; Newman et al. 2006; Baysal, Lawrence, and Ferrell 2007), but rarely results in trans-specific polymorphisms, probably because hosts eventually evolve resistance alleles that become fixed (Hedrick 2004). To our knowledge, there are only two well-documented cases of long-term (i.e. trans-specific) balanced polymorphisms at host defense genes in mammal, the MHC (Figueroa, Gunther, and Klein 1988; Edwards et al. 1997; Gutierrez-Espeleta et al. 2001) and the TRIM5α gene in Old-World monkeys (Newman et al. 2006), making *Oas1b* only the third known case.

Considering the strong signature for balancing selection at *Oas1b*, the presence of premature stop codons and other inactivating mutations in five wild-derived strains of mice seems surprising, as such mutational events are expected for genes under relaxed selection. In addition, the four sequences in our sample that are carrying the premature stop-codon in exon 4 (C57BL/6J, SKIVE, SF and DOT) differ at other positions suggesting that this mutation is not very recent. As susceptible alleles have not been eliminated rapidly from populations by natural selection, the fitness of susceptible mice might not be significantly lower than the fitness of resistant mice, casting doubts about the possible advantage of flavivirus resistance in modern populations. A possible explanation for this observation is that the selective pressure exerted on *Oas1b* has changed recently when mice colonized new geographical areas or new ecological niches. Resistant mice could have been favored in the original range of the species, which was much more limited geographically than its current range. In addition, when mice became commensal with humans, their ecology changed drastically and they might be protected from flavivirus infection in human habitat. Another possible explanation is that inactivating mutations are at such low frequencies in natural populations that they exist only in the heterozygous state. As resistance to flaviviral infection is dominant over susceptibility, selection might not be acting efficiently against susceptible alleles.

There are three main categories of mechanisms that can maintain allelic diversity for long periods of time: frequency dependent selection, heterozygote advantage and selection that varies in time and space. At this point, it is not possible to determine with certainty the mechanism responsible for the long-term maintenance of *Oas1b* alleles and more data about *Oas1b*'s allelic diversity in natural populations are needed. However, our data point to some type of frequency dependent mechanism. Under this model, the resistance allele that protects against the most common pathogenic genotypes will also be the most common one in the host population. In this situation, rare flaviviral genotypes that evade host-defense will be advantaged and increase in frequency. Because hosts carrying rare alleles have in turn the highest fitness, the frequency of the rare allele will increase and the frequency of the common allele will decrease leading to a dynamic polymorphism. The *dn/ds* ratios were plotted against their *ds* values, which is a proxy of the divergence time between alleles (Figure 12). Less divergent (younger) alleles have higher *dn/ds* than older ones. This suggests that balancing selection has favored rare alleles over more common ones, assuming that young alleles are, in general, at lower frequency in populations than older alleles. A similar relationship is also expected if nonsynonymous sites become saturated faster than synonymous sites. However, this seems unlikely because most values of *dn* are lower than 3% and saturation is not expected within this range.

Whatever the mechanism involved in the maintenance of *Oas1b* alleles, a functional difference between the *Oas1b* alleles under balancing selection is implicit. Balancing selection also implies some type of interactions, direct or indirect between *Oas1b* and flaviviruses. The two alleles that have been functionally characterized (the major and minor resistance alleles) differ in the specificity of their response to flaviviruses, the minor resistance allele protecting against a smaller diversity of flaviviral genotypes than the major resistance allele. The amino-acid polymorphisms responsible for these functional differences have yet to be mapped to specific regions of *Oas1b*, as these two alleles differ at a number of sites spread across the entire gene. Although *Oas1b* retained a number of functional features found in other oligoadenylate synthetases, it has lost its synthetase activity and seems to protect against flaviviruses by a still unknown RNase L-independent pathway (Scherbik et al. 2006). As the dsRNA-

binding motifs of *Oas1b* are located at the N-terminus of the protein, balancing selection is probably not caused by the interactions between flaviviral RNA and the *Oas1b* protein. The C-terminus of the *Oas1b* protein contains a domain with strong sequence similarity to an extra-cellular domain of interleukin receptors 3 and 5 (W. Ferguson and S. Boissinot, unpublished observation). This domain encompasses the CFK motif responsible for *Oas1b* tetramerization and plays an important role in protein-protein binding in the interleukin receptor family. This domain is in the immediate vicinity of a number of polymorphic sites including one of the polymorphisms shared between the different subspecies of *M. musculus* and between *M. musculus* and *M. spicilegus*. As this region has the potential to mediate protein-protein interactions, it is plausible that interactions between the *Oas1b* protein and a flaviviral protein, or a host protein which is itself under balancing selection, are responsible for the pattern of balancing selection at *Oas1b*, although this hypothesis needs to be tested experimentally.

ACKNOWLEDGEMENTS

The work was conducted in part with equipment from the Core Facilities for Imaging, Cellular and Molecular Biology at Queens College. We thank Else Fjerdingstad, George Jackman and Peter Novick for their helpful comments on the manuscript. This research was supported by NIH grant AI058968-01A1 and PSC-CUNY grant 66253-0035 to Stéphane Boissinot.

Chapter 2

CONTRASTED SIGNATURE OF SELECTION IN THE ANTI-VIRAL *Oas1* GENE FAMILY IN THE HOUSE MOUSE

William Ferguson [*, †], Nancy Okeke [*], Shira Dvora [*], Ann Duong [*], and Stephane Boissinot [*, †]

[*] Department of Biology, Queens College, the City University of New York, 65-30 Kissena Boulevard, Flushing, NY 11367, USA.

[†] Graduate School and University Center, the City University of New York, 365 Fifth Avenue, New York, NY 10016, USA.

Corresponding author:
Stéphane Boissinot
Department of Biology
Queens College, CUNY
65-30 Kissena Boulevard
Flushing, NY 11367-1597
Tel: 718 997 3437
Fax: 718 997 3445
Email: stephane.boissinot@qc.cuny.edu

Submitted: Journal of Molecular Evolution

ABSTRACT

The oligoadenylate synthetase (OAS) gene family is part of the innate response to viral infection. OAS proteins are activated by the presence of double stranded RNA and then synthesize oligoadenylates which are able to activate Rnase L, thus impeding the infection. There are eleven OAS loci in the mouse genome, seven of which arose after the split from primates but before the divergence from rat. *Oas1b* is a member of the mouse OAS family and is a potent anti-viral gene which specifically inhibits infection by flaviviruses, such as West Nile virus, dengue fever virus, and yellow fever virus. Recent work determined that balancing selection had retained functionally divergent *Oas1b* alleles for over 3 million years. Here, we analyzed two other members of the mouse *Oas1* gene family, *Oas1g* and *Oas1e*. Our results determined that alleles of *Oas1g* have been retained for between 1.5 and 2.8 million years, suggesting this locus is also evolving under balancing selection. In contrast, we found very little diversity at *Oas1e* in *M. musculus* suggesting that *Oas1e* is evolving under purifying selection.

INTRODUCTION

The expansion of a functional gene family can be a potent force to the evolution of an organism, whereby newly generated loci will either continue their functional redundancy or take on a novel role, thus evolving under different selective pressures (Prince VE and FB., 2002). Many extensive gene families are associated with the host defense system and harbor high levels of variation (Ferrer-Admetlla et al., 2008; Newman et al., 2006; Tan et al., 2005). The maintenance of variation between alleles is often referred to as balancing selection and can be the result of frequency-dependent selection, heterozygotic advantage, (Kojima, 1971a) or selection that varies in space and time (Hedrick, 1976). Examining the evolutionary history of these genes can reveal the functional domains and aid in determining the specific selective pressures.

The anti-viral oligoadenylate synthetase (OAS) gene family is dependent on the release of interferon-alpha and -beta for proper upregulation (Hovanessian AG and Justesen J, 2007). The OAS protein remains inactive until it comes into contact with double-stranded RNA (dsRNA), which can be found at relatively high levels as the virus replicates its genome (Samuel CE, 2002). The activated OAS protein then synthesizes oligoadenylates by forming a bond between the 2 and 5 carbon of two ATP molecules. These 2-5 oligoadenylates (2-5A) activate Rnase L, which degrades both viral and cellular RNA, thus preventing further viral infections. Human OAS is involved in the innate response to infection with HIV-1, hepatitis C, vaccinia virus, influenza, and reovirus infections.

The OAS gene family is ubiquitous in animals. Humans have a total of four OAS genes designated OAS1, OAS2, OAS3, and OASL (OAS-like), resulting from the duplication of 1, 2, or 3 enzymatic domains (Justesen J et al., 2000). In mouse, the OAS family is located on chromosome 5 consists of eleven loci. Several duplication events led to a total of eight *Oas1* loci, designated *Oas1a* through *Oas1h* (Figure 13a). Evolutionary analysis of this gene family has only been completed for *Oas1b* (Ferguson W et al., 2008), which determined that long-term balancing selection has been maintaining alleles, resulting in trans-specific polymorphism.

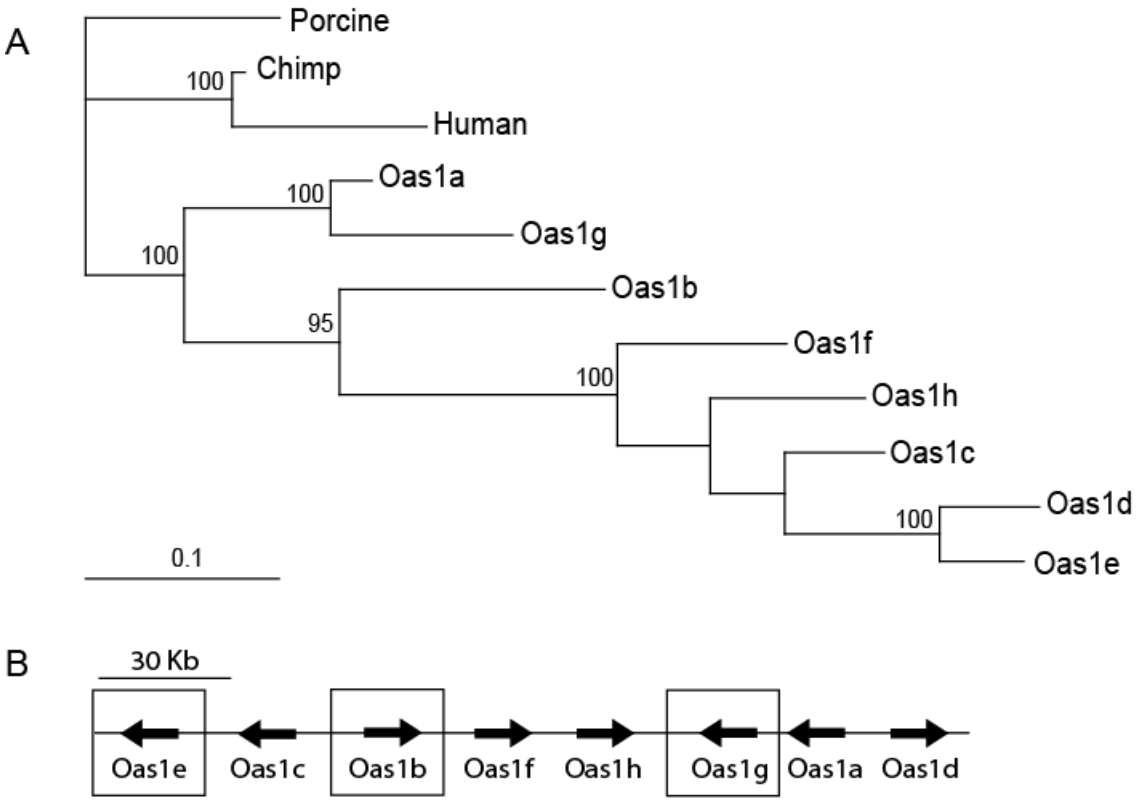

Figure 13: A) Maximum likelihood phylogeny of the *Oas1* gene family in mouse using porcine as the outgroup. Bootstrap values greater than 85 are indicated. B) Organization of the *Oas1* cluster on chromosome 5 of the mouse genome.

The eight copies of *Oas1* have remained relatively conserved; however their functions and expression patterns vary (Mashimo T, 2003). *Oas1a* and *Oas1g* have the ability to produce 2-5A's and are also broadly expressed in many tissue types (Kakuta S et al., 2002; Mashimo T, 2003). *Oas1b* has been shown to specifically prevent infection against flaviviruses, such as dengue fever virus, yellow fever virus and West Nile virus (Mashimo T, 2002; Perelygin AA et al., 2002). Yet the *Oas1b* protein is enzymatically inactive and operates in a Rnase L independent pathway which has yet to be characterized (Scherbik SV et al., 2006). It is unknown whether *Oas1b* interacts with other members of the *Oas1* family, however both *Oas1b* and *Oas1g* are highly upregulated during infection with West Nile virus (Scherbik SV et al., 2007)

Oas1e and *Oas1d* are very similar at the amino acid level (~84%) and are expressed during embryonic development as well as in some sex cells (Yan et al., 2005). Recent work has found that *Oas1d* can inhibit the function of *Oas1a in vivo* by blocking 2-5A production (Yan W, 2005). This same study found that mice with *Oas1d* knocked-out showed reduced fertility, with defects in ovulation and preimplantation. It is hypothesized that both *Oas1d* and *Oas1e* may aid in the survival of crucial sex cells or embryos during viral infection. In this scenario both *Oas1d* and *Oas1e* may work against either the activation of Rnase L or another unknown pathway. The functions of *Oas1c, Oas1f,* and *Oas1h* have yet to be clarified.

We studied the evolution of three members of the mouse *Oas1* gene family. Our original analyses focused on *Oas1b* and determined that balancing selection has maintained functionally divergent alleles for over 3 million years (Ferguson W et al., 2008). We further analyzed two loci which flank *Oas1b* to determine whether balancing selection acts on other *Oas1* loci. *Oas1g* is located ~52 kb downstream of *Oas1b* and has anti-viral properties which are similar to human OAS1, in its ability to bind to RNA and synthesize 2-5A's. Our data determined that alleles of *Oas1g* are also being maintained by balancing selection and are between 1.5 and 2.8 million years old. We further identified the RNA binding region of the protein as the likely site for selection, suggesting that viral RNA is the selective pressure acting on *Oas1g*. *Oas1e* is ~26 kb upstream of *Oas1b* and lacks the ability to synthesize 2-5A's or act in any known anti-

viral capacity. From our results we concluded that *Oas1e* is evolving under positive selection.

MATERIALS AND METHODS

Sampling – The house mouse, *Mus musculus*, originated in India less than one million years ago (Boursot P et al., 1996). Its radiation throughout Europe and Asia yielded three well defined subspecies, *M. m. musculus* from Eastern Europe to Northern China, *M. m. domesticus* in Western Europe, North Africa, and the Middle East; *M. m. castaneus* in South East Asia. Recent hybridizations in Asia between *M. m. musculus* and *M. m. castaneus* produce hybrid populations sometimes referred to as *Mus musculus molossinus*. (Yonekawa and Moriwaki, 1986). We obtained DNA for twelve wild-derived strains of *Mus musculus* from the "Conservatoire de la souris sauvage" (Universite Montpellier II, Montpellier, France), and eight others from the Jackson laboratory (Bar Harbor, ME). Multiple generations of inbreeding have made their genomes homozygous at most loci, facilitating haplotype determination. This study focuses on *M. m. musculus* and *M. m. domesticus*. We also obtained DNA from other species within the genus *Mus* to define the phylogenetic context of the genes analyzed. These samples included three European species, *Mus spretus*, *Mus macedonicus*, and *Mus spicilegus*, all of which diverged from the house mouse ~1.4 MYA. We also examined the Indian species *Mus famulus* (2.8 MYA) as a distant relative and possible outgroup. When *M. famulus* was not appropriate as an outgroup, sequence data from the rat genome was used.

Molecular analysis – All of the exons of *Oas1b, Oas1e,* and *Oas1g* were amplified by polymerase chain reaction (PCR). Intron 4 was also amplified for each gene, as well as intron 6 and a portion of intron 1 of *Oas1g*, intron 5 of *Oas1e*, and portions of all introns in *Oas1b* were also sequenced (*Oas1b* sequences have been previously described in Ferguson et al. 2008). PCR products were purified in our lab and then sequenced by Macrogen (Seoul, Korea). For neutral comparisons, two regions of the mouse genome

were selected at least 1 Mb from any known gene, one on chromosome 12 (64,811,904 to 64,813,152 – February 2006 assembly [http://genome.ucsc.edu]) and one on chromosome 19 (19,726,706 – 19,727,769). These regions were also amplified by PCR and sequenced by Macrogen. A list of primers used for amplification is available upon request to the corresponding author. DNA sequences have been deposited at the EMBL database under GenBank accession numbers AM887890-AM887932,

Data analysis – Sequences were aligned and manipulated using the BioEdit platform (Hall 1996). The Nei and Gojobori (1986) method implemented in MEGA3 (Kumar et al. 2004) was used to calculate the non-synonymous (ka) and synonymous (ks) distances. Phylogenetic analyses were performed using the minimum evolution method (MEGA3) and the maximum likelihood method (PHYML [Guidon et al. 2005]). Nucleotide diversity was estimated using pi (π), and Watterson's theta (θ). Deviation from neutrality was assessed by calculating Tajima's D. The Hudson-Kreitman-Aguade test (HKA test [Hudson et al. 1987]) was used to test for selection for each locus. This test compares the ratio of intraspecific polymorphism to interspecific divergence between a region of interest (*Oas1* genes) and a putatively neutral region of the mouse genome (from chromosome 12 or 19). If the difference between the ratios is statistically significant, then we can reject the null hypothesis of neutrality. We also used the McDonald and Kreitman test (MK) (1991) to test for significant deviations from neutrality. Under neutral conditions it is assumed the ratio of non-synonymous to synonymous fixed differences should be equal to the ratio of non-synonymous to synonymous polymorphisms within a species. The DnaSP 4.5 program was used to calculate population genetics parameters and to perform the HKA and MK tests.

RESULTS

We sequenced the coding region of the *Oas1b, Oas1g*, and *Oas1e* genes in 22 strains of *Mus musculus*. The highest levels of amino acid variation were found in *Oas1b* (28) and *Oas1g* (14) whereas *Oas1e* was not very variable (5) (Figure 14).

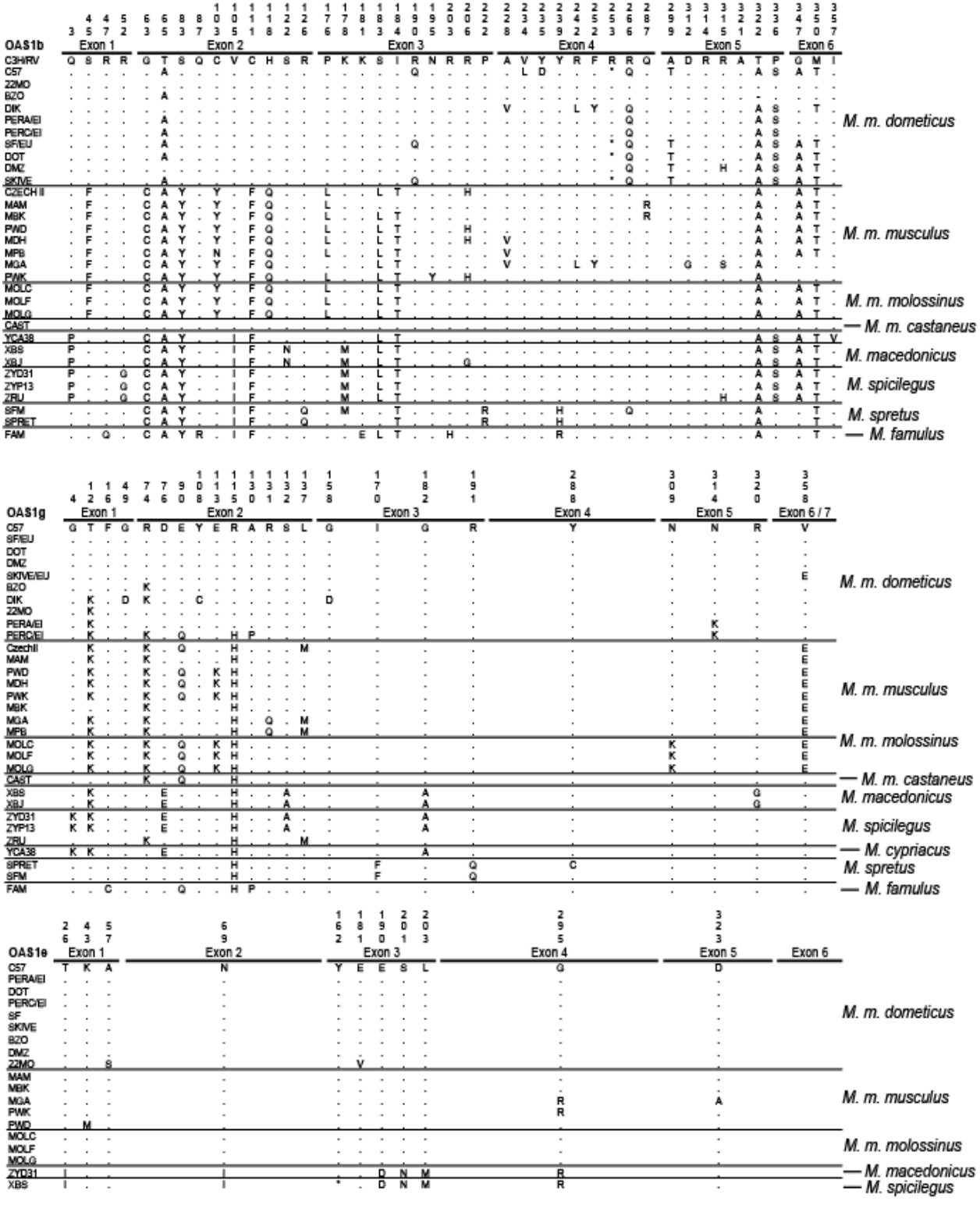

Figure 14: Variable amino acid sites for *Oas1b, Oas1g,* and *Oas1e* for all strains of *M. m. domesticus, M. m. musculus, M. macedonicus,* and *M. spicilegus* analyzed. Dots indicate an identical amino acid to the top line.

During the sequencing of *Oas1b, Oas1g,* and *Oas1e* we found no occurrences of heterozygosity, confirming that the primers amplified specifically the genes they were designed to amplify and not paralogous copies. Within *M. musculus* we found that 7 out of 28 polymorphic sites were shared between subspecies at the *Oas1b* locus. Comparisons with more distant species found 5 polymorphisms were shared between *Mus musculus* and either *Mus spicilegus* or *Mus spretus*. This indicates that at least 5 polymorphisms in *Oas1b* are older than the split between *Mus musculus* and other European species, which occurred about 1.4 million years ago. At the *Oas1g* locus we found a total of 14 variable amino acid polymorphisms in *Mus musculus* with 5 being shared between the different subspecies. We also found 4 polymorphisms shared between *Mus musculus* and *Mus spicilegus*. Of these polymorphic sites 3 out of the 4 occur in either exon 1 or exon 2. This indicates that at least 3 polymorphisms in *Oas1g* are also older than the split between *Mus musculus* and other European species. *Oas1e* showed the least amount of variable amino acid sites (5) within *Mus musculus* and none were shared between subspecies.

To determine the age of the allele variation of these three loci we compared the synonymous divergence (ks) of the *Oas1* genes in *M. musculus* alleles to the divergence of 183 autosomal genes for which a *musculus* and *domesticus* allele were available in GenBank (Figure 15a). The *Oas1b* alleles were more divergent at synonymous sites than most autosomal genes, with levels as high as 4%. Only a few genes showed such high levels of ks, such as members of the MHC and immunoglobulin gene families. Ks values were also elevated in *Oas1g*, and were higher than most autosomal genes, with some alleles differing by 2.25%. Surprisingly high levels of ks were also found in *Oas1e* (as high as 2.35%), which was unexpected due to the relatively low level of amino acid diversity. These results suggest that the age of alleles for all three genes is at the upper limit of genes in the mouse genome.

The age of alleles for *Oas1e, Oas1g,* and *Oas1b* were further assessed by phylogenetic analysis of intron 4 (Figure 16). Intron 4 of *Oas1e* followed the known phylogenetic relationships among species; however the *M. m. domesticus* and *M. m. musculus* lineages are much deeper than expected under neutrality. The clear separation of *M. macedonicus* from *M. musculus* in *Oas1e* and the deep lineages within

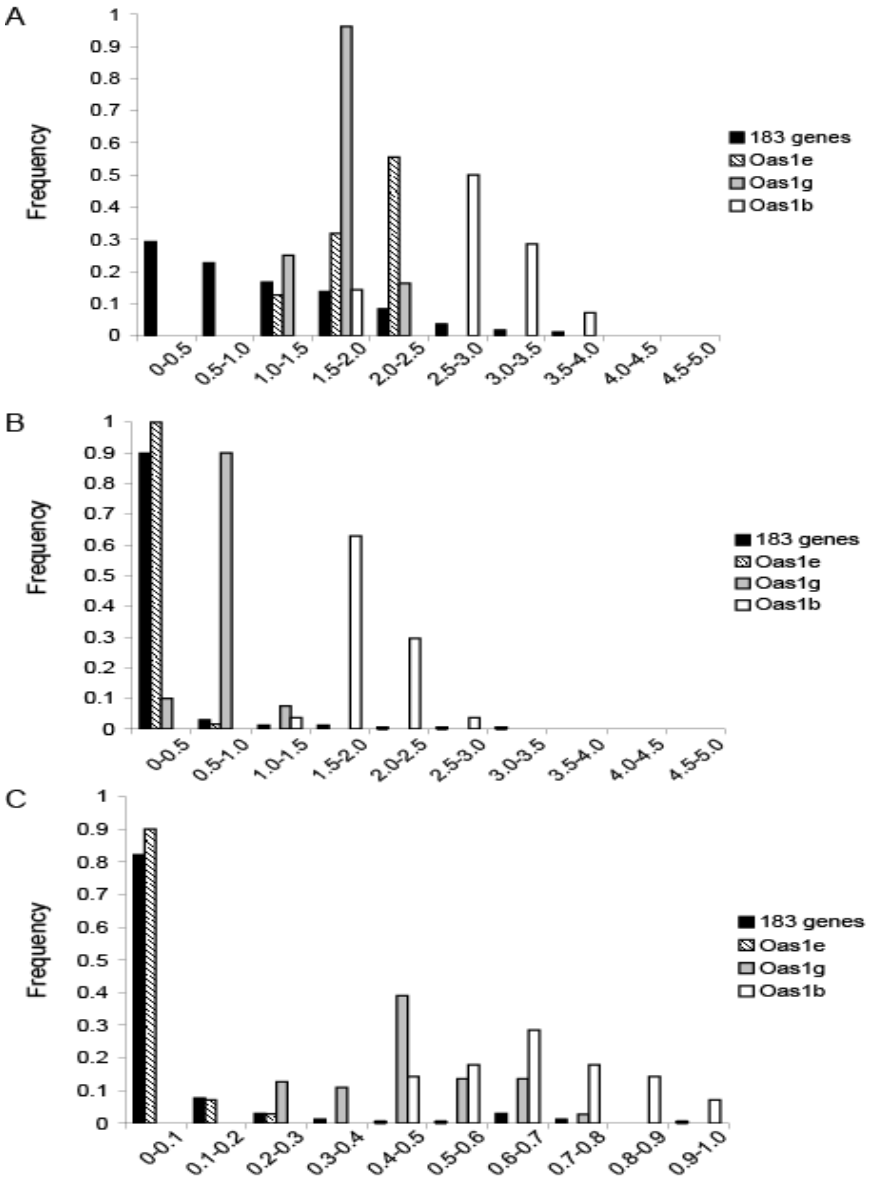

Figure 15: A) The Ks values were calculated for 183 genes collected from GenBank for *Mus musculus*. Ks pairwise analyses were performed on *Oas1e, Oas1g,* and *Oas1g*. B) Ka values for the 183 genes, *Oas1e, Oas1g,* and *Oas1g*. C) Ka/Ks values for the 183 genes, *Oas1e, Oas1g,* and *Oas1g*.

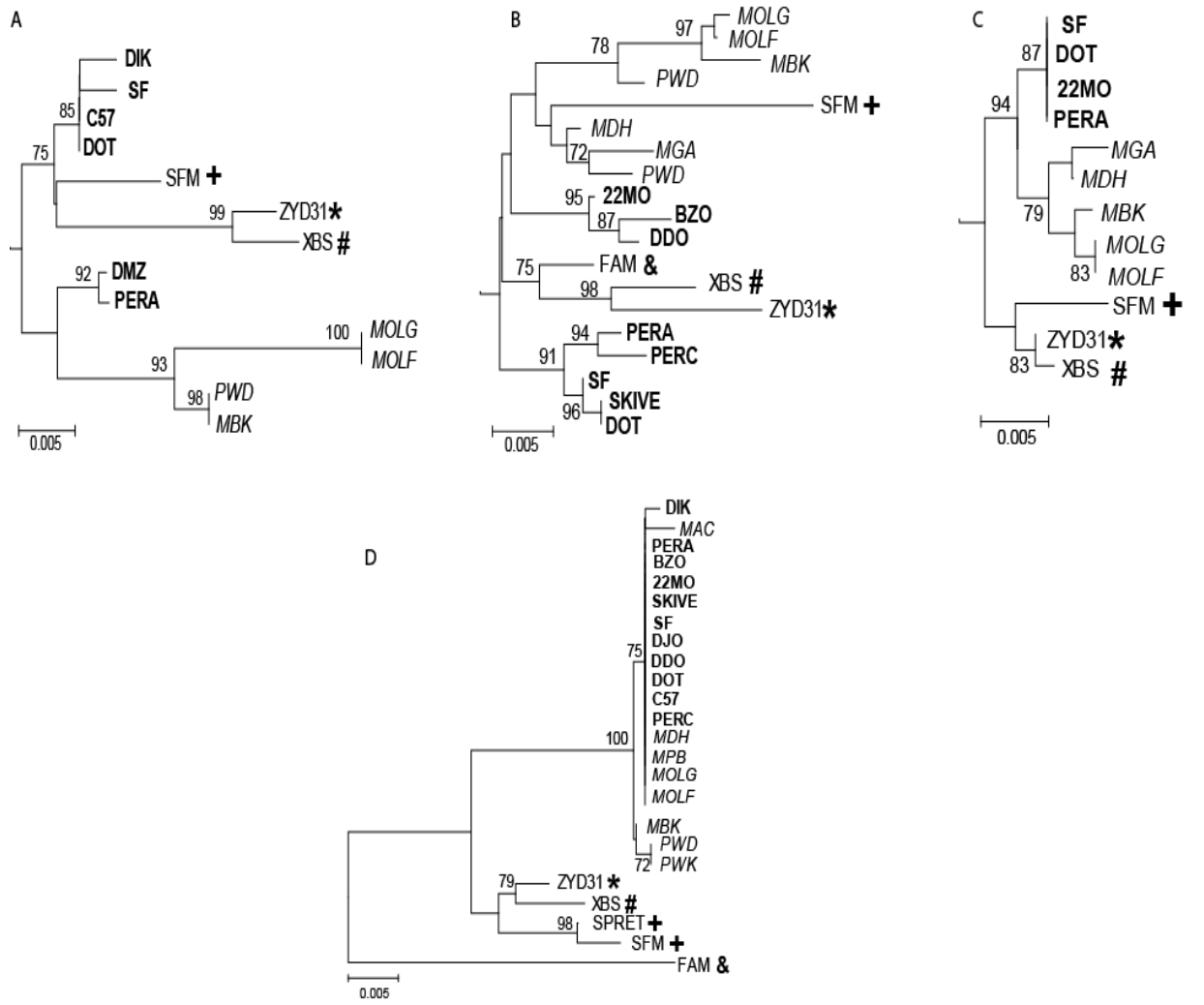

Figure 16: *M. m. musculus* strains are shown in italic, *M. m. domesticus* in bold, *M. spicilegus* are indicated with an asterisk (*), *M. macedonicus* with a number sign (#), *M. spretus* with a plus sign (+), and *M. famulus* with &. Bootstrap values of 75 and over are shown. D) Phylogenetic analysis of putatively neutral region on chromosome 12. A-C) Phylogenetic analysis of intron 4 for *Oas1e, Oas1g,* and *Oas1b*.

M. musculus suggests that the alleles are likely between 0.5 and 1.4 million years old. Intron 4 of *Oas1g* contains two deep lineages within *M. musculus*, with one of the lineages containing strains of *M. macedonicus* and *M. spretus*. The presence of these species within *M. musculus* clades and the fact that *M. famulus* falls out of these lineages suggests that the age of *Oas1g* alleles are between 1.4 and 2.8 million years old. The deepest lineages were found in *Oas1b*, where *M. famulus* grouped with *M. musculus* suggesting the alleles are over 2.8 million years old.

The nucleotidic diversity in intron 4 of *Oas1e* in both *M. m. musculus* and *M. m. domesticus* was only slightly higher than found in the putatively neutral regions of the mouse genome (Table 4). Nucleotidic diversity (π) in intron 4 of *Oas1g* was 2 to 5 times higher than the putatively neutral region, and 5 to 10 times higher in intron 4 of *Oas1b*. Values for Tajima D were negative in intron 4 of *Oas1e* and *Oas1g* for *M. m. domesticus*, with all other Tajima D values positive. Positive Tajima D values indicate that either balancing selection and/or a decrease in population size has lowered the levels of low and high frequency polymorphisms. None of the Tajima D values were significant and therefore there regions do not deviate from neutrality under this test. The HKA test was performed using sections of chromosome 12 and 19 as putatively neutral regions (see Materials and Methods). The null hypothesis of neutrality was rejected in *Oas1g* for *M. m. domesticus* and for both *M. m. domesticus* and *M. m. musculus* in *Oas1b* using the neutral region of chromosome 19. The values were significant for both subspecies in *Oas1g* and *Oas1b* when the region of chromosome 12 was used. A small piece of intron 1 was also tested for *Oas1g* and yielded significant HKA values for both *M. m. domesticus* and *M. m. musculus* using chromosome 19 (*M. m. domesticus P* is 0.0039 and *M. m. musculus P* is 0.0391) and chromosome 12 (*M. m. domesticus P* is 0.0001 and *M. m. musculus P* is 0.0053).

The level of non-synonymous divergence (ka) was calculated for each gene and *Oas1b* was found to have the highest level, with *Oas1e* having the lowest ka (Table 4). The average non-synonymous divergence between *M. m. musculus* and *M. m. domesticus* in *Oas1b* was 1.88% (Figure 15b), which is a higher ka than all but 5 genes analyzed from GenBank. The overall ka in *Oas1g* (0.67%) was lower than *Oas1b*;

		Exons						Intron 4							
												HKA Test - 19		HKA Test - 12	
Gene	**Species**	**n**	**Π(%)**	**Θ(%)**	**Ka(%)**	**Ks(%)**	**Ka/Ks**	**n**	**Π(%)**	**Θ(%)**	**Tajima D**	X^2	***P* value**	X^2	***P* value**
Oas1e	*M.m.domesticus*	8	0.213	0.288	0.0654	0.5951	0.1099	6	0.0450	0.0590	-0.9330	0.204	0.6515	0.220	0.6387
	M.m.musculus	8	0.282	0.264	0.1214	0.7253	0.1674	5	0.4580	0.3880	1.2410	0.400	0.5269	3.136	0.0766
Oas1g	*M.m.domesticus*	10	0.418	0.578	0.3709	0.5388	0.6884	7	1.0060	1.1570	-0.7048	4.285	**0.0384**	9.471	**0.0021**
	M.m.musculus	10	0.363	0.310	0.3121	0.4938	0.6320	8	0.8810	0.7930	0.5379	2.095	0.1478	3.885	**0.0487**
Oas1b	*M.m.domesticus*	10	0.631	0.530	0.5951	0.9294	0.6403	7	1.3820	1.3370	0.1876	7.177	**0.0074**	15.658	**0.0001**
	M.m.musculus	10	0.447	0.553	0.4738	0.3218	1.4723	8	1.3840	1.3150	0.2736	5.391	**0.0202**	9.737	**0.0018**
								Neutral Region							
19	*M.m.domesticus*	-	-	-	-	-	-	8	0.0530	0.0810	-1.3101	-	-	1.117	0.2905
	M.m.musculus	-	-	-	-	-	-	8	0.0600	0.0860	-1.2372	-	-	0.011	0.9161
12	*M.m.domesticus*	-	-	-	-	-	-	8	0.0630	0.0450	1.1665	1.117	0.2905	-	-
	M.m.musculus	-	-	-	-	-	-	8	0.1910	0.1930	-0.0398	0.011	0.9161	-	-

Table 4: Values for pi, theta, Ka, Ks and Tajima D calculated using DnaSP 4.50.2 and Mega3. HKA values were determined using the putatively neutral regions of chromosome 12 and 19 in the mouse genome. Significant values are shown in bold.

however it was still much higher than the majority of the 183 autosomal genes (Figure 15b). *Oas1e* had a very low ka at 0.10%, far below both *Oas1g* and *Oas1b*. High ka/ks ratios indicate that selection is favoring amino acid changes and low ka/ks ratios are presumably due to selection against amino acid changes (purifying selection) (Garrigan and Hedrick). The ka/ks ratio for most genes is extremely low (<0.1 [Gibbs et al. 2004]) and this was confirmed for the majority of the autosomal genes analyzed (Figure 15c). The ka/ks values for *Oas1e* were also very low (<0.1) for most pairwise comparisons, indicating that purifying selection is acting on this gene. These low ka/ks values are in part due to the high level of synonymous mutations found in the *Oas1e* alleles, demonstrating that strong purifying selection has retained this major allele for a long period of *M. musculus's* evolution. The ka/ks ratio was much higher in *Oas1b* (highest pairwise analysis of 0.976) and *Oas1g* (highest pairwise analysis of 0.730) than found in most of the183 autosomal genes analyzed. Other genes approaching the ka/ks level of *Oas1b* and *Oas1g* from the 183 analyzed include the kill T-cell receptor Ly49H (0.997), the MHC class II locus Mb1 (0.636), and the MHC class II antigen A (0.567). The predominance of these high ratios at *Oas1b* and *Oas1g* confirms that selection is favoring non-synonymous mutations and that alleles for both genes are likely maintained through balancing selection.

If *Oas1b* and *Oas1g* are being under balancing selection, then an excess of non-synonymous polymorphisms is expected within species. The largest excess of non-synonymous mutations was found in *M. m. musculus* for both *Oas1b* (80.0% of all mutations are non-synonymous) and *Oas1g* (66.7% of all mutations are non-synonymous), with lower levels in *M. m. domesticus* for *Oas1b* and *Oas1g*, 50.0% and 61.1% respectively (Table 5). The MK test determined that the ratio of non-synonymous to synonymous fixed differences significantly differs from the ratio of non-synonymous to synonymous polymorphisms in *M. m. musculus* for both *Oas1b* and *Oas1g* and for *M. m. musculus* in *Oas1g*. The failure to detect selection in *M. m. domesticus* for *Oas1b* is likely a result of the conservative nature of the MK test. The excess of non-synonymous mutations in both *Oas1b* and *Oas1g* are consistent with signals of balancing selection. To be certain that *Oas1b* and *Oas1g* are evolving independently, we sequenced a region 10 kb downstream of *Oas1b,* which lies between the two genes. The nucleotidic

diversity for this region was 0.372% in *M. m. domesticus* and 0.183% in *M. m. musculus* and Tajima D values were not significant. The failure to reject neutrality confirms that both *Oas1g* and *Oas1b* are independently evolving under balancing selection.

The location of non-synonymous polymorphisms can indicate which region of a gene is evolving under positive selection. Figure 17 shows the introns (solid lines) and exons (boxes) for *Oas1e, Oas1g,* and *Oas1b*. Variation in *Oas1b* seems near ubiquitous in all exons with the highest value of π in exon 5 (2.45%) and exon 2 (2.00%). This is similar to *Oas1g* in that exon 5 and exon 2 show the highest values of π, 1.05% and 1.12% respectively, ignoring exon 7 due to its extremely short length. A more interesting aspect of *Oas1g* is an average ka/ks value above 1.0 in exon 2. Exons

Table 2.
MK Neutrality Tests

	Oas1b			*Oas1g*			*Oas1e*		
	Synonymous	Non-synonymous	*P* Value	Synonymous	Non-synonymous	*P* Value	Synonymous	Non-synonymous	*P* Value
Mus musculus musclulus									
Polymorphic changes	4	16	0.01201	3	6	0.03892	5	3	0.42699
Fixed differences	15	12		10	3		9	12	
Mus musculus domesticus									
Polymorphic changes	15	15	0.41571	7	11	0.01599	6	2	0.10865
Fixed differences	10	16		7	1		8	13	

Table 5: Results of the MK test for both *Oas1b, Oas1g*, and *Oas1e* using *M. famulus* for fixed difference analyses.

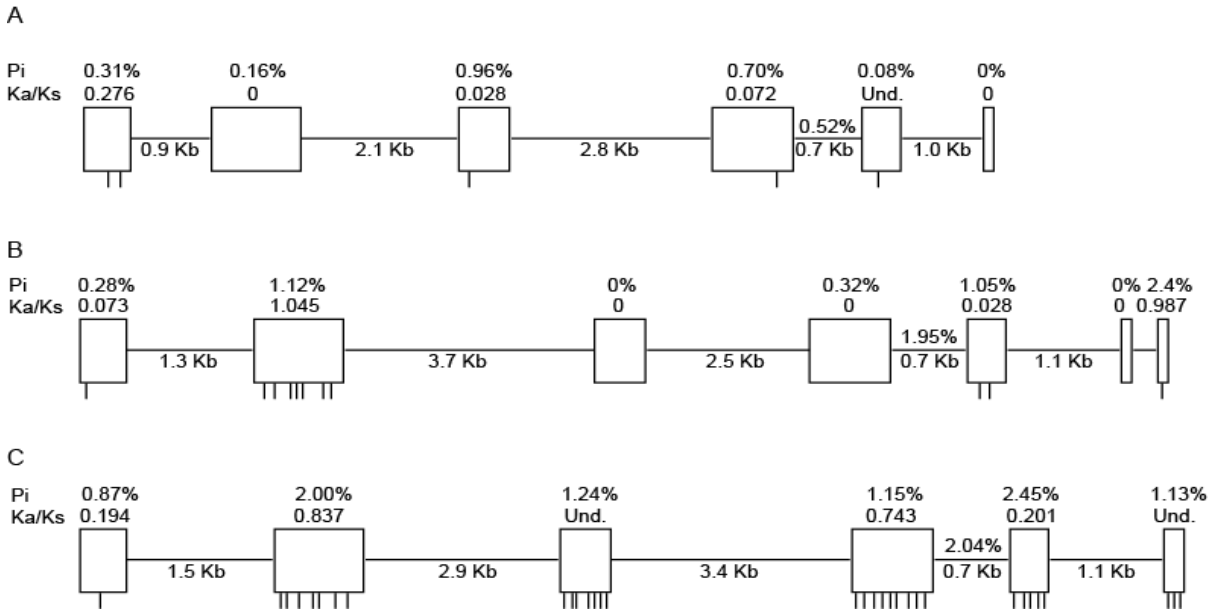

Figure 17: Boxes represent the exons of *Oas1e*, *Oas1g,* and *Oas1b* (A-C) with introns drawn between them. Values for π and ka/ks are given for each exon and values for π are indicated for intron 4 for each all three genes. Tick marks indicate non-synonymous mutations. Distances for each intron is given in kilobase pairs. Exons and introns are not drawn to the same scale.

of *Oas1b* only approach values of 1.0, other than exon 3 which has no synonymous mutations. These data reveal that signals for selection appear in the 5' region of *Oas1g* and in the 3' region of *Oas1g*.

DISCUSSION

Our results found distinct signature of selection within the *Oas1* gene family of the house mouse (*Mus musculus*). The conclusion that balancing selection was acting on both the *Oas1b* and *Oas1g* genes was determined using a variety of different analyses. The nucleotide diversity in intron 4 of *Oas1b* and *Oas1g* is an order of magnitude higher than the diversity in putatively neutral regions of the mouse genome (the regions on chromosome 12 and 19) and these high levels of diversity were substantiated by significant deviations from neutrality. Phylogenetic analysis of intron 4 revealed deep lineages when compared to other species of *Mus* for both genes, allowing us to approximate the age of alleles for *Oas1b* and *Oas1g*, ~2.8 MYO and ~1.4 MYO respectively. Our conclusions were further confirmed by ka/ks values using the coding regions of *Oas1b* and *Oas1g*, which where much higher than most of the 183 other mouse genes investigated and have similar levels to other genes which have been shown to contain ancient alleles. We further evaluated the coding regions using the MK test and found a significant deviation from neutrality in three out of the four MK tests for both genes.

The retention of alleles for long periods of evolutionary time requires strong selective pressures which remain constant within a population. We hypothesize that interaction with either flavivirus proteins or RNA is the most likely selective pressure acting on *Oas1b*, due to its very specific anti-viral nature. *Oas1g* is also an anti-viral gene, however it lacks the viral specificity of *Oas1b*, leaving a variety of viral agents as possible selective pressures. Signals for balancing selection were significant for both *Oas1b* and *Oas1g*; however the polymorphic patters are not identical. Analysis of the *Oas1e* gene found signs of purifying selection, in which one major allele dominates the

population. It is plausible that *Oas1e* may be an important developmental gene, as suggested by its expression patterns, which would explain its conservation.

The signals for balancing selection in *Oas1g* appear in the 5' region of the protein, within the vicinity of the RNA binding domain. All of the 3 shared polymorphisms in exons 1 and 2 of *M. musculus* involve positively charged amino acids (K, R, and H). Positively charged amino acids commonly appear in nucleic acid binding regions of proteins, due to the inherently negative charge of RNA and DNA. The requirement of RNA binding for proper activation of *Oas1g* along with the evolutionary data presented here leads to the conclusion that viral RNA is likely acting as a selective pressure on *Oas1g*. Our lab has recently identified similar evolutionary patterns in chimpanzee OAS1 and further determined that differences in the RNA binding domain leads to varying abilities to bind to viral RNA (in press). In contrast to *Oas1g*, signals for selection in *Oas1b* appear in the 3' region of the gene. Little is known about the functional importance of exons 5 and 6 of *Oas1b*, however it does contain a tetramerization motif (CFK) which is required for enzymatic activity (Ghosh A et al., 1997b). Our lab has identified a region in exon 5 and 6 of *Oas1b* (amino acid 328 to 342) that is homologous to a region in the interleukin receptors-3 and -5. This region of the interleukin receptor is known to create a hydrophobic core which is involved in protein-protein binding (ref). There is also a putative transmembrane domain in exon 6 of *Oas1b* (amino acid 348 to 320).

Although the specific selective pressures remain unclear, our results confirm that balancing selection has been acting on *Oas1b* for over 2.8 million years and *Oas1g* for over 1.4 million years. Our results have eliminated the possibility of gene linkage and resolved the fact that *Oas1b* and *Oas1g* are evolving separately from each other, thus likely having different selective pressures. Within the same gene family we determined that purifying selection had fixed a major allele of *Oas1e* in *Mus musculus*. These contrasted signatures of selection between genes which are highly similar in sequence identity and all closely localized with 100 Kb of each other is very rare. Further analysis is required to determine the types of selection on the other *Oas1* genes and what selective pressures are driving their evolution.

Chapter 3

LONG-TERM BALANCING SELECTION AT THE ANTI-VIRAL GENE OAS1 MAINTAINS FUNCTIONALLY DIFFERENT ALLELES IN CHIMPANZEES

William Ferguson [*, †], Shira Dvora [*], Anne C. Stone [‡] and Stéphane Boissinot [*, †]

[*] Department of Biology, Queens College, the City University of New York, 65-30 Kissena Boulevard, Flushing, NY 11367, USA.
[†] Graduate School and University Center, the City University of New York, 365 Fifth Avenue, New York, NY 10016, USA.
[‡] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ , USA.

Corresponding author:
Stéphane Boissinot
Department of Biology
Queens College, CUNY
65-30 Kissena Boulevard
Flushing, NY 11367-1597
Tel: 718 997 3437
Fax: 718 997 3445
Email: stephane.boissinot@qc.cuny.edu

Key word: Balancing selection, OAS1, chimpanzee

Running head: Balancing selection at OAS1

Abbreviations: OAS, Oligoadenylate Synthetase; MHC, Major Histocompatibility Complex; HIV, Human Immunodeficiency Virus; HLA, Human Leucocyte Antigen; SNP, Single Nucleotide Polymorphism; SARS, Severe Acute Respiratory Syndrome; PCR, Polymerase Chain Reaction; HKA, Hudson-Kreitman-Aguade.

ABSTRACT

Oligoadenylate synthetases (OAS) are interferon-inducible enzymes that participate in the first line of defense against a wide range of viral infection. Upon activation by viral dsRNA, OAS synthesizes (2-5A) oligoadenylates which activate Rnase L, leading to the non-specific degradation of cellular and viral RNA. Association studies in humans suggest that variation at one of the OAS genes, OAS1, could be influencing host susceptibility to viral infection. We assessed the diversity of OAS1 in hominoid primates with a focus on chimpanzees. We found that two very divergent groups of alleles have been maintained in common chimpanzees by long-term balancing selection since before the split between chimpanzees, humans and gorillas. These two classes of alleles differ by a large number of amino acids, including several amino acids putatively involved in RNA binding. We demonstrate experimentally that the two chimpanzee allelic classes have different RNA-binding affinities, resulting in different enzymatic activity. It is therefore very likely that variation at the OAS1 locus affects the innate immune response of individual chimpanzees to viral infection. Our data strongly suggests that interactions between viral RNA and OAS1 are responsible for the maintenance of ancestral polymorphisms at this locus for more than 9 million years.

INTRODUCTION

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

Oligoadenylate synthetases (OAS) are interferon inducible enzymes that participate in the first line of defense against a wide range of viral infections (for reviews on OAS, see (Goodbourn et al., 2000; Hovanessian and Justesen, 2007)). Upon infection, OAS genes are up-regulated by interferon and are activated by a double strand (ds) RNA mediated process. Production of dsRNA by a virus is believed to be the trigger of oligoadenylate synthesis in infected cells. Activated OAS proteins synthesize, from ATP, 2',5'-linked oligoadenylates. Oligoadenylate molecules activate Rnase L which, in turn, degrades cellular and viral RNA thus preventing viral replication. In humans, the OAS gene family has three enzymatically active copies, OAS1, OAS2 and OAS3, resulting from duplication events that occurred before the origin of mammals

(Kumar S et al., 2000). OAS1 is up-regulated during the early stage of infection by a wide range of viruses and has been shown to play an anti-viral role during infection with encephalomyocardiatis virus, vaccinia virus, respiratory syncytial virus and HIV-1 (Behera et al., 2002; Diaz-Guerra et al., 1997; Gribaudo et al., 1991; Schroder et al., 1994). Because OAS proteins interact directly with viral RNA's, it is plausible that changes in the OAS proteins can affect the protection conferred by the OAS pathway. Although the interactions between OAS proteins and dsRNA's seem relatively unspecific, it has been recently suggested that polymorphisms at the OAS1 gene can affect host susceptibility to specific viral infection. Two studies have reported an association between single nucleotide polymorphisms (SNP) in OAS1 and the outcome of infection with hepatitis C and SARS (Hamano, 2005; Knapp S et al., 2003). In another study, a SNP at a splice acceptor site in the OAS1 gene was shown to be responsible for variations of the enzymatic activity of OAS1, thus demonstrating that a polymorphism in the OAS1 gene has a direct phenotypic effect (Bonnevie-Nielsen et al., 2005). However, the significance of these studies is difficult to assess without a better understanding of the evolutionary history of the OAS1 gene in human and related species.

Because OAS1 plays a major role in anti-viral resistance, we decided to examine its evolution and diversity in hominoid primates. OAS1 was sequenced in all species of hominoid primates with a focus on the common chimpanzee, for which individuals from 3 of the 4 sub-species were analyzed. We found that two divergent allelic lineages have been maintained in the common chimpanzee since before the split between chimp, gorilla and human, 9 Million years ago. We determined that this pattern of variation results from the action of long-term balancing selection. We also demonstrated experimentally that these two classes of alleles differ in their RNA binding affinity, suggesting that direct interactions between viral RNA and OAS1 protein drive the evolution of this gene.

MATERIALS AND METHODS

Sampling –We obtained DNA from 37 common chimpanzees (*Pan troglodytes*), including representative of the Central African subspecies, *P. t. troglodytes* (7

individuals), the Eastern African subspecies, *P. t. schweinfurthi* (3 individuals) and the Western African subspecies *P. t. verus* (27 Individuals). When necessary, the sub-specific origin of individuals was confirmed by sequencing of the mitochondrial DNA control region (data not shown). A sample of 15 humans from various ethnic origins was purchased from the Coriell Institute for Medical Research. This sample includes 3 individuals from Africa, 6 from Europe and 6 from Asia. In addition, we analyzed 8 bonobos (*Pan paniscus*), 3 gorillas (*Gorilla gorilla*), 2 orangutans (*Pongo pygmaeus*) and one Rhesus macaque (*Macaca mulatta*). The list of individuals and their origin is presented on Table 6.

Sequencing and haplotype determination– We first verified that OAS1 exists as single-copy by Real-Time PCR using the SYBR Green I chemistry (Roche). Real-Time amplifications were performed on two different sections of the OAS1 gene and compared among hominoid species. The coding sequence of the OAS1 gene was obtained for all individuals. Each exon was amplified independently by Polymerase Chain Reaction (PCR). The PCR products were treated with Shrimp alkaline phosphatase and exonuclease I, followed by phenol/chloroform extraction and ethanol precipitation. The PCR products were directly sequenced using the Big-Dye terminator chemistry by the company Macrogen (Seoul, Korea). Haplotypes were inferred by the method of Stephens et al (Stephens et al., 2001) implemented in the program PHASE. Individuals with the lowest probability of correct haplotype call were selected for further experimental haplotyping. Experimental haplotyping was accomplished by amplifying long fragments that encompass polymorphisms for which the linkage is unclear. Individual clones were then sequenced allowing the determination of the chromosomal phase by comparison with the diploid sequence. The molecular information obtained for the individuals with the lowest PHASE call was then incorporated into the PHASE analysis and an additional round of statistical inference was performed. This procedure was repeated until 95% of the individuals had a phase call with a confidence interval higher than 95%. We also amplified and sequenced partial fragments of introns 1, 2, 3, 4 and 5 of individuals that were homozygous at the OAS1 locus. To assess the level of variation at the OAS1 gene, we needed to compare it with the level of variation in a

neutral region of the chimpanzee genome. We selected an autosomal segment located at least 1Mb from any known gene (on chromosome 4, from nucleotides 129997604 to 129998795 on the February 2006 assembly of the chimpanzee genome at http://genome.ucsc.edu). This fragment was amplified by PCR in each chimpanzee and sequenced. The list of primers used for amplification is available on request from the corresponding author. Sequences are available in the EMBL database under accession number XXXXX – XXXXX (submitted).

Data analysis – Sequences were aligned and manipulated using the BioEdit platform (Hall, 1999). Synonymous (ds) and non-synonymous (dn) distances, as well as their ratio ω, were estimated by maximum-likelihood using the model of Yang and Nielsen (Yang and Nielsen, 2000) implemented in PAML (Yang, 2000). PAML was also used to assess the effect of positive selection. This was done by comparing maximum likelihood models that allow for positive selection with models that do not. In addition, specific codons under positive selection were identified using the NSsites model from the codonml program in PAML. Phylogenetic analyses were performed using the neighbor joining method (as implemented in MEGA3) and the maximum likelihood method (using PHYML (Guindon et al., 2005)). The nucleotide diversity was estimated using the parameters π and Watterson's θ. Tajima's D (Tajima, 1989) was calculated to assess the effect of selection on polymorphisms. We also tested for selection using the Hudson-Kreitman-Aguade test (HKA test; (Hudson et al., 1987)). This test is based on the assumption that, under neutrality, polymorphism and divergence should be the same across the genome. The test compares the ratio of polymorphism to divergence between a gene of interest and a neutral region of the genome. If the difference between the two ratios is significant using a goodness-of-fit test, we can reject the hypothesis of neutrality. As putative neutral regions we used a non-coding fragment of chromosome 4 (see above) as well as 10 of the 50 non-coding segments analyzed by Yu *et al* (Yu et al., 2003). Population genetics calculations were performed using the DnaSP program (Rozas et al., 2005).

Cloning and expression of OAS1 protein – Chimpanzee cell cultures were purchased from the Coriell repository. The two cultures we obtained were from a bonobo (*Pan paniscus*) and from a Western African chimpanzee (*P. troglodytes verus*). The pygmy chimp culture comes from an individual carrying an OAS1 allele which is 99.5% identical at the amino-acid level to class II haplotypes (Figure 18 and 19a) and the Western African chimpanzee carry an haplotype of class I. Cultures were propagated using Dulbecco's modified eagles medium (MDEM) with specific percentages of heat inactivated fetal bovine serum (FBS). The cells were grown until confluent and then induced overnight using recombinant interferon-α. Total RNA was extracted using Trizol, followed by phenol/chloroform extractions and ethanol precipitations. The RNA was treated with DNase and its integrity was checked on a 2% agarose gel stained with ethidium bromide. cDNA was generated by reverse transcription of the total RNA with MMLV reverse transcriptase and then using DNA polymerase for the second strand synthesis. The most common isoform of OAS1 (isoform p42) was amplified by PCR and purified. The OAS1 PCR products were ligated into the pET-41a expression vector (Novagen). The ligated vectors were cloned into Rosetta competent cells (Novagen). Expression of the GST-OAS1 fusion protein was activated by the addition of 2 mM of IPTG overnight at 25°C. Fusion proteins were purified using the MagneGST kit (Promega) and verified through western blot analysis using anti-GST (Bethyl) and anti-Rabbit (Promega) antibodies. Fusion proteins were cleaved using enterokinase (Novagen) and the GST fraction was removed using magnetic anti-GST beads (Promega). Presence of pure OAS1 was verified through PAGE analysis using coommassie blue and western analysis with anti-OAS1 antibody (Abgent).

Functional Assays – We generated RNA's for different regions of HIV-1 (a retrovirus), SV-40 (a dsDNA virus) and the Modoc virus (a positive strand RNA virus). The genomes of the HIV-1 and SV-40 viruses were obtained as clones from AddGene. The Modoc virus was obtained from ATCC and grown on Vero cells. Modoc RNA was extracted using Trizol and cDNA was synthesized by reverse transcription with MMLV reverse transcriptase. Individual genes of the three viruses were amplified by PCR. Viral genes were cloned into expression vectors (pEF6/V5-His-TOPO) and transcribed into

RNA using RNA polymerase I for 2 hours at 42°C. Viral RNA's were labeled with the fluorescent Cy5 dye using the Mirus kit (Invitrogen). The RNA was purified by ethanol precipitation and then incubated with purified protein in RNA-protein binding buffer, with or without polyinosinic:polycytidylic acid (poly I:C) as competitor. Five microliters of the mixture were then run on a 6% PAGE gel in TBE and analyzed using the Storm 860 imager. The enzymatic activity of each OAS1 variant was assayed by incubating purified OAS1 protein with unlabeled viral RNA's (or poly I:C) as well as fluorescently labeled rATP (Jena Bioscience) for 20 to 48 hours at 30˚C in the activity assay buffer (20mM HEPES-KOH pH7.5, 50mM KCL, 25mM $Mg(OAC)_2$, and 7mM β-ME). After incubation, the mixtures were run on a 20% PAGE gel in TBE and analyzed on a Storm 860 imager to quantify the amount of oligoadenylate generated.

RESULTS

We began our investigation of the OAS1 gene by verifying that this gene is single-copy in hominoids. We first performed several BLAST and BLAT searches of the human, chimpanzee, orangutan and macaque genomes. These searches failed to find any novel OAS1 copies in any of these species. Next, we performed quantitative real-time PCR on DNA samples from several chimpanzees and humans, with the primers used to amplify and sequence the OAS1 gene. These experiments did not reveal any additional OAS1 copies in either species (see supplementary materials). Therefore, we are confident that the OAS1 gene exists as a single copy in hominoids.

The sequencing of the OAS1 coding regions revealed an extremely high level of polymorphism in common chimpanzees (sample listed on Table 6). We identified 29 single nucleotide polymorphisms (SNP) at 28 positions in chimps. Twenty of those SNP's result in changes at the amino-acid level. As the OAS1 gene is 335 amino-acids long, we estimated that 6% of all amino acid positions are variable. In contrast, the human and bonobo samples did not show a similarly high level of polymorphism, with only 6 and 2 SNP's respectively. Combinations of polymorphisms result into 13 haplotypes in chimpanzees that can be classified into two main classes, named class I and II (Figure 18). Class I haplotypes are found in all three subspecies of chimpanzees

whereas class II haplotypes are nearly absent from the Western chimpanzees sample (*P. t. verus*). Our sample contained homozygous individuals for each haplotypic class and heterozygous individuals, confirming that OAS1 is a single copy gene. The two classes differ by 17 fixed mutations, translating into twelve amino-acid differences. We calculated the ratio ω between these classes and we found that it was significantly higher than 1 (mean estimate of ω=1.82). A value of ω higher than 1 indicates that some form of selection in favor of amino-acid changes (i.e. positive selection) is acting on a gene and is usually taken as evidence of adaptive evolution. The mean value of ω within each class was also relatively high (ω=0.60 and ω=0.90, within class II and class I, respectively). Although those values are lower than 1, they are still higher than the mean estimates of ω calculated for more than 10,000 orthologues between human and chimpanzee (ω=0.23) and human and macaque (ω=0.25) (Consortium, 2005; Gibbs et al., 2007). Using the codonml program, we obtained a significant signal of positive selection at 11 codons (Figure 18). Nine of the positively selected sites involve amino acids that are positively charged and have RNA binding properties. This suggests that positive selection at OAS1 could be driven by interactions between the OAS1 protein and viral dsRNA.

The average sequence divergence at synonymous sites (ds) between class I and II is also surprisingly elevated (1.3%) and is higher than the mean ds between human

Species	Assigned #	ISIS# / Name	Origin	Allele Type
P. t. verus	Pt81	Studbook #380	Riverside Zoo, records suggest captured in Sierra Leone	A/A
	Pt82	Studbook #341	Riverside Zoo, records suggest captured in Sierra Leone	A/A
	Pt87	ISIS # 1142	New Iberia Primate Center	A/A
	Pt97	ISIS #2036	Primate Foundation of Arizona	A/A
	Pt101	ISIS #3214	Southwest Foundation for Biomedical Research	A/A
	Pt102	ISIS #1068	Southwest Foundation for Biomedical Research	A/B
	Pt109	ISIS #1851	Detroit Zoo	A/A
	Pt114	ISIS #2412	New Iberia Primate Center	A/A
	Pt115	ISIS #2738	New Iberia Primate Center	A/A
	Pt117	ISIS #1641	New Iberia Primate Center	A/A
	Pt120	ISIS #2216	New Iberia Primate Center	A/A
	Pt122	ISIS #2417	New Iberia Primate Center	A/A
	Pt124	ISIS #2404	New Iberia Primate Center	A/A
	Pt125	ISIS #2554	New Iberia Primate Center	A/D
	Pt126	ISIS #1818	New Iberia Primate Center	A/D
	Pt1	Banana	Yerkes	A/A
	Pt2	Foxy	Yerkes	A/C
	Pt3	Scott	Yerkes	A/A
	Pt6	Sheila	Yerkes	A/D
	Pt7	Mary	Yerkes	A/A
	Pt8	Artifer	Yerkes	A/D
	Pt9	Teppi	Yerkes	A/A
	Pt11	Beleka	Yerkes	A/D
	Pt12	Adu	Yerkes	A/A
	Pt13	Ossabaw	Yerkes	A/A
	Pt14	Lee	Yerkes	A/I
	Pt15		Yerkes	A/D
P. t. schweinfurthii	Pt96	ISIS #3020	Primate Foundation of Arizona	A/F
	Pt147	Mgbado	CIRMF	A/A
	Pt161	ISIS #925	Primate Foundation of Arizona	A/K
P. t. troglodytes	Pt137	Henri	CIRMF	J/J
	Pt140	Ayrton	CIRMF	A/K
	Pt142	Ponia	CIRMF	C/H
	Pt143	Nestor	CIRMF	C/I
	Pt409	Oliver	Primarily Primates	E/G
	Pt10	Sparkle	Yerkes	A/L
	Pt16	Harv	Yerkes	I/I
P. paniscus	Pt4	Panbanishe	Yerkes	
	Pp05	Studbook #58	Cincinnati Zoo	
	Pp07	Psuke	Georgia State Language Research Center	
	Pp01	Kidajo	Yerkes	
	Pp02	Zalia	Yerkes	
	Pp03	Linda	Yerkes	
	Pp04	Jill	Yerkes	

Table 6: Chimpanzee samples used in this study.

			Exon 1						Exon 2										Exon 3								Exon 4											Exon 5			
Class	Allele	Probability	71	72	83	140	160	161	192	205	281	286	323	349	379	383	410	426	482	485	487	499	525	526	538	601	706	728	739	745	756	764	838	867	870	877	878	909	1008	1089	1090
I	A	1.000	A	G	A	A	C	A	A	A	G	G	C	T	T	A	T	C	C	A	G	G	C	G	T	C	C	A	G	G	A	A	G	A	G	A	G	G	A	T	G
I	B	1.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-
I	C	1.000	-	-	-	-	-	-	-	-	A	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I	D	1.000	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I	E	1.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-
I	F	0.635	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I	G	1.000	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-
I	H	1.000	-	-	-	-	-	-	-	-	A	-	-	-	A	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
II	I	0.962	C	A	T	G	T	G	-	-	-	-	-	-	A	-	-	T	-	G	-	-	T	-	-	G	-	G	A	C	-	G	A	C	-	G	A	A	G	C	-
II	J	1.000	C	A	T	G	T	G	-	-	A	-	-	-	A	-	-	T	-	G	-	-	T	-	-	G	-	G	A	C	-	G	A	C	-	G	A	A	G	C	-
II	K	1.000	C	-	C	G	T	G	-	-	-	A	-	-	A	-	A	T	-	G	-	-	T	-	-	G	-	G	A	C	-	G	A	C	-	G	A	A	G	C	
II	L	0.841	C	-	C	G	T	G	-	-	A	G	-	-	A	-	-	T	-	G	-	-	T	-	-	G	-	G	A	C	-	G	A	C	-	G	A	A	G	C	-
	Paniscus	1.000	C	-	C	G	T	G	-	-	-	-	T	-	-	-	-	T	T	-	-	-	-	-	-	G	-	G	A	C	C	G	A	C	-	G	A	A	G	C	-
	Human	1.000	C	-	T	G	T	G	-	-	-	-	-	-	-	-	-	-	-	-	A	A	-	-	G	-	-	G	A	C	-	G	A	-	-	-	-	-	G	-	-
	Gorilla	1.000	C	-	T	G	T	G	G	G	-	-	-	G	-	G	-	-	-	-	-	A	-	A	-	-	-	-	A	C	-	G	A	-	A	-	G	-	G	C	A
			K	-	K	Q	H		-	T	R	G	A	S	W	D	L	-	T	D	G	D	-	E	Y	Q	-	Q	E	D	-	E	E	R	-	T		-	'	-	E
	Resulting amino acid changes		T	-	T	R	C		-	A	H	R	V	A	R	R	Q	-	I	G	S	N	-	K	D	E	-	R	K	H	-	G	K	S	-	E		-	W	-	K
					M																															R					
	Amino acid position		24	-	28	47	54		-	69	94	96	108	117	127	128	137	-	161	162	163	167	-	176	180	201	-	243	247	249	-	255	280	289	-	293		-	336	-	364
	RNA binding properties		+		+	+	+				+	+			+													+	+	+				+		+					+
	Positive Selection (* > .95, ** > .99)				*	*	**				**	**			**					*						*		*	**							**					

Figure 18: Variation at the OAS1 gene in common chimpanzees. The human, gorilla and pygmy chimpanzee sequences are OAS1 consensus sequences. The PHASE call probability of each haplotype is indicated.

and chimpanzee genes (1.2%; (Consortium, 2005)), suggesting that the two haplotypic classes diverged before the split between human and chimpanzees. A phylogenetic analysis performed on OAS1 coding sequences (Figure 19a) showed that the two classes of haplotypes form two well-supported monophyletic groups. The branches leading to each groups are as long as the branches separating them from human and gorilla, supporting the idea that the two classes are as old or older than the divergence between African apes (chimpanzees, gorillas and humans). Class II haplotypes seem more closely related to bonobo, human and gorilla sequences than they are to class I, although this phylogenetic proximity is not strongly supported by the bootstrap analysis. We further examined the possibility of trans-specific polymorphism by analyzing intronic sequences phylogenetically. On Figure 19b, the tree built with intron 3 sequences shows that class I sequences form a strongly supported clade with the human and gorilla

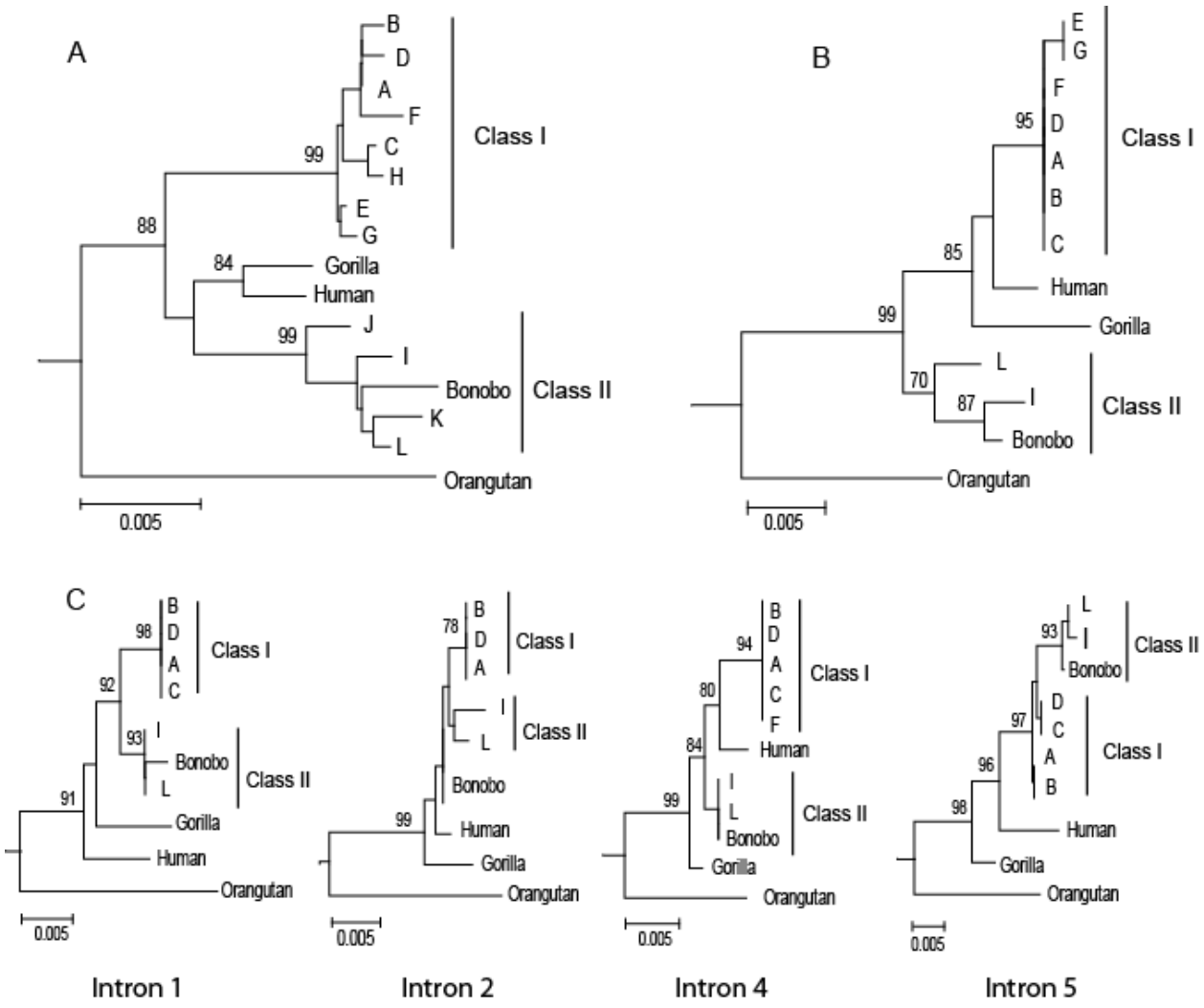

Figure 19: Maximum-likelihood phylogeny of OAS1 haplotypes based on (A) coding sequences, (B) intron 3 and (C) introns 1, 2, 4 and 5. Bootstrap values higher than 75% are indicated.

sequence, whereas class II sequences branch outside of this clade. This demonstrates that the classes separated before the split between humans, chimpanzees and gorillas, 9 Million years ago, and that they were maintained in chimpanzee populations for that period of time. Trans-specific polymorphism was also apparent on the intron 4 tree but not on the trees based on introns 1, 2 and 5. On those trees, class I and II sequences form a monophyletic group relative to human and gorilla, although they are still separated by relatively long branches. The persistence of two groups of haplotypes over such an extended period of evolutionary time is extremely unusual and strongly suggests that balancing selection has been acting on this locus.

Intronic sequences showed an extremely high level of variation consistent with balancing selection. The nucleotidic diversity in intron 3 ($\pi = 1.22$ and $\theta = 1.05$; Table 7) is an order of magnitude higher than the one reported for 50 non-coding fragments of the chimpanzee genome (mean $\pi = 0.132$ and $\theta = 0.194$) (Yu et al., 2003). The nucleotide diversity in other introns (ranging from $\pi = 0.42$ for intron 2 to $\pi = 0.60$ for intron 1) is also larger than the average diversity in chimpanzee, although the difference is not as considerable as for intron 3. Except for intron 5, the values of Tajima's D are all positive. No statistically significant deviation from neutrality was detected, yet positive values of D are consistent with balancing selection. Using the HKA test, we determined if the excess of polymorphisms in introns was significantly different from the level of intraspecific variation expected under neutrality. The HKA test was first performed using a 1.2Kb non-coding region of the genome as neutral region of reference. We were able to reject the null hypothesis of neutrality for introns 3 and 4 but not for introns 1, 2 and 5 (Table 7). We also perform multiple HKA tests using 10 non-coding regions sequenced by Yu et al. as neutral region of reference (Yu et al., 2003). We rejected neutrality for introns 3 (all tests) and for introns 4 (4 out of 10 tests; see supplementary material).

The presence of a peak of diversity in intron 3 suggests that balancing selection is acting on sites that are closely linked to this intron. Indeed, exons 3 and 4 contain 5 positively selected sites. As three of these sites are presumptive dsRNA binding sites, we speculated that the alleles under balancing selection encode for proteins that differ in their dsRNA binding affinities and, possibly, in their enzymatic activity. We first verified that chimpanzee OAS1 proteins have the ability to bind RNA.

REGIONS	L[a]	N[b]	S[c]	π(%)[d]	θ(%)[e]	D[f]	HKA chr4g	
							χ^2	*P*-value
INTRON 1	837	7	8	0.602	0.525	1.44	2.99	0.084
INTRON 2	787	7	6	0.424	0.416	0.18	2.16	0.14
INTRON 3	1414	7	27	1.216	1.053	1.59	6.21	**0.013**
INTRON 4	738	7	5	0.452	0.37	2.12	4.1	**0.043**
INTRON 5	955	7	7	0.493	0.513	-0.39	2.79	0.095
CHROMOSOME 4	1195	7	4	0.182	0.183	-0.06	-	-

[a] Length of the fragment in base pairs
[b] Number of individuals sequenced
[c] Number of segregating sites
[d] Nucleotidic diversity
[e] Watterson's estimator of nucleotidic diversity
[f] Tajima's D
[g] Results of the HKA test using the region on chromosome 4 as neutral region of reference

Table 7: Summary statistics for OAS1 introns. P values <0.05 are indicated in bold.

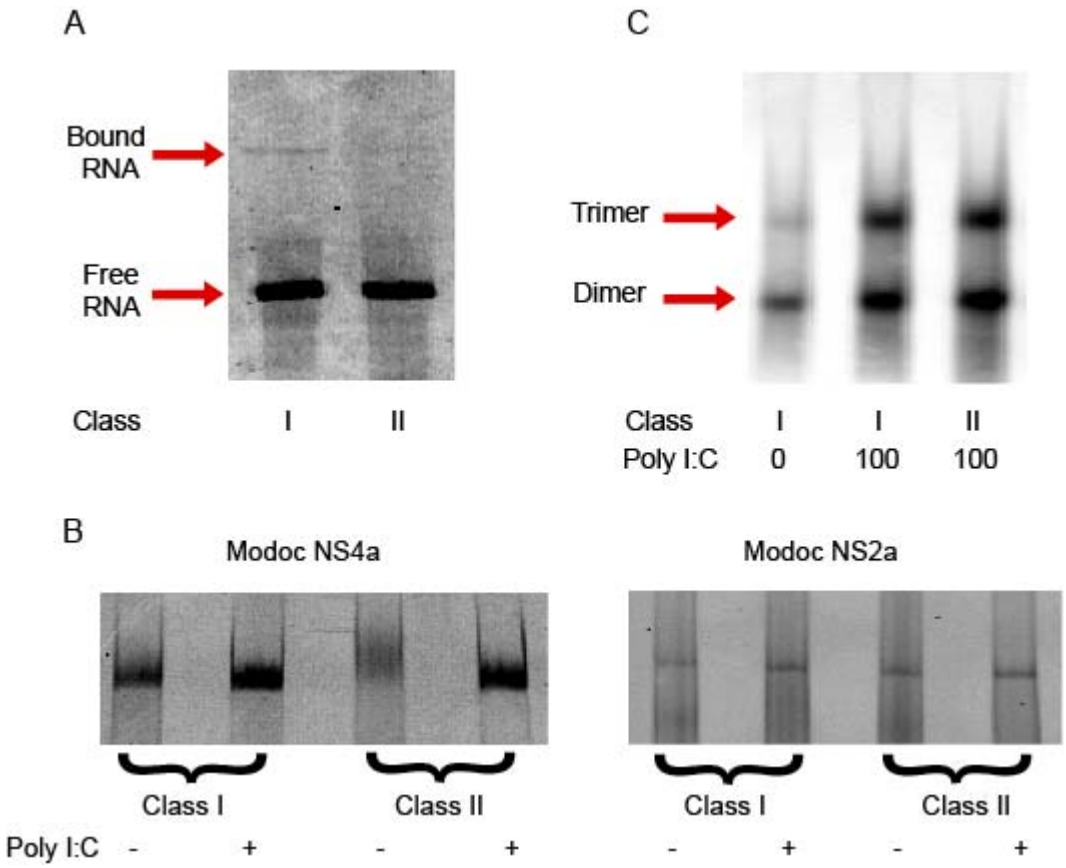

Figure 20: A. RNA binding of HIV-1 *nef* RNA to OAS1 proteins. B. Band shift assay with Modoc NS4a and NS2a RNA, with or without competitor. C. Detection of oligoadenylates generated by class I and II OAS1 proteins in the presence of synthetic RNA (poly I:C).

Figure 20a displays the result of a binding assay with an excess of labeled HIV-1 *nef* RNA. On this gel, a band corresponding to the RNA-protein complex appears, confirming the RNA-binding capabilities of the two alleles. We then performed a band shift assay with or without competitor (poly I:C). The principle of this assay is that protein-bound RNA will migrate slower than unbound RNA. In absence of competitor, the band corresponding to the free labeled RNA should be less intense than the same band incubated with competitor because a fraction of the RNA is bound to OAS1 protein. For some RNA's, like the Modoc NS2a RNA, we observed only a minimal difference of intensity, with or without competitors, indicating that OAS1 does not have a strong affinity to these RNA's (Figure 20b). In contrast, the band corresponding to the Modoc NS4a RNA is less intense in absence of competitor than with competitor, suggesting that a significant fraction of this RNA is bound to the OAS1 protein (Figure 20b). Interestingly, we observed a difference in binding affinity between class I and class II OAS1 proteins. In absence of competitor, 35% of the Modoc NS4a RNA binds to class I OAS1 whereas most (70%) of the RNA binds to the class II OAS1, leading to the near disappearance of the band. This demonstrates that the two types of OAS1 differ in their RNA binding affinity, as predicted by the evolutionary analysis. Out of the 5 RNA's for which we detected binding to OAS1 by band shift assay, two showed a significant difference between class I and class II alleles, the Modoc NS4a and the *nef* gene of HIV-1. In the case of *nef*, it is the class I allele that shows the highest binding affinity.

We examined if differences in RNA binding cause variations in enzymatic activity. We first verified that the chimpanzee OAS1s were enzymatically active by incubating each allele with poly I:C (Figure 20c). We obtained similar amount of oligoadenylate dimer and trimer with both alleles indicating that they are both active and that their basal enzymatic activity does not differ significantly. The enzymatic activity was then assessed by measuring the quantity of oligoadenylate trimers produced when each allele is incubated with different RNA's. Figure 21a shows that the amount of trimer synthesized differ drastically among RNA's and correlates well with the binding affinity of the RNA to OAS1. For instance, very few trimers are produced when OAS1 is incubated with Modoc's NS2a RNA, a RNA with low binding affinity to OAS1, whereas

incubation with high-affinity RNA's produced comparatively a large amount of trimers. Significant

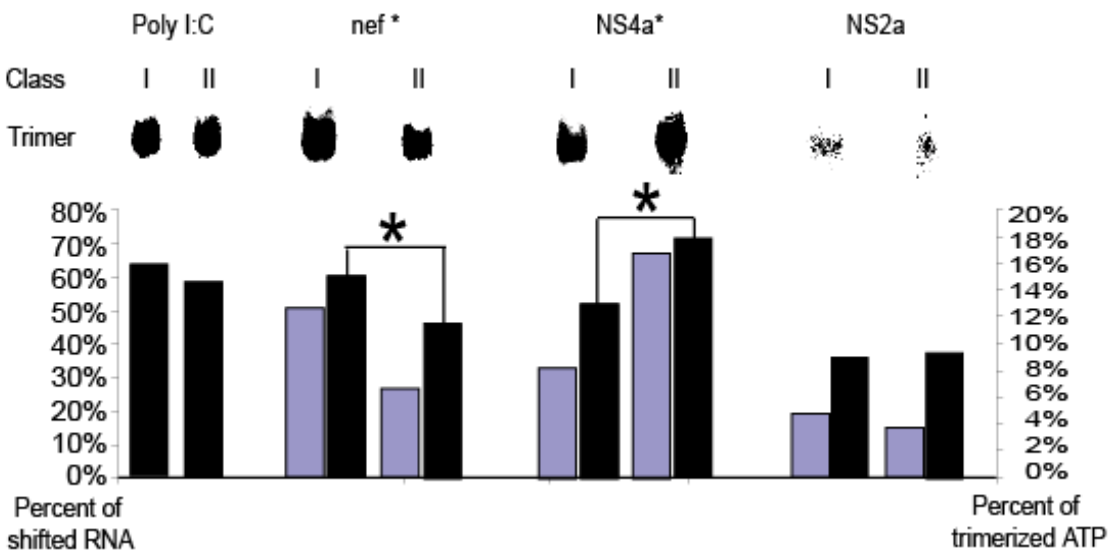

Figure 21: Quantification of trimeric oligoadenylate synthesized by class I and class II OAS1. The star indicates that the difference in the amount of trimer produced is significant. The grey bars indicate the percent of RNA that is bound to OAS1 in the gel shift assay and the open bars the % of rATP trimerized.

differences in trimer production by class I and class II alleles were observed for the NS4a and for the *nef* RNA's. The class I OAS1 allele produces more trimers than class II allele in presence of *nef* RNA while the opposite is true with NS4a RNA (Figure 21). Those differences are consistent with the band shift assays and prove that the differential binding affinities of the two alleles cause variations in the production of oligoadenylates.

DISCUSSION

We demonstrated that two deeply divergent haplotypic lineages are present at the OAS1 gene in common chimpanzees. Phylogenetic analysis revealed that these two lineages are older than the split between humans, chimpanzees and gorillas that occurred 9 million years ago, making this polymorphism the oldest one in hominoids outside the MHC (Lawlor et al., 1988; Mayer et al., 1988). The persistence of alleles over such a long period of time is extremely unusual and strongly suggests that OAS1 has evolved under balancing selection. We provide several lines of evidence that support this hypothesis, including an extremely high nucleotidic diversity in introns, a significant deviation from neutrality as revealed by the HKA test, and a high ratio of non-synonymous to synonymous substitutions between lineages. Balanced polymorphisms are very rare in nature and their persistence across species is even more uncommon (Charlesworth, 2006). This is because the signature of balancing selection can easily be erased by recombination and drift. The evolution of novel alleles that sweep the population and replace pre-existing polymorphisms as well as drastic reductions in population size can similarly reduce the diversity of loci that have evolved under balancing selection. In hominoids, the only well-documented case of long-term balancing selection resulting in trans-specific polymorphism involves genes of the MHC (Lawlor et al., 1988; Mayer et al., 1988). A recent screen of the human genome failed to find a single other gene evolving under balancing selection (Bubb et al., 2006), reinforcing the view that the conditions for a polymorphism to be maintained by balancing selection are extremely restrictive. This does not mean that balancing selection has not played a significant role in hominoid evolution. Humans went through a drastic population bottleneck that caused a reduction in polymorphisms, including

balanced ones. Other hominoids (e.g. chimpanzees) do not seem to have suffered a population bottleneck to the same extent humans have (Becquet et al., 2007; Fischer et al., 2006; Stone et al., 2002) and are more likely to have retained ancestral polymorphisms, as exemplified by our analysis of OAS1.

The observation that OAS1 is evolving under balancing selection implies some functional differences between class I and II haplotypes. Of the 19 polymorphic amino-acids, 12 involve positively charged amino-acids and 9 are located in the vicinity of a positive groove in the OAS1 protein responsible for RNA binding and subsequent activity (Hartmann R et al., 2003). The location and class of polymorphic amino acids suggests that balancing selection at OAS1 is driven by interactions between the OAS1 protein and viral RNA. We demonstrated experimentally that it is likely to be the case as class I and II OAS1 differ in their RNA binding affinity. It is therefore plausible that OAS1 polymorphisms confer different level of protection during viral infection. Our data indicate that the OAS1 alleles produce variable amount of oligoadenylates when incubated with different viral RNA's. However, it is unclear if the relatively minor differences in enzymatic activity we observe would lead to different levels of Rnase L activity. It is possible that other pathways are involved. The existence of a Rnase L-independent pathway has been recently proposed (Scherbik et al., 2006) because an enzymatically inactive mouse OAS1 homologue, *Oas1b*, confers specific resistance against flavivirus infection (Mashimo T, 2002; Perelygin AA et al., 2002). Interestingly, we recently discovered that long-term balancing selection is also acting on this gene in mice (Ferguson W et al., 2008).

While the exact antiviral mechanism conferred by OAS1 in chimpanzee remains to be clarified, we can still address how this polymorphism was maintained in chimp populations. Balancing selection occurs when heterozygotes have a selective advantage over homozygotes, when rare alleles are advantaged over frequent ones (i.e. frequency-dependent selection), or when the nature or intensity of selection varies in time or space (Hedrick, 2007; Hedrick et al., 1976; Kojima, 1971b). We speculate that each haplotypic class confers a better resistance against some viral strains than others. Under the overdominance scenario, heterozygous individuals would benefit from the protection of both haplotype and be protected against a larger diversity of viruses.

Overdominance implies that both classes will be at high frequency in populations because heterozygous individuals have a higher fitness. Under the frequency dependent scenario (and selection that varies in time and space), the frequency of each haplotype should vary cyclically and/or geographically. In our limited sample of Central and East African chimpanzee, each class is equally represented, suggesting that none of the haplotypic classes is rare. Although this is consistent with the overdominance hypothesis, more data relative to the frequency of each haplotypic class in natural chimpanzee populations are needed to solve this question.

ACKNOWLEDGEMENTS

The work was conducted in part with equipment from the Core Facilities for Imaging, Cellular and Molecular Biology at Queens College. This research was supported by NIH grant AI058968-01A1 and PSC-CUNY grant 69552-00 38 to Stéphane Boissinot.

Bibliography

Abel, K., Alegria-Hartman, M. J., Rothaeusler, K., Marthas, M. and Miller, C. J. (2002). The relationship between simian immunodeficiency virus RNA levels and the mRNA levels of alpha/beta interferons (IFN-alpha/beta) and IFN-alpha/beta-inducible Mx in lymphoid tissues of rhesus macaques during acute and chronic infection. *J Virol* **76**, 8433-45.

Ajariyakhajorn, C., Mammen, M. P., Jr., Endy, T. P., Gettayacamin, M., Nisalak, A., Nimmannitya, S. and Libraty, D. H. (2005). Randomized, placebo-controlled trial of nonpegylated and pegylated forms of recombinant human alpha interferon 2a for suppression of dengue virus viremia in rhesus monkeys. *Antimicrob Agents Chemother* **49**, 4508-14.

Andrews DM, Matthews VB, Sammels LM, Carrello AC and McMinn PC. (1999). The severity of murray valley encephalitis in mice is linked to neutrophil infiltration and inducible nitric oxide synthase activity in the central nervous system. *J Virol.* **73**, 8781-90.

Apperson CS, Harrison BA, Unnasch TR, Hassan HK, Irby WS, Savage HM, Aspen SE, Watson DW, Rueda LM, Engber BR et al. (2002). Host-feeding habits of Culex and other mosquitoes (Diptera: Culicidae) in the Borough of Queens in New York City, with characters and techniques for identification of Culex mosquitoes. *J Med Entomol.* **39**, 777-85.

Arias CF, Preugschat F and Strauss JH. (1993). Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. *Virology* **193**, 888-99.

Arunachalam N, Samuel PP, Hiriyan J, Rajendran R and Dash AP. (2005). Short report: observations on the multiple feeding behavior of Culex tritaeniorhynchus (Diptera: culicidae), the vector of Japanese encephalitis in Kerala in southern India. *Am J Trop Med Hyg.* **72**, 198-200.

Bagley CJ, Woodcock JM, Stomski FC and Lopez AF. (1997). The Structural and Functional Basis of Cytokine Receptor Activation: Lessons From the Common beta Subunit of the Granulocyte-Macrophage Colony-Stimulating Factor, Interleukin-3 (IL-3), and IL-5 Receptors. *Blood* **89**, 1471-1482.

Bartholomeusz AI and Wright PJ. (1993). Synthesis of dengue virus RNA in vitro: initiation and the involvement of proteins NS3 and NS5. *Arch Virol.* **128**, 111-21.

Basler, C. F., Wang, X., Muhlberger, E., Volchkov, V., Paragas, J., Klenk, H. D., Garcia-Sastre, A. and Palese, P. (2000). The Ebola virus VP35 protein functions as a type I IFN antagonist. *Proc Natl Acad Sci U S A* **97**, 12289-94.

Becquet, C., Patterson, N., Stone, A. C., Przeworski, M. and Reich, D. (2007). Genetic structure of chimpanzee populations. *PLoS Genet* **3**, 0617-0625.

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

Bell JR, Kinney RM, Trent DW, Lenches EM, Dalgarno L and Strauss JH. (1985). Amino-terminal amino acid sequences of structural proteins of three flaviviruses. *Virology* **143**, 224-9.

Bhamarapravati N. (1997). Pathology of dengue infections. *In: Gubler DJ, Kuno G, eds. Dengue and dengue haemorrhagic fever.*, 115–32.

Boege U, Heinz FX, Wengler G and Kunz C. (1983). Amino acid compositions and amino-terminal sequences of the structural proteins of a flavivirus, European Tick-Borne Encephalitis virus. *Virology* **126**, 651-7.

Bonnevie-Nielsen, V., Field, L. L., Lu, S., Zheng, D. J., Li, M., Martensen, P. M., Nielsen, T. B., Beck-Nielsen, H., Lau, Y. L. and Pociot, F. (2005). Variation in antiviral 2',5'-oligoadenylate synthetase (2'5'AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene. *Am J Hum Genet* **76**, 623-33.

Bonnevie-Nielsen, V., Larsen, M. L., Frifelt, J. J., Michelsen, B. and Lernmark, A. (1989). Association of IDDM and attenuated response of 2',5'-oligoadenylate synthetase to yellow fever vaccine. *Diabetes* **38**, 1636-42.

Boursot P, Din W, Anand R, Darviche D, Dod B, Von Deimling F, Talwar GP and Bonhomme F. (1996). Origin and radiation of the house mouse: mitochondrial DNA phylogeny. *J. Evol. Biol.* **9**, 391-415.

Brinton MA. (2002). The molecular biology of West Nile virus: a new invader of the Western hemisphere. *Annu. Rev. Microbiol.* **56**, 371-402.

Brinton MA and Dispoto JH. (1988). Sequence and secondary structure analysis of the 5'-terminal region of flavivirus genome RNA. *Virology* **162**, 290-9.

Brinton MA and Fernandez AV. (1983). A replication-efficient mutant of West Nile virus is insensitive to DI particle interference. *Virology* **129**, 107-115.

Brinton MA and Perelygin AA, S. S., Zhulin IB, Stockman BM, Li Y, Brinton MA. (2003). Genetic resistance to flaviviruses. *Adv Virus Res.* **60**, 43-85.

Bubb, K. L., Bovee, D., Buckley, D., Haugen, E., Kibukawa, M., Paddock, M., Palmieri, A., Subramanian, S., Zhou, Y., Kaul, R. et al. (2006). Scan of human genome reveals no new Loci under ancient balancing selection. *Genetics* **173**, 2165-77.

Calattini, S., Nerrienet, E., Mauclere, P., Georges-Courbot, M. C., Saib, A. and Gessain, A. (2006). Detection and molecular characterization of foamy viruses in Central African chimpanzees of the Pan troglodytes troglodytes and Pan troglodytes vellerosus subspecies. *J Med Primatol* **35**, 59-66.

Cardenas, W. B., Loo, Y. M., Gale, M., Jr., Hartman, A. L., Kimberlin, C. R., Martinez-Sobrido, L., Saphire, E. O. and Basler, C. F. (2006). Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. *J Virol* **80**, 5168-78.

Carlson CB, Stephens OM and Beal PA. (2003). Recognition of double-stranded RNA by proteins and small molecules. *Biopolymers.* **70**, 86-102.

Castle E, Leidner U, Novak T and Wengler G. (1986). Primary structure of the West Nile flavivirus genome coding for all nonstructural proteins. *Virology* **149**, 10-26.

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

Chen CJ, Kuo MD, Chien LJ, Hsu SL, Wang YM and Lin JH. (1997). RNA-protein interactions: involvement of NS3, NS5, and 3' noncoding regions of Japanese encephalitis virus genomic RNA. *J Virol.* **71**, 3466-73.

Chen Y, Maguire T and Marks RM. (1996). Demonstration of binding of dengue virus envelope protein to target cells. *J Virol.* **70**, 8765–8772.

Christensen HA, de Vasquez AM and Boreham MM. (1996). Host-feeding patterns of mosquitoes (Diptera: Culicidae) from central Panama. *Am J Trop Med Hyg.* **52**, 202-8.

Consortium, C. S. a. A. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69-87.

Davis JW and Hardy JL. (1973). In vitro studies with Modoc Virus in Vero cells: plaque assay and kinetics of growth, neutralization, and thermal inactivation. *Applied Microbiology* **3**, 344-348.

Davis JW and Hardy JL. (1974). Characterization of persistent Modoc viral infections in Syrian hamsters. *Infect Immun.* **2**, 328-34.

Davis JW, Hardy JL and Reeves WC. (1974). Modoc viral infections in the deer mouse Peromyscus maniculatus. *Infect Immun.* **6**, 1362-9.

Desai SY, Patel RC, Sen GC, Malhotra P, Ghadge GD and Thimmapaya B. (1995). Activation of interferon-inducible 2'-5' oligoadenylate synthetase by adenoviral VAI RNA. *J Biol Chem.* **7**, 3454-61.

Diaz-Guerra, M., Rivas, C. and Esteban, M. (1997). Inducible expression of the 2-5A synthetase/RNase L system results in inhibition of vaccinia virus replication. *Virology* **277**, 220-228.

Edwards, S. V., Chesnut, K., Satta, Y. and Wakeland, E. K. (1997). Ancestral polymorphism of Mhc class II genes in mice: implications for balancing selection and the mammalian molecular clock. *Genetics* **146**, 655-68.

Enserink, M. (2004). SARS treatment. Interferon shows promise in monkeys. *Science* **303**, 1273-5.

Eskildsen S, Hartmann R, Kjeldgaard N and Justesen J. (2002). Gene structure of the murine 2'-5'-oligoadenylate synthetase family. *Cellular and Molecular Life Sciences* **59**, 1212-1222.

Falcone, V., Schweizer, M., Toniolo, A., Neumann-Haefelin, D. and Meyerhans, A. (1999). Gamma interferon is a major suppressive factor produced by activated human peripheral blood lymphocytes that is able to inhibit foamy virus-induced cytopathic effects. *J Virol* **73**, 1724-8.

Feng, Z., Cerveny, M., Yan, Z. and He, B. (2007). The VP35 protein of Ebola virus inhibits the antiviral effect mediated by double-stranded RNA-dependent protein kinase PKR. *J Virol* **81**, 182-92.

Ferguson W, Dvora S, Gallo J, Orth A and Boissinot S. (2008). Long-Term Balancing Selection at the West-Nile Virus Resistance Gene, Oas1b, Maintains Trans-Specific Polymorphisms in the House Mouse. *Mol Biol Evol.*

Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marques-Bonet, T., Ramirez-Soriano, A., Muntasell, A., Navarro, A., Lazarus, R., Calafell, F., Bertranpetit, J. et al. (2008). Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol* **181**, 1315-22.

Figueroa, F., Gunther, E. and Klein, J. (1988). MHC polymorphism pre-dating speciation. *Nature* **335**, 265-7.

Fischer, A., Pollack, J., Thalmann, O., Nickel, B. and Paabo, S. (2006). Demographic history and genetic differentiation in apes. *Curr Biol* **16**, 1133-1138.

Formenty, P., Boesch, C., Wyers, M., Steiner, C., Donati, F., Dind, F., Walker, F. and Le Guenno, B. (1999). Ebola virus outbreak among wild chimpanzees living in a rain forest of Cote d'Ivoire. *J Infect Dis* **179 Suppl 1**, S120-6.

Gamian, A., Chomik, M., Laferriere, C. A. and Roy, R. (1991). Inhibition of influenza A virus hemagglutinin and induction of interferon by synthetic sialylated glycoconjugates. *Can J Microbiol* **37**, 233-7.

Gaunt MW, Sall AA, de Lamballerie X, Falconar AK, Dzhivanian TI and Gould EA. (2001). Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J Gen Virol.* **82**, 1867-76.

Ghosh A, Desai SY, Sarkar SN, Ramaraj P, Ghosh SK, Bandyopadhyay S and Sen GC. (1997a). Effects of mutating specific residues present near the amino terminus of 2'-5'-oligoadenylate synthetase. *J Biol Chem.* **24**, 15452-8.

Ghosh A, Sarkar SN, Guo W, Bandyopadhyay S and Sen GC. (1997b). Enzymatic activity of 2'-5'-oligoadenylate synthetase is impaired by specific mutations that affect oligomerization of the protein. *J Biol Chem.* **52**, 33220-6.

Ghosh SK, Kusari J, Bandyopadhyay SK, Samanta H, Kumar R and Sen GC. (1991). Cloning, sequencing, and expression of two murine 2'-5'-oligoadenylate synthetases. Structure-function relationships. *J. Biol. Chem.* **266**, 15293-15299.

Gibbs, R. A. Rogers, J. Katze, M. G. Bumgarner, R. Weinstock, G. M. Mardis, E. R. Remington, K. A. Strausberg, R. L. Venter, J. C. Wilson, R. K. et al. (2007). Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222-34.

Goodbourn, S., Didcock, L. and Randall, R. E. (2000). Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures. *J Gen Virol* **81**, 2341-64.

Goodman GT and Koprowski H. (1962). Study of the mechanism of innate resistance to virus infection. *Journal of Cellular and Comparative Physiology* **59**, 333-373.

Gribaudo, G., Lembo, D., Cavallo, G., Landolfo, S. and Lengyel, P. (1991). Interferon action: binding of viral RNA to the 40-kilodalton 2'-5'-oligoadenylate synthetase in interferon-treated HeLa cells infected with encephalomyocarditis virus. *J Virol.* **65**, 1748-1757.

Gubler, D. (1998a). Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev.* **11**, 480-96.

Guindon, S., Lethiec, F., Duroux, P. and Gascuel, O. (2005). PHYML Online - A web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* **33**.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Hamano, E., Hijikata, M., Itoyama, S., Quy, T., Phi, N. C., Long, H. T., Ha le, D., Ban, V. V., Matsushita, I., Yanai, H. et al. (2005). Polymorphisms of interferon-inducible genes OAS-1 and MxA associated with SARS in the Vietnamese population. *Biochem Biophys Res Commun* **329**, 1234-9.

Hamano, E., M. Hijikata, S. Itoyama, T. Quy, N.C. Phi, H.T. Long, D. Ha le, V.V. Ban, I. Matsushita, H. Yanai, F. Kirikae, T. Kirikae, T. Kuratsuji, T. Sasazuki, and N. Keicho. (2005). Polymorphisms of interferon-inducible genes OAS-1 and MxA

associated with SARS in the Vietnamese population. *Biochem Biophys Res Commun* **329**, 1234-1239.

Hartmann R, Justesen J, Sarkar SN, Sen GC and Yee VC. (2003). Crystal structure of the 2'-specific and double-stranded RNA-activated interferon-induced antiviral protein 2'-5'-oligoadenylate synthetase. *Mol Cell.* **5**, 1173-85.

Hartmann R, Norby PL, Martensen PM, Jorgensen P, James MC, Jacobsen C, Moestrup SK, Clemens MJ and Justesen J. (1998). Activation of 2'-5' oligoadenylate synthetase by single-stranded and double-stranded RNA aptamers. *J Biol Chem.* **6**, 3236-46.

Hartmann R, Walko G and Justesen J. (2001). Inhibition of 2'-5' oligoadenylate synthetase by divalent metal ions. *FEBS Lett.* **1**, 54-8.

Hassel B, Zhou A, Sotomayor C, Maran A and Silverman R. (1993). A dominant negative mutant of 2-5A-dependent RNase suppresses antiproliferative and antiviral effects of interferon. *EMBO J.* **8**, 3297-304.

Hedrick, P. W. (1976). Genetic variation in a heterogeneous environment. II. Temporal heterogeneity and directional selection. *Genetics* **84**, 145-57.

Hedrick, P. W. (2007). Balancing selection. *Curr Biol* **17**, R230-1.

Hedrick, P. W., Ginevan, M. and Ewing, E. (1976). Genetic polymorphism in heterogeneous environment. *Annual Review in Ecology and Systematics* **7**, 1-32.

Henchal EA and Putnak JR. (1990). The dengue viruses. *Clin Microbiol Rev.* **3**, 376-396.

Hill, A. (2006). Aspects of genetic susceptibility to human infectious diseases. *Annual Review of Genetics* **40**, 469-486.

Hovanessian, A. G. (1991). Interferon-induced and double-stranded RNA-activated enzymes: a specific protein kinase and 2',5'-oligoadenylate synthetases. *J Interferon Res* **11**, 199-205.

Hovanessian AG and Justesen J. (2007). The human 2'-5'oligoadenylate synthetase family: unique interferon-inducible enzymes catalyzing 2'-5' instead of 3'-5' phosphodiester bond formation. *Biochimie* **89**, 779-88.

Hudson, R. R., Kreitman, M. and Aguade, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153-9.

Hurrelbrink R and McMinn C. (2003). Molecular determinants of virulence: the structural and functional basis for flavivirus attenuation. *Advances in Virus Research* **60**, 1 - 42.

Jahrling, P. B., Geisbert, T. W., Geisbert, J. B., Swearengen, J. R., Bray, M., Jaax, N. K., Huggins, J. W., LeDuc, J. W. and Peters, C. J. (1999). Evaluation of immune globulin and recombinant interferon-alpha2b for treatment of experimental Ebola virus infections. *J Infect Dis* **179 Suppl 1**, S224-34.

Justesen J, Hartmann R and Kjeldgaard NO. (2000). Gene structure and function of the 2'-5'-oligoadenylate synthetase family. *Cell Mol Life Sci.* **57**, 1593-612.

Kakuta S, Shibata S and Iwakura Y. (2002). Genomic structure of the mouse 2',5'-oligoadenylate synthetase gene family. *J Interferon Cytokine Res.* **22**, 981-93.

Keele, B. F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M. L., Bibollet-Ruche, F., Chen, Y., Wain, L. V., Liegeois, F. et al. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523-6.

Kerr IM and Brown RE. (1978). pppA2'p5'A2'p5'A: an inhibitor of protein synthesis synthesized with an enzyme fraction from interferon-treated cells. *Proc Natl Acad Sci* **1**, 256-60.

Kim, M. J., Latham, A. G. and Krug, R. M. (2002). Human influenza viruses activate an interferon-independent transcription of cellular antiviral genes: outcome with influenza A virus is unique. *Proc Natl Acad Sci U S A* **99**, 10096-101.

Kim, S. Y., Li, J., Bentsman, G., Brooks, A. I. and Volsky, D. J. (2004). Microarray analysis of changes in cellular gene expression induced by productive infection of primary human astrocytes: implications for HAD. *J Neuroimmunol* **157**, 17-26.

Knapp S, Yee LJ, Frodsham AJ, Hennig BJ, Hellier S, Zhang L, Wright M, Chiaramonte M, Graves M, Thomas HC et al. (2003). Polymorphisms in interferon-induced genes and the outcome of hepatitis C virus infection: roles of MxA, OAS-1 and PKR. *Genes Immun.* **6**, 411-9.

Kojima, K. (1971a). The distribution and comparison of "genetic loads" under heterotic selection and simple frequency-dependent selection in finite populations. *Theor Popul Biol* **2**, 159-73.

Kojima, K. (1971b). Is there a constant fitness for a given genotype? No! *Evolution Int J Org Evolution* **25**, 281-285.

Kuhn RJ, Zhang W, Rossmann MG, Pletnev SV, Corver J, Lenches E, Jones CT, Mukhopadhyay S, Chipman PR, Strauss EG et al. (2002). Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717-25.

Kumar S, Mitnik C, Valente G and Floyd-Smith G. (2000). Expansion and molecular evolution of the interferon-induced 2'-5' oligoadenylate synthetase gene family. *Mol Biol Evol.* **17**, 738-50.

Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N and Cropp CB. (1998). Phylogeny of the genus Flavivirus. *J Virol.* **72**, 73-83.

Lawlor, D. A., Ward, F. F., Ennis, P. D., Jackson, A. P. and Parham, P. (1988). HLA-A, -B polymorphisms predate the divergence of humans and chimpanzees. *Nature* **335**, 268-271.

Leendertz, F. H., Junglen, S., Boesch, C., Formenty, P., Couacy-Hymann, E., Courgnaud, V., Pauli, G. and Ellerbrok, H. (2004). High variety of different simian T-cell leukemia virus type 1 strains in chimpanzees (Pan troglodytes verus) of the Tai National Park, Cote d'Ivoire. *J Virol* **78**, 4352-6.

Leyssen P, de Clercq E and Neyts J. (2000). Perspectives for the treatment of infections with Flaviviridae. *Clin Microbiol Rev.* **13**, 67-82.

Leyssen P, Van Lommel A, Drosten C, Schmitz H, de Clercq E and Neyts J. (2001). A novel model for the study of the therapy of flavivirus infections using the Modoc virus. *Virology* **279**, 27-37.

Leyssen, P., De Clercq, E. and Neyts, J. (2000). Perspectives for the treatment of infections with Flaviviridae. *Clin Microbiol Rev* **13**, 67-82, table of contents.

Lindenbach BD and Rice CM. (1997). trans-Complementation of yellow fever virus NS1 reveals a role in early RNA replication. *J Virol.* **71**, 9608-17.

Lindenbach BD and Rice CM. (2001). Flaviviridae: the viruses and their replication. *Fields Virology*, 991-1041.

Lindenbach, B. D., Meuleman, P., Ploss, A., Vanwolleghem, T., Syder, A. J., McKeating, J. A., Lanford, R. E., Feinstone, S. M., Major, M. E., Leroux-Roels, G. et al. (2006). Cell culture-grown hepatitis C virus is infectious in vivo and can be recultured in vitro. *Proc Natl Acad Sci U S A* **103**, 3805-9.

Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., MacDonald, M. E., Stuhlmann, H., Koup, R. A. and Landau, N. R. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367-77.

Liu, W. J., Chen, H. B. and Khromykh, A. A. (2003). Molecular and functional analyses of Kunjin virus infectious cDNA clones demonstrate the essential roles for NS2A in virus assembly and for a nonconservative residue in NS3 in RNA replication. *J Virol* **77**, 7804-13.

Liu, W. J., Chen, H. B., Wang, X. J., Huang, H. and Khromykh, A. A. (2004). Analysis of adaptive mutations in Kunjin virus replicon RNA reveals a novel role for the flavivirus nonstructural protein NS2A in inhibition of beta interferon promoter-driven transcription. *J Virol* **78**, 12225-35.

Lo MK, T. M., Bernard KA, Shi PY. (2003). Functional analysis of mosquito-borne flavivirus conserved sequence elements within 3' untranslated region of West Nile virus by use of a reporting replicon that differentiates between viral translation and RNA replication. *J Virol.* **77**, 10004-14.

MacDonald, D. M., Holmes, E. C., Lewis, J. C. and Simmonds, P. (2000). Detection of hepatitis B virus infection in wild-born chimpanzees (Pan troglodytes verus): phylogenetic relationships with human and other primate genotypes. *J Virol* **74**, 4253-7.

Mackenzie JM, W. E. (2001). Assembly and maturation of the flavivirus Kunjin virus appear to occur in the rough endoplasmic reticulum and along the secretory pathway, respectively. *J Virol.* **75**, 10787-99.

Maitra RK, McMillan N, D. S., McSwiggen J, Hovanessian AG, Sen G, Williams B and R, S. (1994). HIV-1 TAR RNA Has an Intrinsic Ability to Activate Interferon-Inducible Enzymes. *Virology* **204**, 823-827.

Marie, I., Rebouillat, D. and Hovanessian, A. G. (1999). The expression of both domains of the 69/71 kDa 2',5' oligoadenylate synthetase generates a catalytically active enzyme and mediates an anti-viral response. *Eur J Biochem* **262**, 155-65.

Marin MS, Z. P., Gritsun TS, Gould EA. (1995). Phylogeny of TYU, SRE, and CFA virus: different evolutionary rates in the genus Flavivirus. *Virology.* **206**, 1133-9.

Montagutelli, X., Despres, P. and Guenet, J. L. (2003). Structural and functional genomics and evolutionary relationships in the cluster of genes encoding murine 2',5'-oligoadenylate synthetases. *Genomics* **82**, 537-52.

Mashimo T, G. P., Lucas M, Simon-Chazottes D, Ceccaldi PE, Montagutelli X, Despres P, Guenet JL. (2003). Structural and functional genomics and evolutionary relationships in the cluster of genes encoding murine 2',5'-oligoadenylate synthetases. *Genomics.* **82**, 537-52.

Mashimo T, L. M., Simon-Chazottes D, Frenkiel MP, Montagutelli X, Ceccaldi PE, Deubel V, Guenet JL, Despres P. (2002). A nonsense mutation in the gene encoding 2'-5'-oligoadenylate synthetase/L1 isoform is associated with West Nile virus susceptibility in laboratory mice. *Proc Natl Acad Sci* **99**, 11311-6.

Mason PW, McAda PC, Mason TL and Fournier MJ. (1987). Sequence of the dengue-1 virus genome in the region encoding the three structural proteins and the major nonstructural protein NS1. *Virology* **161**, 262-267.

Matthews V, R. T., Kendrick T, Abdo M, Papadimitriou J, McMinn P. (2000). Morphological features of Murray Valley encephalitis virus infection in the central nervous system of Swiss mice. *Int J Exp Pathol.* **81**, 31-40.

Mayer, W. E., Jonker, D., Klein, D., Ivanyi, P., van Seventer, G. and Klein, J. (1988). Nucleotide sequence of chimpanzee MHC class I alleles: evidence for transspecies mode of evolution. *EMBO J.* **7**, 2765-2774.

McMinn PC. (1997). The molecular basis of virulence of the encephalitogenic flaviviruses. *J Gen Virol.* **78**, 2711-22.

McMinn PC, D. L., Weir RC. (1996). A comparison of the spread of Murray Valley encephalitis viruses of high or low neuroinvasiveness in the tissues of Swiss mice after peripheral inoculation. *Virology* **2**, 414-23.

Modis Y, O. S., Clements D, Harrison SC. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**, 313-319.

Monath TP and Heinz FX. (1994). Dengue: the risk to developed and developing countries. *Proc Natl Acad Sci* **91**, 2395–2400.

Monath TP and Heinz FX. (1996). Flavivirus. *Fields Virology* **Third Edition**, 961-1034.

Muñoz-Jordán JL, Sánchez-Burgos GG, Laurent-Rolle M and García-Sastre A. (2003). Inhibition of interferon signaling by dengue virus. *Proc Natl Acad Sci* **100**, 14333–14338.

Nerrienet, E., Meertens, L., Kfutwah, A., Foupouapouognigni, Y., Ayouba, A. and Gessain, A. (2004). Simian T cell leukaemia virus type I subtype B in a wild-caught gorilla (Gorilla gorilla gorilla) and chimpanzee (Pan troglodytes vellerosus) from Cameroon. *J Gen Virol* **85**, 25-9.

Newman, R. M., Hall, L., Connole, M., Chen, G. L., Sato, S., Yuste, E., Diehl, W., Hunter, E., Kaur, A., Miller, G. M. et al. (2006). Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5alpha. *Proc Natl Acad Sci U S A* **103**, 19134-9.

Niyomrattanakit P, Winoyanuwattikun P, Chanprapaph S, Angsuthanasombat C, Panyim S and Katzenmeier G. (2004). Identification of residues in the dengue virus type 2 NS2B cofactor that are critical for NS3 protease activation. *J Virol.* **78**, 13708-16.

Oyanagi S, Ikuta F and Ross ER. (1969). Electron microscopic observations in mice infected with Japanese encephalitis. *Acta Neuropathol (Berl).* **13**, 169-91.

Parquet MC, Kumatori A, Hasebe F, Morita K and Igarashi A. (2001). West Nile virus-induced bax-dependent apoptosis. *FEBS Lett.* **500**, 17-24.

Perelygin AA, Scherbik SV, Zhulin IB, Stockman BM, Li Y and Brinton MA. (2002). Positional cloning of the murine flavivirus resistance gene. *Proc Natl Acad Sci* **99**, 9322-7.

Prince VE and FB., P. (2002). Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet.* **3**, 827-37.

Pryor MJ, Gualano RC, Lin B, Davidson AD and Wright PJ. (1998). Growth restriction of dengue virus type 2 by site-specific mutagenesis of virus-encoded glycoproteins. *J Gen Virol.* **79**, 2631-9.

Rebouillat D and Hovanessian AG. (1999). The human 2',5'-oligoadenylate synthetase family: interferon-induced proteins with unique enzymatic properties. *J. Interferon. Cytokine Res.* **19**, 295-308.

Rhodes-Feuillette, A., Lasneret, J., Paulien, S., Ogunkolade, W., Peries, J. and Canivet, M. (1990). Effects of human recombinant alpha and gamma and of highly purified natural beta interferons on simian Spumavirinae prototype (simian foamy virus 1) multiplication in human cells. *Res Virol* **141**, 31-43.

Rowe, T., Gao, G., Hogan, R. J., Crystal, R. G., Voss, T. G., Grant, R. L., Bell, P., Kobinger, G. P., Wivel, N. A. and Wilson, J. M. (2004). Macaque model for severe acute respiratory syndrome. *J Virol* **78**, 11401-4.

Rozas, J., Sanchez-DelBarrio, J., Messeguer, X. and Rozas, R. (2005). DnaSP version 4.10.4, (ed. U. d. Barcelona).

Ryter JM and Schultz SC. (1998). Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J.* **17**, 7505-13.

Sabin AB. (1952). Genetic, hormonal and age factors in natural resistance to certain viruses. *Ann N Y Acad Sci.* **6**, 936-44.

Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C. et al. (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722-5.

Samuel CE. (2002). Host genetic variability and West Nile virus susceptibility. *Proc Natl Acad Sci* **99**, 11555-7.

Sangster MY, Heliams DB, MacKenzie JS and Shellam GR. (1993). Genetic studies of flavivirus resistance in inbred strains derived from wild mice: evidence for a new resistance allele at the flavivirus resistance locus (Flv). *J Virol.* **67**, 340-7.

Sangster MY, Mackenzie JS and Shellam GR. (1998). Genetically determined resistance to flavivirus infection in wild Mus musculus domesticus and other taxonomic groups in the genus Mus. *Arch Virol.* **143**, 697-715.

Sangster MY, Urosevic N, Mansfield JP, Mackenzie JS and Shellam GR. (1994). Mapping the Flv locus controlling resistance to flaviviruses on mouse chromosome 5. *J Virol.* **68**, 448-52.

Santiago, M. L., Lukasik, M., Kamenya, S., Li, Y., Bibollet-Ruche, F., Bailes, E., Muller, M. N., Emery, M., Goldenberg, D. A., Lwanga, J. S. et al. (2003). Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (Pan troglodytes schweinfurthii). *J Virol* **77**, 7545-62.

Sarkar SN, Ghosh A, Wang HW, Sung SS and Sen GC. (1999). The nature of the catalytic domain of 2'-5'-oligoadenylate synthetases. *J Biol Chem.* **36**, 25535-42.

Sarkar SN, B. S., Ghosh A, Sen GC. (1999). Enzymatic characteristics of recombinant medium isozyme of 2'-5' oligoadenylate synthetase. *J Biol Chem.* **3**, 1848-55.

Sarkar, S. N., Miyagi, M., Crabb, J. W. and Sen, G. C. (2002). Identification of the substrate-binding sites of 2'-5'-oligoadenylate synthetase. *J Biol Chem* **277**, 24321-30.

Sarkar, S. N. and Sen, G. C. (1998). Production, purification, and characterization of recombinant 2', 5'-oligoadenylate synthetases. *Methods* **15**, 233-42.

Savage HM, N. M., Smith GC, Mitchell CJ, Craig GB Jr. (1993). Host-feeding patterns of Aedes albopictus (Diptera: Culicidae) at a temperate North American site. *J Med Entomol.* **30**, 27-34.

Scherbik SV, Paranjape JM, Stockman BM, Silverman RH and MA., B. (2006). RNase L plays a role in the antiviral response to West Nile virus. *J Virol* **80**, 2987-99.

Scherbik SV, Stockman BM and MA, B. (2007). Differential expression of interferon (IFN) regulatory factors and IFN-stimulated genes at early times after West Nile virus infection of mouse embryo fibroblasts. *J Virol.* **81**, 12005–12018.

Schroder, H., Kelve, M. and Muller, W. (1994). The 2-5A system and HIV infection. *Prog Mol Subcell Biol.* **14**, 176-197.

Shellam GR, S. M., Urosevic N. (1998). Genetic control of host resistance to flavivirus infection in animals. *Rev Sci Tech.* **17**, 231-48.

Simmonds, P. (2001). Reconstructing the origins of human hepatitis viruses. *Philos Trans R Soc Lond B Biol Sci* **356**, 1013-26.

Spanggord, R. J. and Beal, P. A. (2000). Site-specific modification and RNA crosslinking of the RNA-binding domain of PKR. *Nucleic Acids Res* **28**, 1899-905.

Speight G, Coia G, Parker MD and Westaway EG. (1988). Gene mapping and positive identification of the non-structural proteins NS2A, NS2B, NS3, NS4B and NS5 of the flavivirus Kunjin and their cleavage sites. *J Gen Virol.* **69**, :23-34.

Stadler K, Allison SL, Schalich J and Heinz FX. (1997). Proteolytic activation of tick-borne encephalitis virus by furin. *J Virol.* **71**, 8475-81.

Starkman, S. E., MacDonald, D. M., Lewis, J. C., Holmes, E. C. and Simmonds, P. (2003). Geographic and species association of hepatitis B virus genotypes in non-human primates. *Virology* **314**, 381-93.

Stefl R, Skrisovska L and FH., A. (2005). RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO J.* **6**, 33-8.

Kalayanarooj S., Endy T. P., Libraty D. H., Nisalak A., Innis, B. L. et al. (2002). HLA-A and -B allele associations with secondary dengue virus infections correlate with disease severity and the infecting viral serotype in ethnic Thais. *Tissue Antigens* **60**, 309-18.

Stephens, M., Smith, N. J. and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-89.

Stone, A. C., Griffiths, R. C., Zegura, S. L. and Hammer, M. F. (2002). High levels of Y-chromosome nucleotide diversity in the genus Pan. *Proc Natl Acad Sci* **99**, 43-48.

Bhullar V., Beer, B. E., Vallet, D., Gautier-Hion, A. et al. (2005). Ancient co-speciation of simian foamy viruses and primates. *Nature* **434**, 376-80.

Taguchi T, Nagano-Fujii M, Akutsu M, Kadoya H, Ohgimoto S, Ishido S and Hotta H. (2004). Hepatitis C virus NS5A protein interacts with 2',5'-oligoadenylate synthetase and inhibits antiviral activity of IFN in an IFN sensitivity-determining region-independent manner. *J Gen Virol.* **85**, 959-69.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

Takeda, T., Kuroki, T., Fukuda, K., Yabusako, T., Nishiguchi, S., Nakajima, S., Shiomi, S., Seki, S., Monna, T., Sakurai, M. et al. (1993). Long-term therapeutic

efficacy of interferon for patients with chronic hepatitis C. *Gastroenterol Jpn* **28 Suppl 5**, 104-8.

Tan BH, Fu J, Sugrue RJ, Yap EH, Chan YC and Tan YH. (1996). Recombinant dengue type 1 virus NS5 protein expressed in Escherichia coli exhibits RNA-dependent RNA polymerase activity. *Virology* **216**, 317-25.

Tan, Z., Shon, A. M. and Ober, C. (2005). Evidence of balancing selection at the HLA-G promoter region. *Hum Mol Genet* **14**, 3619-28.

Taylor, M. D., Korth, M. J. and Katze, M. G. (1998). Interferon treatment inhibits the replication of simian immunodeficiency virus at an early stage: evidence for a block between attachment and reverse transcription. *Virology* **241**, 156-62.

Trent DW, Kinney RM, Johnson BJ, Vorndam AV, Grant JA, Deubel V, Rice CM and Hahn C. (1987). Partial nucleotide sequence of St. Louis encephalitis virus RNA: structural proteins, NS1, ns2a, and ns2b. *Virology* **156**, 293-304.

Urosevic, N. (2003). Is flavivirus resistance interferon type I-independent? *Immunol Cell Biol* **81**, 224-9.

Urosevic N, Silvia OJ, Sangster MY, Mansfield JP, Hodgetts SI and Shellam GR. (1999). Development and characterization of new flavivirus-resistant mouse strains bearing Flv(r)-like and Flv(mr) alleles from wild or wild-derived mice. *J Gen Virol.* **80**, 897-906.

Vainio J and Cutts F. (1998). Yellow fever. *World Health Organization*, 9-45.

Waked, I., Amin, M., Abd-el Fattah, S. A., Osman, L. and Sabbour, M. (1988). Interferon in chronic active hepatitis B. Preliminary results. *Chemioterapia* **7**, 198-202.

Warke, R. V., Xhaja, K., Martin, K. J., Fournier, M. F., Shaw, S. K., Brizuela, N., de Bosch, N., Lapointe, D., Ennis, F. A., Rothman, A. L. et al. (2003). Dengue virus induces novel changes in gene expression of human umbilical vein endothelial cells. *J Virol* **77**, 11822-32.

Webster LT. (1937). Inheritance of resistance of mice to enteric bacterial and neurotropic virus infections. *J. Exp. Med.* **65**, 261-286.

Wengler G and Wengler G. (1993). The NS 3 nonstructural protein of flaviviruses contains an RNA triphosphatase activity. *Virology* **197**, 265-73.

Westaway EG, Brinton MA, Gaidamovich SY, Horzinek MC, Igarashi A, Kaariainen L, Lvov DK, Porterfield JS, Russell PK and Trent DW. (1985). Flaviviridae. *Intervirology.* **24**, 183-92.

WHO. (1997). Dengue Haemorrhagic Fever, Diagnosis, Treatment, Prevention & Control. 18–23.

Wieland, S., Thimme, R., Purcell, R. H. and Chisari, F. V. (2004). Genomic analysis of the host response to hepatitis B virus infection. *Proc Natl Acad Sci U S A* **101**, 6669-74.

Wiens M, Kuusksalu A, Kelve M and Muller WE. (1999). Origin of the interferon-inducible (2'-5')oligoadenylate synthetases: cloning of the (2'-5')oligoadenylate synthetase from the marine sponge Geodia cydonium. *FEBS Lett.* **462**, 12-8.

Woolhouse, M. E., Webster, J. P., Domingo, E., Charlesworth, B. and Levin, B. R. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet* **32**, 569-77.

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

Yan, W., Ma, L., Stein, P., Pangas, S. A., Burns, K. H., Bai, Y., Schultz, R. M. and Matzuk, M. M. (2005). Mice deficient in oocyte-specific oligoadenylate synthetase-like protein OAS1D display reduced fertility. *Mol Cell Biol* **25**, 4615-24.

Yan W, M. L., Stein P, Pangas SA, Burns KH, Bai Y, Schultz RM, Matzuk MM. (2005). Mice deficient in oocyte-specific oligoadenylate synthetase-like protein OAS1D display reduced fertility. *Mol Cell Biol.* **25**, 4615-24.

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

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

Yonekawa, H. and Moriwaki, K. (1986). [Origins and development of laboratory mice: application of wild mice to breeding of new laboratory strains]. *Tanpakushitsu Kakusan Koso* **31**, 1151-70.

Yu, N., Jensen-Seaman, M. I., Chemnick, L., Kidd, J. R., Deinard, A. S., Ryder, O., Kidd, K. K. and Li, W.-H. (2003). Low nucleotide diversity in chimpanzees and bonobos. *Genetics* **164**, 1511-1518.

Yue D, Maizels N and Weiner AM. (1996). CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile Sulfolobus shibatae. *RNA* **2**, 895-908.

Zanotto PM, G. E., Gao GF, Harvey PH, Holmes EC. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci* **93**, 548-53.

Zeng L, Falgout B and Markoff L. (1998). Identification of Specific Nucleotide Sequences within the Conserved 3'-SL in the Dengue Type 2 Virus Genome Required for Replication. *Journal of Virology* **72**, 7510-7522.

Zhang Y, C. J., Chipman PR, Zhang W, Pletnev SV, Sedlak D, Baker TS, Strauss JH, Kuhn RJ, Rossmann MG. (2003). Structures of immature flavivirus particles. *EMBO J.* **22**, 2604-13.

Zhang, Y., Corver, J., Chipman, P. R., Zhang, W., Pletnev, S. V., Sedlak, D., Baker, T. S., Strauss, J. H., Kuhn, R. J. and Rossmann, M. G. (2003). Structures of immature flavivirus particles. *EMBO J* **22**, 2604-13.

Zhou, A., Paranjape, J., Brown, T. L., Nie, H., Naik, S., Dong, B., Chang, A., Trapp, B., Fairchild, R., Colmenares, C. et al. (1997). Interferon action and apoptosis are defective in mice devoid of 2',5'-oligoadenylate-dependent RNase L. *EMBO J* **16**, 6355-63.