



Published in final edited form as:

Gene. 2007 August 1; 397(1-2): 38–50.

Evolution of the interferon alpha gene family in eutherian mammals

Christopher H. Woelk^{1,*}, Simon D. W. Frost¹, Douglas D. Richman^{1,2,3}, Prentice E. Higley¹, and Sergei L. Kosakovskiy¹

¹ Department of Pathology, University of California San Diego, La Jolla, CA, 92093, USA

² Department of Medicine, University of California San Diego, La Jolla, CA, 92093, USA

³ Veterans Affairs San Diego Healthcare System, San Diego, 92161, CA, USA

Abstract

Interferon alpha (IFNA) genes code for proteins with important signaling roles during the innate immune response. Phylogenetically, IFNA family members in eutherians (placental mammals) cluster together in a species-specific manner except for closely related species (i.e. *Homo sapiens* and *Pan troglodytes*) where gene-specific clustering is evident. Previous research has been unable to clarify whether gene conversion or recent gene duplication accounts for gene-specific clustering, partly because the similarity of members of the IFNA family within species has made it historically difficult to identify the exact composition of IFNA gene families. IFNA gene families were fully characterized in recently available genomes from *Canis familiaris*, *Macaca mulatta*, *Pan troglodytes* and *Rattus norvegicus*, and combined with previously characterized IFNA gene families from *Homo sapiens* and *Mus musculus*, for the analysis of both whole and partial gene conversion events using a variety of statistical methods. Gene conversion was inferred in every eutherian species analyzed and comparison of the IFNA gene family locus between primate species revealed independent gene duplication in *M. mulatta*. Thus, both gene conversion and gene duplication have shaped the evolution of the IFNA gene family in eutherian species. Scenarios may be envisaged whereby the increased production of a specific IFN- α protein would be beneficial against a particular pathogenic infection. Gene conversion, similar to duplication, provides a mechanism by which the protein product of a specific IFNA gene can be increased.

Keywords

Interferon alpha; gene conversion; gene duplication; innate immunity

1. Introduction

Factors induced by cytokines encoded by the interferon alpha (IFNA) gene family establish the first wave of the innate immune defense against pathogen infection (Levy and Garcia-Sastre, 2001). Following established nomenclature (Allen and Diaz, 1994), IFNA genes code

* Corresponding author: Christopher H. Woelk. University of San Diego California, Department of Pathology Stein Clinical Research Building, Rm. 326, 9500 Gilman Drive # 0679, La Jolla CA 92093-0679, Tel: 858-552-8585 extn 7193, Fax: 858-552-7445, E-mail: cwoelk@ucsd.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

for IFN- α proteins, which are classified as Type I interferons (IFNs). In eutherians (placental mammals), a number of related cytokines have also been classified as Type I IFNs and fall into an additional six classes: IFN- β , IFN- δ , IFN- ϵ , IFN- κ , IFN- τ , IFN- ω and IFN- ζ (also known as limitin). In humans, the IFN- β , IFN- ϵ , IFN- κ and IFN- ω classes are represented by a single functional gene whereas IFN- α consists of a 13 member multigene family (Pestka et al., 2004). Members of the IFNA gene family in different species normally reside in close proximity within a single chromosome. Other classes of Type I IFN are not found in humans but are present in pigs (IFN- δ), ruminants (IFN- τ) and mice (IFN- ζ). It has been postulated that infection by a particular pathogen results in the induction of specific IFNs whereas other IFNs (i.e. IFN- δ and - τ) have diverged to such an extent that they are no longer involved in immunity (Diaz et al., 1994; Harrison et al., 2003).

Type I IFNs are predominantly produced by leukocytes in response to virus infection, the presence of double stranded RNA (dsRNA), or the recognition of pathogen associated molecular patterns by Toll-like receptors (Goodbourn et al., 2000; Ozato et al., 2002; Woelk et al., 2004; Garcia-Sastre and Biron, 2006). IFN- α is synthesized as a non-glycosylated monomer, which is secreted from the cell and binds the Type I IFN receptor (Stark et al., 1998). Receptor binding leads to intracellular signaling predominantly via the Jak-Stat pathway resulting in the induction of hundreds of interferon stimulated genes (ISGs) (Der et al., 1998). The protein products of ISGs have multiple pleiotropic effects that are capable of creating an antiviral state (Kornbluth et al., 1989; Goodbourn et al., 2000), preventing proliferation (Sen, 2001) and stimulating the adaptive immune response (Levy and Garcia-Sastre, 2001). In the human host, the culmination of the interferon response often results in flu-like symptoms, diarrhea, fatigue and depression (Sleijfer et al., 2005).

The best-characterized antiviral proteins of the interferon mediated innate immune response include protein kinase R, the 2'-5' oligoadenylate synthetase/RNaseL system and the Mx proteins (reviewed in Stark et al., 1998). However, increased study of interferon mediated immunity has identified a number of other proteins with antiviral properties including: ISG15 (Kunzi and Pitha, 1996), ISG20 (Espert et al., 2003) and APOBEC3G (Chen et al., 2006). More undoubtedly exist. For an intracellular pathogen to establish a successful infection, it must circumvent the interferon response. Viruses have been shown to interfere with the functioning of the IFN system at many levels including IFN signaling, dsRNA signaling, induction of ISGs and the functioning of the induced proteins (Sen, 2001).

Gene duplication and gene conversion are evolutionary processes that affect multigene families like the Type I IFNs (Ohno, 1970; Baltimore, 1981; Miyata et al., 1985; Hurles, 2004). Gene duplication generates a new gene locus but gene conversion does not. A gene can be duplicated by retrotransposition, by segmental duplication as a result of homologous or non-homologous recombination, or by duplication of a whole chromosome or genome (Hurles, 2004). A duplicated gene may: (1) remain similar to the progenitor if increased copy number is beneficial, (2) acquire beneficial mutations and derive new functionality (neofunctionalization), (3) undergo evolution in the promoter region leading to changes in gene regulation (subfunctionalization), or (4) acquire disablements through deleterious mutation and become a pseudogene. Homologous recombination may also lead to gene conversion, which is the non-reciprocal transfer of sequence data, where a gene sequence is partially or completely replaced by the corresponding sequence of another gene (Baltimore, 1981; Hurles, 2004). Gene conversion can contribute to concerted evolution by homogenizing paralogous genes so that gene family members may be highly similar within species but continue to diverge between species. Miyata et al. (1985) identified four different types (I–IV) of gene conversion. Since, IFNA genes are intron-less in every species analyzed to date they are only subjected to whole (Type I) and partial (Type III) gene conversion.

In the 1980s, evolutionary analysis of human and mouse IFNA genes favored the hypothesis of gene conversion (Weissmann et al., 1982; Gillespie and Carter, 1983; Shaw et al., 1983; Wilson et al., 1983; Todokoro et al., 1984; Golding and Glickman, 1985) over recent duplication (Miyata and Hayashida, 1982) to explain the species-specific clustering of this family. However, a subsequent analysis by Miyata et al. (1985) suggested that these two evolutionary mechanisms are not mutually exclusive and could both contribute to the evolution of IFNA gene families. The last study to examine this issue failed to find evidence of gene conversion in humans but did find limited gene conversion in mouse (Hughes, 1995). It was concluded that species-specific clustering was the result of recent gene duplication, which had occurred independently in each eutherian species, and that any role of gene conversion was minor. These studies suffered from a number of problems: (1) analysis of gene families from a single species, (2) use of IFNA multigene families that had not been fully defined for the species under analysis, and (3) inclusion of duplicate genes and/or allelic variants (Miyata and Hayashida, 1982; Hughes, 1995).

To resolve the respective contributions of gene conversion and gene duplication to the evolution of the IFNA gene families in eutherian species, we have carefully curated currently available IFNA genes. Genomic sequences were used to fully resolve the IFNA gene family members from chimpanzee (*Pan troglodytes*, *Ptr*), dog (*Canis familiaris*, *Cfa*), human (*Homo sapiens*, *Hsa*), mouse (*Mus musculus*, *Mmus*), rat (*Rattus norvegicus*, *Rno*) and rhesus macaque (*Macaca mulatta*, *Mmul*), to help overcome the limitations of previous studies. A number of statistical methods were applied to detect instances of partial gene conversion and a novel approach using phylogenies of the MER106B repeat element was used to identify whole gene conversion events. Although gene duplication undoubtedly led to the creation of eutherian IFNA gene families, gene conversion plays a ubiquitous role in their evolution and has contributed to both creating and maintaining species-specific phylogenetic clustering.

2. Materials and methods

2.1. Sequence curation

Sequence data for IFNs was retrieved from a variety of sources including: genome repositories (i.e. University of California, Santa Cruz Genome Bioinformatics, <http://genome.ucsc.edu/>), manual transcription directly from published material (i.e. CriIFNA3, Bluysen et al., 1995), and from the NCBI GenBank and Trace Archives (EST, WGS and Other). The IFNA gene family from human and mouse have already been characterized extensively (Pestka et al., 2004; Van Pesch et al., 2004). The availability of sequenced genomes for the chimpanzee, dog, rat and rhesus macaque, facilitated characterization of IFNA gene families from these species. The availability of finalized genomes allows the exact number of genes comprising the IFNA family to be determined, and the assignment of allelic variants that were previously identified as novel IFNA genes to the same gene position in the IFNA family under study (Van Pesch et al., 2004).

A conservative approach was adopted for including IFNA genes from species whose genomes have not been sequenced. In general, sequence of an IFNA gene was only included if it: (1) clustered with an IFNA from a species whose gene family had been fully characterized, (2) was derived from a publication adopting rigorous sequencing controls (Auricchio et al., 2001), or (3) had supporting sequence data that could be derived from an alternative source (i.e. the same sequence could be identified from a GenBank entry and from a genome project in the scaffold stage). Sequence data from the NCBI Trace Archives were treated with caution after observing a number of inaccurate sequence entries in reference to fully characterized IFNA gene families from species with sequenced genomes. In several cases, multiple variants of a particular IFNA sequence were removed from analysis since it was impossible to assess their validity without a fully sequenced genome for the species in question (see Text S1 in the

supplemental material). Exceptions to this rule included an IFNA21 variant designate IFNA21.1 whose sequence was confirmed in two different orangutan species, *Pongo pygmaeus pygmaeus* and *P. pygmaeus abelii*. Similarly, IFNA23 from baboon was confirmed using sequence data from two different species (*Papio anubis* and *P. cynocephalus*). The two IFNA2 variants from the cotton-top Tamarin (*Saguinus oedipus*) designated SoeIFNA2a and 2b were also included since their validity was confirmed through multiple separate cloning experiments (Auricchio et al., 2001). A complete description of the curation procedure for each species analyzed in this paper can be found in the supplemental material.

A total of 156 IFNA sequences were curated for phylogenetic and gene conversion analysis. An additional 6 data sets from species with complete genomes and hence fully delineated IFNA gene families were compiled for analysis of gene conversion: chimpanzee (N=12), dog (N=9), human (N=13), mouse (N=14), rat (N=18), and rhesus macaque (N=14). All alignments are available as NEXUS files for download at <http://www.hyphy.org/pubs/IFN/>.

2.2. Phylogenetic reconstruction

IFNA gene and MER106B repeat element sequence data was aligned using ClustalX (Chenna et al., 2003) and the resulting alignment edited manually using Se-AL (<http://evolve.zoo.ox.ac.uk/>). Phylogenetic trees were constructed using PhyML version 2.4.4 (Guindon and Gascuel, 2003) and the HKY85 substitution model (Hasegawa et al., 1985) with among-site rate variation described through a discrete approximation of the gamma distribution with four rate categories. For each tree, one thousand bootstrap samples were generated and evaluated by PhyML under the same model. MER106B phylogenies were rooted using sequence data obtained from RepBase (<http://www.girinst.org/repbase/update/index.html>) and human interferon omega (HsaIFNW) was used as an outgroup for IFNA gene phylogenies based on prior analysis by Pestka et al. (2004).

2.3. Detection of partial gene conversion

Initially, our Genetic Algorithm for Recombination Detection (GARD, <http://www.datamonkey.org/GARD/>) was used to detect the presence or absence of recombination in the data sets as described previously (Kosakovsky Pond et al., 2006a; Kosakovsky Pond et al., 2006b). GARD has been shown to outperform existing methods in terms of both Type I and Type II errors when detecting the presence of recombination. This method marries a maximum likelihood (ML) model based framework with a genetic algorithm to detect discordant phylogenetic signals in DNA sequence alignments and estimate the number and location of breakpoints. GARD uses an aggressive population-based hill climber to quickly search the space of locations of possible recombination breakpoints in a given sequence alignment (Eshelman, 1991; Kosakovsky Pond and Frost, 2005). In brief, for each data set, GARD starts with 0 breakpoints and the number of breakpoints is increased by 1 for subsequent GA runs until the AIC_c score of the best model stops decreasing as a function of the number of breakpoints.

A one-tailed KH test was used to assign further statistical significance to the signal of recombination identified by GARD (Kishino and Hasegawa, 1989). The KH test was implemented under the HyPhy package (Kosakovsky Pond et al., 2005). For a given data set, phylogenies were constructed from the partitions either side of the breakpoint using the HKY85 model of nucleotide substitution. Incongruence in topology was then tested in both directions using 1000 bootstrap replicates.

IFNA data sets were further analyzed using GENECONV version 1.81 (Sawyer, 1989), which is a well-established method for detecting partial gene conversion (Posada, 2002). GENECONV implements an algorithm that determines whether some fragment that is shared

by a pair of DNA sequences in a multiple alignment has more consecutive identical silent polymorphic sites in common than would be expected by chance. GENECONV assess the significance of fragments that have potentially undergone gene conversion using both pairwise and global p -values. Global p -values compare each fragment with all possible fragments for the entire multiple alignment and are taken as primary evidence for gene conversion since they aim to eliminate the false detection of similarities that may have arisen by chance. Global simulated p -values based on 10,000 permutations have a built-in multiple-comparison correction for all sequence pairs in the alignment. Bonferroni-corrected global Karlin-Altschul p -values are generally more conservative and are 100–1000 times as large as global simulated p -values. Although global p -values are normally superior to pairwise p -values, in some cases Bonferroni-corrected pairwise simulated p -values are more significant according to the GENECONV documentation (<http://www.math.wustl.edu/~sawyer/geneconv/gconvdoc.html>). All three types of p -value were recorded during the analysis of IFNA gene families and strong evidence of gene conversion noted when a fragment had a p -value that was less than 0.05 for at least two different types of statistical test. All polymorphic sites were tested for evidence of gene conversion using mismatch penalties (g -scale) of 0, 1 or 2. A penalty of 0 prohibits shared fragments from having internal mismatches whereas more ancient gene conversion events can be detected with penalties of 1 and 2 because this allows fragments to accumulate mutations following the gene conversion event.

Two control analyses were performed with the GENECONV package. Analysis was repeated using the `-randomize_sites` option, which permutes polymorphic sites. In this case few or no fragments should exhibit a significant p -value indicating gene conversion. If a fragment identified previously was also identified under the `-randomize_sites` option it was removed from the results. For the second control, GENECONV was rerun on IFNA alignments after those sequences harboring fragments with a significant signal for gene conversion were removed. If all fragments that underwent a gene conversion event had been identified previously then this repeated analysis should not identify new fragments.

2.4. Detection of whole gene conversion

Serendipitously, the 5' non-coding region of IFNA genes from the chimpanzee, dog, human and rhesus macaque genome contained a copy of the conserved repeat element MER106B. The most parsimonious explanation for such a coordinated relationship of MER106B with IFNA genes is that this repeat element was duplicated along with the gene sequence during the expansion of the gene family. Therefore, evidence of whole gene conversion was identified by locating significant incongruence between the IFNA and MER106B phylogenetic trees, where discordant clades had high bootstrap support (>75) in each tree. Bootstrapped ML phylogenetic trees were constructed from IFNA gene and MER106B repeat element alignments using the methods and parameters already described. To maximize the MER106B alignment used for phylogenetic reconstruction, HsaMER2, PtrMER2 and MmulMER2 were removed because they represent a small fragment of the complete MER106B repeat element. Correspondingly, their gene equivalents (HsaIFNA2, PtrIFNA2 and MmulIFNA2) were also removed from the IFNA alignment to facilitate straightforward comparison of the two trees. PtrIFNA8 was removed from analysis since it was a pseudogene in the chimpanzee genome (see supplemental material for more details).

2.5. Synteny evaluation

Advanced PipMaker (<http://pipmaker.bx.psu.edu/cgi-bin/pipmaker?advanced>) was used to align both genic and intergenic regions of the chimpanzee and rhesus macaque IFNA gene family locus to the human locus (Schwartz et al., 2000). Dot plots were obtained using “*search one strand*” and “*single coverage*” options. All other parameters were set to their defaults.

3. Results and discussion

3.1. Eutherian IFNA phylogenetic analysis

Pestka et al. (2004) recently performed a phylogenetic analysis of all classes of Type I IFN, which provided a good starting point for examining the evolution of eutherian IFNA genes. We have improved upon their phylogenetic analysis of IFNA genes by: (1) using sequenced genomes to fully characterize IFNA gene families for chimpanzee, dog, rat and rhesus macaque (Figure 1), (2) removing allelic variants, erroneous sequences and duplicate genes, (3) adding new IFNA gene sequence data for the cat (*Felis catus*), hamster (*Cricetidae sp.*), Himalayan woodchuck (*Marmota himalayana*), and two species of orangutan (Figure 1), and by (4) providing reference to the source from which IFNA sequence data were obtained (i.e. GenBank accession number or genome position, Table S1 in the supplemental material). Using HsaIFNW as an outgroup (Pestka et al., 2004), phylogenetic analysis of 156 IFNA genes from 17 eutherian species confirmed that IFNA genes form species-specific clusters except for the primate section of the tree (Figure 1). For primate species, the majority of IFNA genes form gene-specific clusters although a few instances of species-specific clustering are still evident (PtrIFNA4 and 10, HsaIFNA4 and 10, HsaIFNA1 and 13, MmulIFNA1 and 13, MmulIFNA23-29). The longest IFNA (603 nts) belongs to the woodchuck (*Marmota monax*) and the shortest (543 nts) to the pig (*Sus scrofa*).

Due to low divergence of IFNA genes within species, genome assemblies are required to identify all members of the IFNA gene family and discriminate duplicate GenBank entries and allelic variants. We encourage this type of curation to be repeated whenever new species genomes become available in the future. Of the species whose IFNA gene family we fully characterized, the rat has the largest family consisting of 18 members that separate into two clusters separated by high bootstrap support (Figure 1). The most obvious sequence difference between these two clusters is the deletion of an “RNKRAF” motif in the cluster containing RnoIFNA5, 10, and 13–18. In addition, the rat genome at the UCSC Genome Bioinformatics browser (<http://genome.ucsc.edu/>) and an independent version at the Medical College of Wisconsin, designated the Rat Genome Database (<http://rgd.mcw.edu/>), both indicate a small IFNA cluster on chromosome 3 consisting of RnoIFNA2, 4, and a pseudogene, which is an exact duplicate of the same region on chromosome 5.

3.2. Partial gene conversion

Gene conversion analysis was focused on species (chimpanzee, dog, human, mouse, rat and rhesus macaque) whose genome sequence was available and hence whose IFNA gene family had been fully characterized. Prominent reviews assessing the performance of algorithms and statistical tests used to detect gene conversion suggest that no single method should be relied upon (Drouin et al., 1999; Posada, 2002). Therefore a number of methods were used starting with GARD (Kosakovsky Pond et al., 2006a), which found a signal of phylogenetic incongruence in every data set except for dog. In the other species the introduction of breakpoints produced a model with lower small sample Akaike Information Criterion (AIC_c) scores (Figure 2). The lowest AIC_c score recorded for the mouse IFNA gene family corresponded to a model with 4 breakpoints, while the chimpanzee, human, rat and rhesus macaque IFNA gene families have 2 breakpoints.

GARD can find significant breakpoints not only when there is a change in tree topology across a putative breakpoint, but also if branch lengths are sufficiently different on either side of the breakpoint. This could be caused, for instance, by local heterotachy or an elevation/reduction in substitution rates. To confirm the significance of the signal of gene conversion in the chimpanzee, human, mouse, rat and rhesus macaque data sets, separate phylogenetic trees were inferred from alignment partitions either side of the breakpoint as estimated by GARD. Tree

topologies were compared using a one-tailed Kishino-Hasegawa (KH) test (Table 1). The null hypothesis that the topologies were the same and no recombination had occurred between partitions was rejected (p -value <0.01) for every data set when comparing partitions in both directions. Therefore, additional statistical significance can be assigned to the signal of gene conversion detected by GARD analysis. Unfortunately, there were too many IFNA sequences in our total data set ($N=156$) for computationally efficient GARD analysis. However, it could be subjected to single breakpoint analysis and showed an AIC score improvement of 777.274 with a breakpoint at nucleotide position 351. The phylogenetic trees generated separately from partitions around this breakpoint were also significantly different from each other as assessed by the KH test (Table 1).

Specific sequences and fragments involved in gene conversion events were then identified in IFNA gene data sets (Table 2) using GENECONV (Sawyer, 1989). Mismatch penalties were adjusted so that results were obtained for both recent and ancient gene conversion events. It is immediately obvious from Table 2 that every IFNA gene data set (except dog) contains sequences with shared fragments that had a significant signal for gene conversion. Previous analyses indicating that HsaIFNA14 and MmusIFNA11 were involved in a partial gene conversion event were supported (Weissmann et al., 1982; Golding and Glickman, 1985; Miyata et al., 1985; Hughes, 1995). The rhesus macaque IFNA gene family appears to have the greatest number of partial gene conversion events with the fragment shared between MmulIFNA14 and 23 exhibiting a p -value <0.05 for all three types of statistical test. Furthermore, there are many instances of gene conversion between IFNA sequences that were only supported by one of the three statistical tests. The fragments identified in these cases should still be considered candidates for having undergone gene conversion. Inferred pairwise gene conversions may be difficult to assemble into a coherent picture. For example, HsaIFNA14 may have partially converted the ancestor of HsaIFNA10, 17 and 4 over the 347–567 fragment, and HsaIFNA10 may have more recently converted HsaIFNA7 (or vice versa) over the 221–415 fragment (Table 2). As the number of inferred pairwise interactions increases, individual sequences may be involved in multiple gene conversion events leading mosaic sequences.

To determine if GENECONV analysis had identified all putative gene conversion events, one of the pair of sequences identified as undergoing gene conversion was removed and the analysis repeated for each IFNA data set. In each case, apart from the mouse IFNA gene family no further recombinants were identified within the data sets (data not shown). When MmusIFNA6T, 7/10 and 9 were removed from the mouse data set, a gene conversion event was identified between MmusIFNA1 and 5 but this was only supported by a Bonferroni-corrected pairwise simulated p -value of 0.0495 (data not shown). In summary, it appears that GENECONV captured the majority of partial gene conversion events in the IFNA data sets.

3.3. Synteny evaluation

Alignment dot plots of genic and intergenic regions in the IFNA family locus generated using Advanced PipMaker (Schwartz et al., 2000) show that synteny is conserved between human and chimpanzee (Figure 3A). Each IFNA gene in human is in exactly the same position of the family locus compared to each gene in chimpanzee. This suggests that all IFNA genes existed in the MRCA of human and chimpanzee approximately five million years ago (Kumar and Hedges, 1998) and the whole family was then inherited by each species. Therefore, IFNA genes from humans and chimpanzees should cluster together in a gene-specific manner. However, as previously noted, species-specific clustering is seen for IFNA1 and 13, and for IFNA4 and 10 (Figure 1). This breakdown in gene-specific clustering can be explained by recent gene duplication events or by gene conversion. Gene duplication is unlikely because it would require that both humans and chimpanzees duplicated the same gene to exactly the same new position and this whole process needed to occur twice in each species. Gene conversion appears to be

the more parsimonious explanation for the breakdown in gene-specific clustering in these cases.

Comparison of gene synteny in the IFNA gene family locus between humans and rhesus macaque confirms that recent gene duplication does not conserve the position of IFNA genes (Figure 3B). Synteny is only conserved for a part of the IFNA gene family locus suggesting that the MRCA of human, chimpanzee and rhesus macaque (≈ 23 million years ago) contained a subset of IFNA genes, possibly IFNA1, 8, 2, 13 and 6, which then gave rise to two different IFNA gene families through species-specific duplication and gene conversion. On the branch leading to chimpanzee and human these genes duplicated to give rise to the IFNA gene family seen in each of these species today. In rhesus macaque the IFNA gene family probably expanded differently to produce MmulIFNA23-29.

3.4. Whole gene conversion

A phylogenetic tree of the MER106B repeat element, which is located at the same position upstream of every IFNA gene, was constructed for sequences from dog, human, chimpanzee and rhesus macaque, for comparison to the IFNA gene tree for these species (Figure 4). Duplication in conjunction with IFNA genes is the most parsimonious explanation for the ubiquitous presence of the MER106B repeat element since the alternative hypothesis would indicate that each repeat element independently inserted into the same position in front of each IFNA gene in each species multiple times. Furthermore, transposition of MER repeats is thought to have occurred during early mammalian radiation before the origin of the *Alu* family (Jurka et al., 1996) and thus transposition is unlikely to have resulted in the ubiquitous presence of the MER106B element in species that have diverged more recently (i.e. human, chimp and rhesus macaque). These observations made it possible to identify instances of whole gene conversion as bootstrap supported differences in topology between the MER106B repeat element and IFNA gene trees (Figure 4).

Todokoro et al. (1984) hypothesized that the similarity exhibited by HsaIFNA1 and 13 was the result of recent whole gene conversion and the primate clade of IFNA1 and IFNA13 sequences provides the best evidence of whole gene conversion in our study. Prior gene synteny analysis established that IFNA1 and 13 were present in the MRCA of humans, chimpanzee and rhesus macaque, and thus should cluster on a gene-specific basis (Figure 3). Such gene-specific clustering is confirmed in this clade with high bootstrap support when considering phylogenies of the MER106B element (Figure 4A). However, the IFNA gene phylogeny for this clade (Figure 4B) depicts bootstrap supported species-specific clustering such that HsaIFNA1 clusters with HsaIFNA13, and MmulIFNA1 with MmulIFNA13. The most parsimonious explanation is that whole gene conversion events have homogenized these gene sequences within species. Whether PtrIFNA1 and 13 have been affected by a whole gene conversion event is less clear. Another way to demonstrate whole gene conversion is to apply GARD to the concatenated sequences of the MER106B repeat element and IFNA gene sequences for IFNA1 and 13 from human, chimpanzee and rhesus macaque. If the only supported recombination breakpoint falls on the boundary between the MER106B repeat elements and the IFNA genes, then partial gene conversion in either the repeat element or the gene can be ruled out as statistically unlikely. Indeed, the only topological change is observed at the variable site nearest this boundary (AIC_c improvement of 123, KH p-value <0.01), which separates gene-specific (exhibited by MER106B) from species-specific (exhibited by IFNA) clustering (data not shown). Phylogenies for the IFNA gene and MER106B element also exhibit bootstrap supported incongruence for IFNA4, 5, and 10 from humans and chimpanzee, for IFNA14 from all three primate species, and for IFNA7 from dog. Shifts in topology for these IFNA genes may represent partial or whole gene conversion events that occurred a long time ago (i.e. IFNA5 and 14) or more recently (i.e. IFNA4 and 10). Finally, the observation that the

MER106B elements associated with MmulIFNA23 through 29 form a bootstrapped supported clade (Figure 4A) suggests that these MmulIFNA genes were created by recent gene duplication. If gene conversion had created the MmulIFNA23 through 29 cluster, then the associated MER106B elements should cluster with their human/chimp counterparts, but this is not the case.

What role could gene conversion have played in the evolution of IFNA gene families? The whole gene conversion event between HsaIFNA1 and 13 may provide a specific example (Mogensen et al., 1999). HsaIFN- α 1 has been characterized as: (1) having low antiviral activity compared to other IFN- α 's (Sperber et al., 1992; Sperber et al., 1993), (2) being the major component of the IFN mixture induced by parainfluenza viruses in peripheral blood leukocytes (Pestka, 1983), (3) having low affinity for the human IFN receptor (Uze et al., 1985), and (4) capable of binding the soluble B18R protein expressed by vaccinia virus, which is a type I IFN receptor homolog (Liptakova et al., 1997). These observations led Mogensen et al. (1999) to suggest that HsaIFN- α 1 is expressed during innate immunity in order to negate the effects of virus encoded type I IFN receptor decoy molecules. This would allow other Type I IFNs with greater antiviral activity to escape virus inhibition and signal through the cellular receptor in order to induce ISGs contributing to innate immunity. Gene conversion has effectively produced two sources of HsaIFN- α 1 (HsaIFNA1 and 13) providing a possible selective advantage to the host in the form of protection from viruses which express type I IFN receptor decoy molecules.

Little is known about the 1059 nucleotide long MER106B repeat element except for the sequence data present in RepBase (<http://www.girinst.org/repbase/update/browse.php>). In fact a PubMed search recorded no entries for this repeat element. MER106B repeat elements belong to one of forty different families exhibiting medium reiteration frequency (MER) in the human genome (Jurka et al., 1996; Smit, 1996). MER106B repeat elements were also detected upstream of IFNA genes in genomic sequence fragments from cow, horse and pig, but not in rodent lineages (data not shown). Therefore, it will be possible to identify further instances of whole gene conversion in non-rodent eutherian orders as their genomes become available. Interestingly, the *Alu* repeat element has been implicated in unequal crossing over and generation of segmental duplication due to its enrichment at the junctions between duplicated and single copy sequences (Bailey et al., 2003). Genome rearrangements have also been attributed to the MER family since MER4 was shown to mediate multigene deletion on human chromosome 14 (Jurka, 1990). Whether, the MER106B element has contributed to the evolution of IFNA gene families in similar ways remains to be determined.

A graphical depiction of the evolutionary processes affecting eutherian IFNA gene families is presented in Figure 5. This figure shows the resulting phylogenetic structure when duplication precedes speciation (Scenario 1) compared to when duplication occurs after speciation (Scenario 2). In the first scenario the phylogeny contains gene-specific clusters with mixed species. Such a phylogeny can be seen when considering the majority of IFNA genes in the primate clade (IFNA2, 5, 6, 7, 8, 14, 16, 17 and 21, Figure 1). The second scenario gives rise to phylogenies with species-specific clusters. An example of this type of clustering can be seen when comparing the human and mouse IFNA gene families where gene duplication is thought to have occurred independently in each species following speciation. However, partial and whole gene conversion following speciation has the effect of converting gene-specific clusters into species-specific clusters. Examples of this can again be seen when comparing IFNA gene families between chimpanzee and human where, in the case of IFNA1 with 13, and IFNA4 with 10, gene-specific clusters have been converted into species-specific clusters (Figure 4B).

3.5. Conclusion

Partial gene conversion was initially detected using GARD (Figure 2) and GENECONV (Table 2) in every species analyzed apart from dog. Partial and whole gene conversion events were further confirmed in every species (human, chimpanzee, rhesus macaque and dog) for which MER106B repeat elements were available for topology comparison to IFNA gene trees (Figure 4). The extent of partial and whole gene conversion detected in this study is in stark contrast to the analysis of Hughes et al. (1995) who failed to detect gene conversion in human and detected only limited gene conversion in mouse. The different results obtained by Hughes et al. (1995) may stem from the use of a human IFNA gene data set that was both incomplete and contained multiple duplicate sequences of the same gene. Furthermore, their work chose arbitrary upstream and downstream breakpoints in untranslated regions (UTRs) to investigate gene conversion, which are probably different from actual gene conversion boundaries. Therefore, our results agree best with the early conclusions of Miyata et al. (1985) positing that both duplication and gene conversion have shaped the evolution of IFNA gene families from placental mammals. However, our analyses suggest that gene conversion is operating at a much higher level than previously thought.

Gene duplication has undoubtedly played an essential role in the expansion of the IFNA gene family within species. This appears true even for relatively closely related species like chimpanzee, human and rhesus macaque. Our analysis of synteny of the IFNA gene family locus (Figure 3) and of the phylogenetic relationship of MER106B repeat elements between these species (Figure 4A) suggests that they all inherited a core of around 5 IFNA gene sequences but that this core then duplicated differently along the evolutionary path leading to rhesus macaque and the branch leading to human and chimpanzee. Further fully characterized IFNA gene families from closely related species would provide great insights into the extent of recent gene duplication. All of the fully characterize gene families presented here appear to have different numbers of IFNA genes but the same overall structure starting with a large group of IFNA genes in the reverse orientation and a smaller subset in the forward orientation (data not shown). It is plausible that the MRCA of eutherian mammals inherited a small core of IFNA genes, which then evolved differently through gene duplication, deletion, conversion and pseudogene formation to give rise to the gene families seen in each species today (Figure 1).

IFNA gene families within species must maintain antiviral signaling through a single receptor complex at the same time as avoiding pathogen countermeasures. Both gene duplication and gene conversion provide mechanisms whereby the IFNA gene family can increase copy number of useful genes and create genetic material from which new IFNA variants may arise. In addition, if a particular IFNA gene has evolved in response to the selective force of a distinct pathogenic infection, but then this selective force is removed due to pathogen eradication or migration of the host species, then gene conversion may provide a mechanism by which the IFNA can be reclaimed to produce a more useful product.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are thankful to Val Terri for useful comments pertaining to the presentation of our data and to Drs. Guy Drouin and Michael Worobey for discussions concerning gene synteny and conversion. This work was supported by grants AI043638, AI047745, AI065242 and AI07384 from the National Institutes of Health and grant IS02-SD-701 from the University of California Universitywide AIDS Research Program (UARP). The University of California San Diego Center for AIDS Research (AI36214) further supported this work via the Genomics Core and a Developmental Award to SLKP. This work was performed in space provided by the San Diego Veterans Affairs Healthcare System.

References

- Allen G, Diaz MO. Nomenclature of the human interferon proteins. *J Interferon Res* 1994;14:223–226. [PubMed: 7529811]
- Aurisicchio L, Ceccacci A, La Monica N, Palombo F, Traboni C. Tamarin alpha-interferon is active in mouse liver upon intramuscular gene delivery. *J Gene Med* 2001;3:394–402. [PubMed: 11529669]
- Bailey JA, Liu G, Eichler EE. An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* 2003;73:823–834. [PubMed: 14505274]
- Baltimore D. Gene conversion: some implications for immunoglobulin genes. *Cell* 1981;24:592–594. [PubMed: 7018693]
- Bluyssen HA, Nakamura N, Vlietstra RJ, Smit EM, Hagemeyer A, Trapman J. Isolation, properties and chromosomal localization of four closely linked hamster interferon-alpha-encoding genes. *Gene* 1995;158:295–300. [PubMed: 7607557]
- Chen K, Huang J, Zhang C, Huang S, Nunnari G, Wang FX, Tong X, Gao L, Nikisher K, Zhang H. Alpha interferon potently enhances the anti-human immunodeficiency virus type 1 activity of APOBEC3G in resting primary CD4 T cells. *J Virol* 2006;80:7645–7657. [PubMed: 16840343]
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 2003;31:3497–3500. [PubMed: 12824352]
- Der SD, Zhou A, Williams BR, Silverman RH. Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci U S A* 1998;95:15623–15628. [PubMed: 9861020]
- Diaz MO, Pomykala HM, Bohlander SK, Maltepe E, Malik K, Brownstein B, Olopade OI. Structure of the human type-I interferon gene cluster determined from a YAC clone contig. *Genomics* 1994;22:540–552. [PubMed: 8001965]
- Drouin G, Prat F, Ell M, Clarke GD. Detecting and characterizing gene conversions between multigene family members. *Mol Biol Evol* 1999;16:1369–1390. [PubMed: 10563017]
- Eshelman, LJ. The CHC adaptive search algorithm: how to do safe search when engaging in nontraditional genetic recombination. In: Rawlines, GJE., editor. *Foundations of genetic algorithms*. Morgan Kaufmann Publishers; San Mateo, CA: 1991. p. 265-283.
- Espert L, Degols G, Gongora C, Blondel D, Williams BR, Silverman RH, Mechti N. ISG20, a new interferon-induced RNase specific for single-stranded RNA, defines an alternative antiviral pathway against RNA genomic viruses. *J Biol Chem*. 2003
- Garcia-Sastre A, Biron CA. Type 1 interferons and the virus-host relationship: a lesson in detente. *Science* 2006;312:879–882. [PubMed: 16690858]
- Gillespie D, Carter W. Concerted evolution of human interferon alpha genes. *J Interferon Res* 1983;3:83–88. [PubMed: 6842040]
- Golding GB, Glickman BW. Sequence-directed mutagenesis: evidence from a phylogenetic history of human alpha-interferon genes. *Proc Natl Acad Sci U S A* 1985;82:8577–8581. [PubMed: 3866242]
- Goodbourn S, Didcock L, Randall RE. Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures. *J Gen Virol* 2000;81:2341–2364. [PubMed: 10993923]
- Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;52:696–704. [PubMed: 14530136]
- Harrison GA, Young LJ, Watson CM, Miska KB, Miller RD, Deane EM. A survey of type I interferons from a marsupial and monotreme: implications for the evolution of the type I interferon gene family in mammals. *Cytokine* 2003;21:105–119. [PubMed: 12697149]
- Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 1985;22:160–174. [PubMed: 3934395]
- Hughes AL. The evolution of the type I interferon gene family in mammals. *J Mol Evol* 1995;41:539–548. [PubMed: 7490767]
- Hurles M. Gene duplication: the genomic trade in spare parts. *PLoS Biol* 2004;2:E206. [PubMed: 15252449]
- Jurka J. Novel families of interspersed repetitive elements from the human genome. *Nucleic Acids Res* 1990;18:137–141. [PubMed: 2106661]

- Jurka J, Kapitonov VV, Klonowski P, Walichiewicz J, Smit AF. Identification of new medium reiteration frequency repeats in the genomes of Primates, Rodentia and Lagomorpha. *Genetica* 1996;98:235–247. [PubMed: 9204548]
- Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol* 1989;29:170–179. [PubMed: 2509717]
- Kornbluth RS, Oh PS, Munis JR, Cleveland PH, Richman DD. Interferons and bacterial lipopolysaccharide protect macrophages from productive infection by human immunodeficiency virus in vitro. *J Exp Med* 1989;169:1137–1151. [PubMed: 2466937]
- Kosakovsky P, Frost SD. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol* 2005;22:478–485. [PubMed: 15509724]
- Kosakovsky P, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 2005;21:676–679. [PubMed: 15509596]
- Kosakovsky P, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 2006a;23:1891–1901. [PubMed: 16818476]
- Kosakovsky P, Posada D, Gravenor MB, Woelk CH, Frost SD. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 2006b;22:3096–3098. [PubMed: 17110367]
- Kumar S, Hedges SB. A molecular timescale for vertebrate evolution. *Nature* 1998;392:917–920. [PubMed: 9582070]
- Kunzi MS, Pitha PM. Role of interferon-stimulated gene ISG-15 in the interferon-omega-mediated inhibition of human immunodeficiency virus replication. *J Interferon Cytokine Res* 1996;16:919–927. [PubMed: 8938567]
- Levy DE, Garcia-Sastre A. The virus battles: IFN induction of the antiviral state and mechanisms of viral evasion. *Cytokine Growth Factor Rev* 2001;12:143–156. [PubMed: 11325598]
- Liptakova H, Kontseikova E, Alcamí A, Smith GL, Kontsek P. Analysis of an interaction between the soluble vaccinia virus-coded type I interferon (IFN)-receptor and human IFN- α 1 and IFN- α 2. *Virology* 1997;232:86–90. [PubMed: 9185591]
- Miyata T, Hayashida H. Recent divergence from a common ancestor of human IFN- α genes. *Nature* 1982;295:165–168. [PubMed: 6173759]
- Miyata T, Hayashida H, Kikuno R, Toh H, Kawade Y. Evolution of interferon genes. *Interferon* 1985;6:1–30. [PubMed: 2410368]
- Mogensen KE, Lewerenz M, Reboul J, Lutfalla G, Uze G. The type I interferon receptor: structure, function, and evolution of a family business. *J Interferon Cytokine Res* 1999;19:1069–1098. [PubMed: 10547147]
- Ohno, S. *Evolution by gene duplication*. Springer-Verlag; Berlin, Germany: 1970.
- Ozato K, Tsujimura H, Tamura T. Toll-like receptor signaling and regulation of cytokine gene expression in the immune system. *Biotechniques Suppl* 2002;66–68. 70–72.passim
- Pestka S. The human interferons--from protein purification and sequence to cloning and expression in bacteria: before, between, and beyond. *Arch Biochem Biophys* 1983;221:1–37. [PubMed: 6187286]
- Pestka S, Krause CD, Walter MR. Interferons, interferon-like cytokines, and their receptors. *Immunol Rev* 2004;202:8–32. [PubMed: 15546383]
- Posada D. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol Biol Evol* 2002;19:708–717. [PubMed: 11961104]
- Sawyer S. Statistical tests for detecting gene conversion. *Mol Biol Evol* 1989;6:526–538. [PubMed: 2677599]
- Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res* 2000;10:577–586. [PubMed: 10779500]
- Sen GC. Viruses and interferons. *Annu Rev Microbiol* 2001;55:255–281. [PubMed: 11544356]
- Shaw GD, Boll W, Taira H, Mantei N, Lengyel P, Weissmann C. Structure and expression of cloned murine IFN- α genes. *Nucleic Acids Res* 1983;11:555–573. [PubMed: 6188104]

- Sleijfer S, Bannink M, Van Gool AR, Kruit WH, Stoter G. Side effects of interferon-alpha therapy. *Pharm World Sci* 2005;27:423–431. [PubMed: 16341948]
- Smit AF. The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev* 1996;6:743–748. [PubMed: 8994846]
- Sperber SJ, Gocke DJ, Haberzettl C, Kuk R, Schwartz B, Pestka S. Anti-HIV-1 activity of recombinant and hybrid species of interferon-alpha. *J Interferon Res* 1992;12:363–368. [PubMed: 1331260]
- Sperber SJ, Hunger SB, Schwartz B, Pestka S. Anti-rhinoviral activity of recombinant and hybrid species of interferon alpha. *Antiviral Res* 1993;22:121–129. [PubMed: 8279808]
- Stark GR, Kerr IM, Williams BR, Silverman RH, Schreiber RD. How cells respond to interferons. *Annu Rev Biochem* 1998;67:227–264. [PubMed: 9759489]
- Todokoro K, Kioussis D, Weissmann C. Two non-allelic human interferon alpha genes with identical coding regions. *Embo J* 1984;3:1809–1812. [PubMed: 6479148]
- Uze G, Mogensen KE, Aguet M. Receptor dynamics of closely related ligands: “fast” and “slow” interferons. *Embo J* 1985;4:65–70. [PubMed: 3160580]
- Van Pesch V, Lanaya H, Renaud JC, Michiels T. Characterization of the murine alpha interferon gene family. *J Virol* 2004;78:8219–8228. [PubMed: 15254193]
- Weissmann, C.; Nagata, S.; Boll, W.; Fountoulakis, M.; Fujisawa, A.; Fujisawa, J.; Haynes, J.; Henco, K.; Mantei, N.; Ragg, H.; Schein, J.; Schmid, J.; Shaw, G.; Streuli, M.; Taira, H.; Todokoro, K.; Weidle, U. Primary and tertiary structure of nucleic acids and cancer research. Miwa, M., editor. Japan Scientific Societies Press; Tokyo: 1982. p. 1-22.
- Wilson V, Jeffreys AJ, Barrie PA, Boseley PG, Slocombe PM, Easton A, Burke DC. A comparison of vertebrate interferon gene families detected by hybridization with human interferon DNA. *J Mol Biol* 1983;166:457–475. [PubMed: 6306247]
- Woelk CH, Ottone F, Plotkin CR, Du P, Royer DR, Rought SE, Lozach J, Sasik R, Kornbluth RS, Richman D, Corbeil J. Interferon gene expression following HIV type 1 infection of monocyte-derived macrophages. *AIDS Res Hum Retroviruses* 2004;20:1210–1222. [PubMed: 15588343]

Abbreviations used

| | |
|------------------------|---|
| AIC_c | small sample Akaike Information Criterion |
| dsRNA | double stranded RNA |
| IFNA | interferon alpha |
| ISG | interferon stimulated gene |
| KH | Kishino-Hasegawa |
| ML | maximum likelihood |
| MRCA | most recent common ancestor |
| UTR | untranslated region |

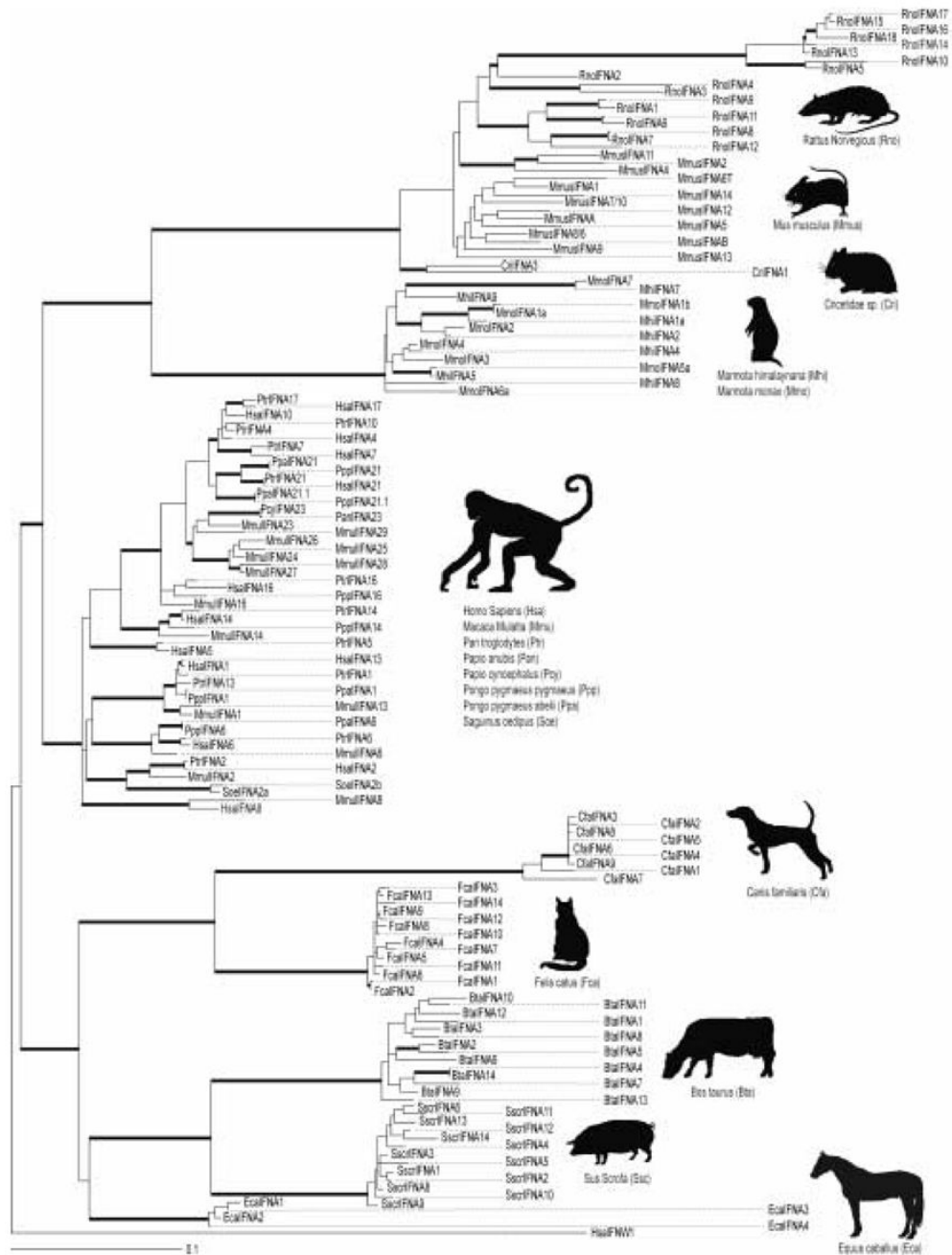


Figure 1. ML phylogenetic tree of 156 eutherian (placental mammals) IFNA gene sequences constructed using the HKY85 substitution model. The tree is rooted using human interferon omega gene (HsaIFNW) as an outgroup based off of previous analysis by Pestka et al. (2004) and all horizontal branches are drawn to scale. Thicker branches in bold refer to those with bootstrap support >75%. Dashed lines represent placeholders for gene names. Expansion of abbreviations used in the gene names is shown underneath the black and white pictogram depicting the species.

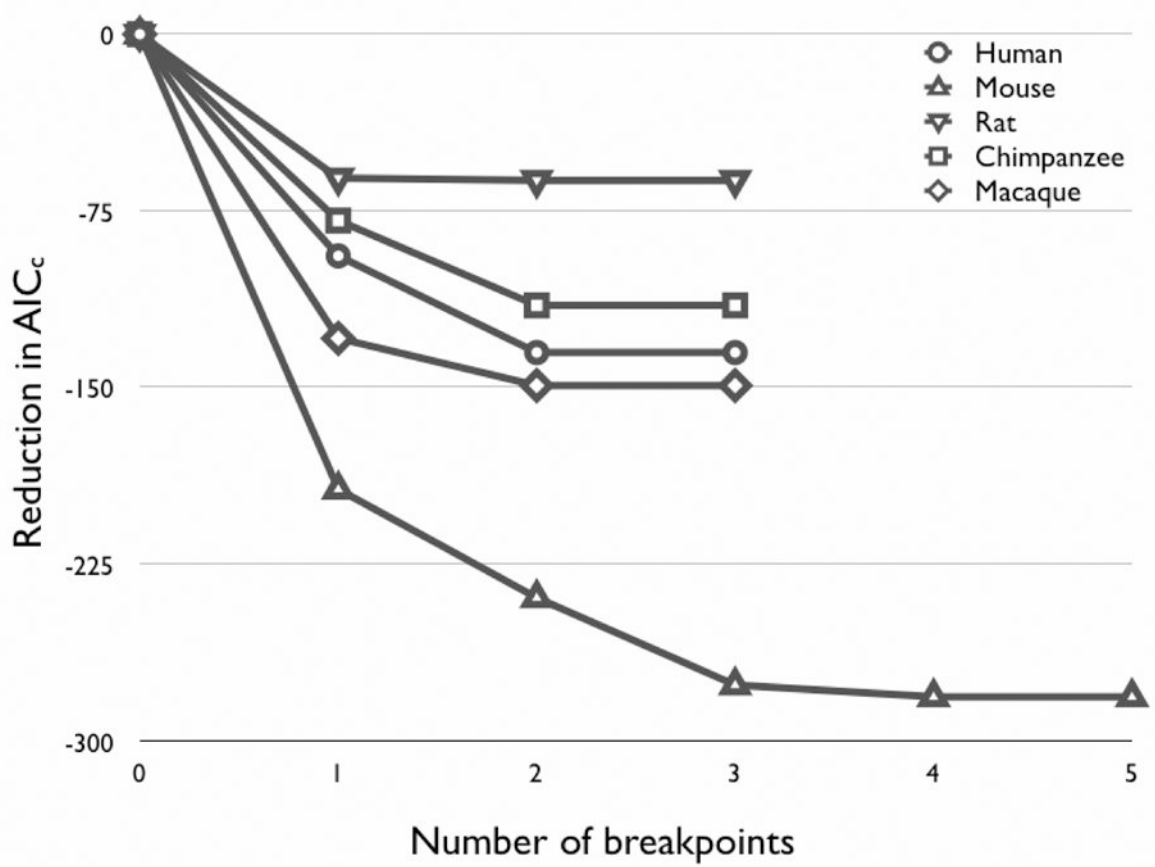


Figure 2. GARD detection of breakpoints indicative of partial gene conversion in IFNA gene data sets. Small sample Akaike Information Criterion (AIC_c) scores were plotted until no improvement in score was seen with increasing number of breakpoints (i.e. 2 points with the same score AIC_c). Therefore the penultimate point in each plot represents the model

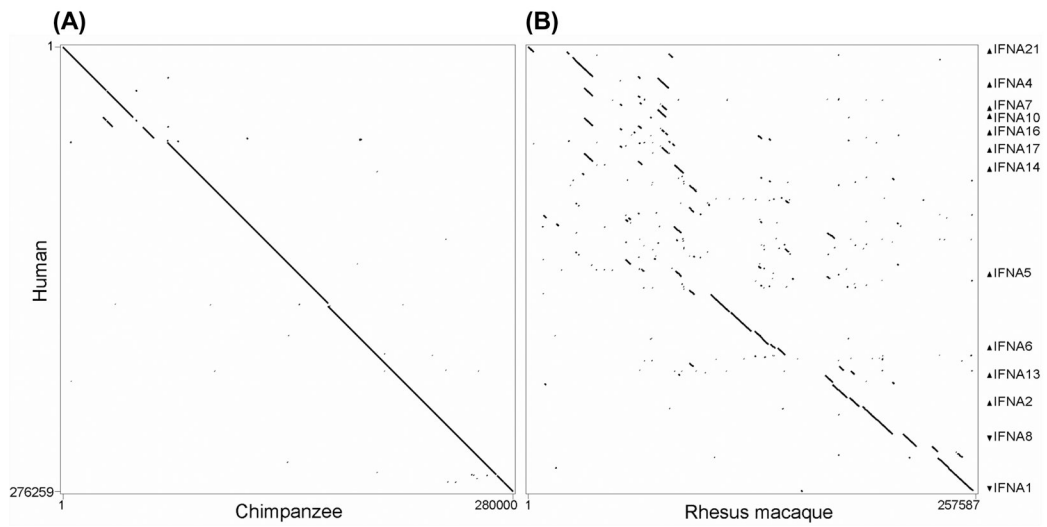


Figure 3. Alignment dot plots generated by Advanced PipMaker show that the IFNA gene family locus is more syntenous between (A) human and chimpanzee than between (B) human and rhesus macaque.

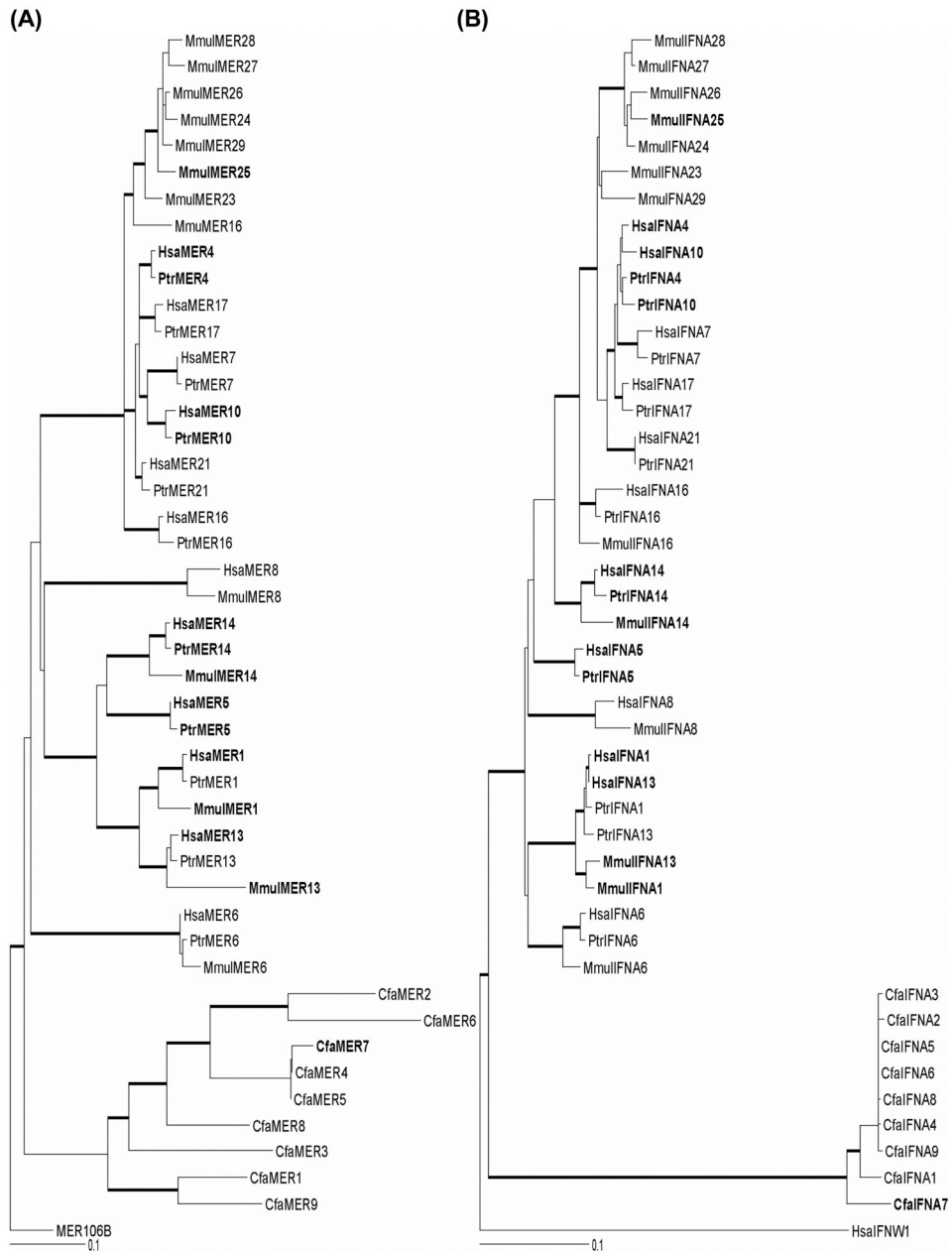


Figure 4. ML phylogenetic trees constructed using the HKY85 substitution model of (A) MER106B repeat elements associated with (B) IFNA gene sequences. Sequences in bold represent those whose topology is different between the two phylogenetic trees and thus indicative of whole gene conversion in the IFNA gene tree. Sequence data from RepBase was used as an outgroup or the MER106B tree and HsaIFNW was used to root the IFNA gene tree. All horizontal branches are drawn to scale and thicker branches in bold refer to those with bootstrap support >75%.

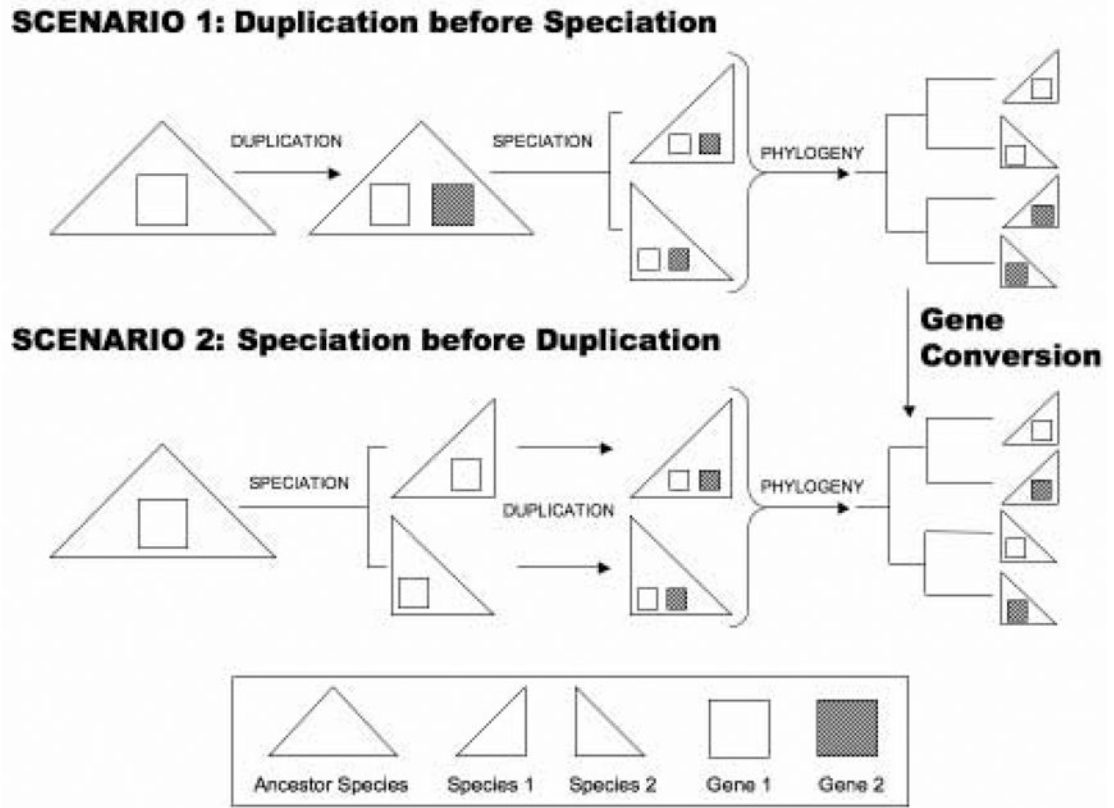


Figure 5. Evolutionary processes affecting IFNA gene families. Scenario 1 results in gene-specific clusters whereas scenario 2 produces species-specific clusters. Recent gene duplication following speciation and gene conversion are both capable of producing similar topologies.

Table 1
KH testing confirmed the significance of breakpoints estimated by GARD analysis¹.

| Data set | Number of breakpoints | AIC _c improvement | Breakpoint location | p-value | |
|----------------|-----------------------|------------------------------|---------------------|----------------------|----------------------|
| | | | | LHS vs. RHS <0.01 | RHS vs. LHS <0.01 |
| Human | 2 | 133.4 | 247 | " | " |
| Mouse | 4 | 279.0 | 504 | " | " |
| | | | 134 | " | " |
| | | | 226 | " | " |
| | | | 346 | " | " |
| Chimpanzee | 2 | 118.9 | 476 | " | " |
| | | | 267 | " | " |
| Rat | 2 | 63.9 | 354 | " | " |
| | | | 186 | " | " |
| Rhesus macaque | 2 | 147.6 | 322 | " | " |
| | | | 235 | " | " |
| Total | 1 | 777.3 | 431 | " | " |
| | | | 351 | " | " |

¹ For each data set, the number of breakpoints are reported for the model showing the greatest reduction in AIC_c as estimated by GARD. A one-tailed KH test was used in both directions to compare phylogenies constructed from the alignment segment to the left hand side (LHS) and right hand side (RHS) of each estimated breakpoint. In all cases the null hypothesis that there was no difference in topology was rejected.

Table 2

Partial gene conversion detection using GENECONV¹.

| Dataset | Gene 1 | Gene 2 | g-scale | SIMP | BC SIMP | BC KA P | Begin | End | L | Diff. | Total |
|----------------|------------|--------------|---------------|---------------|---------------|---------------|-------|-----|-----|-------|-------|
| Human | HsaIFNA4 | HsaIFNA14 | 0 | 0.1505 | 0.2808 | 0.6285 | 364 | 430 | 67 | 0 | 50 |
| | | | 1 | 1.0000 | 0.0078 | 0.5137 | 347 | 567 | 221 | 6 | 50 |
| | HsaIFNA7 | HsaIFNA10 | 2 | 0.1831 | 0.3276 | 0.9055 | 364 | 430 | 67 | 0 | 50 |
| | | | 0 | 0.0461 | 0.0234 | 0.2279 | 221 | 415 | 195 | 0 | 23 |
| | HsaIFNA7 | HsaIFNA14 | 1 | 1.0000 | 0.1092 | 1.0000 | 221 | 464 | 244 | 1 | 23 |
| | | | 2 | 0.0852 | 0.0312 | 0.4888 | 221 | 415 | 195 | 0 | 23 |
| | HsaIFNA10 | HsaIFNA14 | 0 | 0.2919 | 0.7176 | 1.0000 | 364 | 415 | 52 | 0 | 56 |
| | | | 1 | 1.0000 | 0.0078 | 0.1721 | 347 | 567 | 221 | 7 | 56 |
| | | | 2 | 0.2526 | 0.6162 | 1.0000 | 364 | 543 | 180 | 4 | 56 |
| | | | 0 | 0.0959 | 0.1872 | 0.4320 | 364 | 430 | 67 | 0 | 52 |
| HsaIFNA10 | HsaIFNA14 | 1 | 1.0000 | 0.0078 | 0.2406 | 347 | 567 | 221 | 6 | 52 | |
| | | 2 | 0.1388 | 0.2418 | 0.7315 | 364 | 430 | 7 | 0 | 52 | |
| Chimpanzee | PtrIFNA14 | PtrIFNA16 | 0 | 0.9998 | 1.0000 | 1.0000 | 364 | 416 | 53 | 0 | 41 |
| | | | 1 | 0.0345 | 0.0066 | 0.5120 | 220 | 567 | 348 | 10 | 41 |
| | PtrIFNA14 | PtrIFNA17 | 2 | 0.7767 | 1.0000 | 1.0000 | 442 | 567 | 126 | 1 | 41 |
| | | | 0 | 0.9249 | 1.0000 | 1.0000 | 364 | 416 | 53 | 0 | 50 |
| | PtrIFNA16 | PtrIFNA21 | 1 | 0.0404 | 0.0594 | 0.5559 | 364 | 567 | 204 | 5 | 50 |
| | | | 2 | 0.7607 | 1.0000 | 1.0000 | 466 | 567 | 102 | 1 | 50 |
| | PtrIFNA16 | PtrIFNA21 | 0 | 0.8860 | 1.0000 | 1.0000 | 134 | 237 | 104 | 0 | 30 |
| | | | 1 | 1.0000 | 0.0066 | 1.0000 | 1 | 294 | 294 | 4 | 30 |
| | PtrIFNA16 | PtrIFNA21 | 2 | 0.1826 | 0.4158 | 1.0000 | 1 | 294 | 263 | 4 | 30 |
| | | | 0 | 0.3663 | 0.4158 | 1.0000 | 32 | 294 | 263 | 3 | 30 |
| Mouse | MmusIFNA2 | MmusIFNA7/10 | 0 | 0.3098 | 0.5642 | 1.0000 | 311 | 363 | 53 | 0 | 66 |
| | | | 1 | 0.0295 | 0.0273 | 0.6639 | 311 | 538 | 228 | 7 | 66 |
| | MmusIFNA6T | MmusIFNA11 | 2 | 0.2241 | 0.4004 | 0.9791 | 416 | 416 | 106 | 2 | 66 |
| | | | 0 | 0.0124 | 0.0091 | 0.0624 | 462 | 569 | 108 | 0 | 63 |
| | MmusIFNA9 | MmusIFNA14 | 1 | 0.0835 | 0.0819 | 1.0000 | 417 | 569 | 153 | 2 | 63 |
| | | | 2 | 0.0178 | 0.0182 | 0.1208 | 462 | 569 | 108 | 0 | 63 |
| | MmusIFNA9 | MmusIFNA14 | 0 | 0.0296 | 0.0546 | 0.1237 | 221 | 303 | 83 | 0 | 50 |
| | | | 1 | 0.1613 | 0.0910 | 1.0000 | 172 | 323 | 152 | 4 | 50 |
| | MmusIFNA9 | MmusIFNA14 | 2 | 0.0390 | 0.0637 | 0.2243 | 221 | 303 | 83 | 0 | 50 |
| | | | 0 | 0.0270 | 0.0765 | 0.1785 | 455 | 542 | 88 | 0 | 79 |
| Rat | RnoIFNA3 | RnoIFNA9 | 1 | 1.0000 | 0.1989 | 0.9939 | 435 | 542 | 108 | 1 | 79 |
| | | | 2 | 0.0344 | 0.1377 | 0.2870 | 455 | 542 | 88 | 0 | 79 |
| | RnoIFNA10 | RnoIFNA13 | 0 | 0.0165 | 0.0306 | 0.1251 | 1 | 120 | 120 | 0 | 52 |
| | | | 1 | 1.0000 | 0.8568 | 1.0000 | 1 | 120 | 120 | 0 | 52 |
| | RnoIFNA10 | RnoIFNA14 | 2 | 0.0337 | 0.0459 | 0.2764 | 1 | 120 | 120 | 0 | 52 |
| | | | 0 | 0.0165 | 0.0306 | 0.1251 | 1 | 120 | 120 | 0 | 52 |
| RnoIFNA10 | RnoIFNA14 | 1 | 1.0000 | 0.9792 | 1.0000 | 1 | 120 | 120 | 0 | 52 | |
| | | 2 | 0.0337 | 0.0459 | 0.2764 | 1 | 120 | 120 | 0 | 52 | |
| Rhesus macaque | MmulIFNA1 | MmulIFNA8 | 0 | 0.0187 | 0.0637 | 0.1173 | 422 | 535 | 114 | 0 | 75 |
| | | | 1 | 0.0614 | 0.2184 | 1.0000 | 422 | 535 | 114 | 0 | 75 |
| | MmulIFNA14 | MmulIFNA16 | 2 | 0.0201 | 0.0728 | 0.1574 | 422 | 535 | 114 | 0 | 75 |
| | | | 0 | 0.0000 | 0.0091 | 0.0000 | 236 | 463 | 228 | 0 | 42 |
| | MmulIFNA14 | MmulIFNA23 | 1 | 0.0000 | 0.0091 | 0.0065 | 221 | 567 | 347 | 3 | 42 |
| | | | 2 | 0.0000 | 0.0091 | 0.0000 | 236 | 567 | 332 | 2 | 42 |
| | MmulIFNA14 | MmulIFNA23 | 0 | 0.6735 | 1.0000 | 1.0000 | 484 | 559 | 76 | 0 | 58 |
| | | | 1 | 0.0153 | 0.0091 | 0.3819 | 356 | 559 | 204 | 6 | 58 |
| | MmulIFNA14 | MmulIFNA27 | 2 | 0.0777 | 0.1547 | 0.4810 | 400 | 559 | 160 | 2 | 58 |
| | | | 0 | 0.0026 | 0.0091 | 0.0239 | 432 | 567 | 136 | 0 | 58 |

| Dataset | Gene 1 | Gene 2 | g-scale | SIMP | BC SIMP | BC KA P | Begin | End | L | Diff. | Total |
|--------------------|------------|------------|---------|---------------|---------------|---------------|-------|-----|-----|-------|-------|
| | | | 1 | 0.0027 | 0.0091 | 0.1173 | 401 | 567 | 167 | 1 | 58 |
| | | | 2 | 0.0015 | 0.0091 | 0.0175 | 401 | 567 | 167 | 1 | 58 |
| | MmulIFNA23 | MmulIFNA25 | 0 | 0.9610 | 1.0000 | 1.0000 | 219 | 327 | 109 | 0 | 28 |
| | | | 1 | 0.4456 | 0.0273 | 1.0000 | 1 | 327 | 327 | 5 | 28 |
| | | | 2 | 0.2548 | 1.0000 | 1.0000 | 87 | 327 | 241 | 2 | 28 |
| | MmulIFNA27 | MmulIFNA29 | 0 | 0.7536 | 1.0000 | 1.0000 | 110 | 215 | 106 | 0 | 32 |
| | | | 1 | 0.2481 | 0.0091 | 1.0000 | 1 | 305 | 305 | 4 | 32 |
| | | | 2 | 0.0652 | 0.0546 | 0.3963 | 26 | 305 | 280 | 3 | 32 |
| | MmulIFNA24 | MmulIFNA29 | 0 | 0.7024 | 1.0000 | 1.0000 | 217 | 327 | 111 | 0 | 32 |
| | | | 1 | 0.3206 | 0.0091 | 1.0000 | 85 | 327 | 243 | 2 | 32 |
| | | | 2 | 0.0200 | 0.0091 | 0.1539 | 110 | 327 | 218 | 1 | 32 |
| | MmulIFNA29 | MmulIFNA28 | 0 | 0.0241 | 0.0182 | 0.1450 | 82 | 215 | 134 | 0 | 36 |
| | | | 1 | 0.0297 | 0.0091 | 0.5867 | 82 | 305 | 224 | 1 | 36 |
| | | | 2 | 0.0019 | 0.0091 | 0.0234 | 82 | 305 | 224 | 1 | 36 |
| Total ² | MhiIFNA1a | MhiIFNA7 | 0 | 0.0360 | 1.0000 | 0.8538 | 1 | 99 | 99 | 0 | 61 |
| | | | 1 | 1.0000 | 1.0000 | 1.0000 | 146 | 321 | 176 | 3 | 55 |
| | | | 2 | 0.1131 | 1.0000 | 1.0000 | 1 | 99 | 99 | 0 | 61 |
| | MmusIFNA2 | RnoIFNA3 | 0 | 0.0090 | 1.0000 | 0.2458 | 494 | 580 | 87 | 0 | 84 |
| | | | 1 | 1.0000 | 1.0000 | 1.0000 | 481 | 580 | 100 | 1 | 84 |
| | | | 2 | 0.0220 | 1.0000 | 0.7786 | 494 | 580 | 87 | 0 | 84 |

¹ *p*-values in bold are <0.05. *p*-values were derived from three types of statistical test: *SIMP*, global simulated *p*-values based on 10,000 permutations; *BC SIMP*, Bonferroni-corrected pairwise simulated *p*-values based on 10,000 permutations; *BC KA P*, Bonferroni-corrected global Karlin-Altschul *p*-values. Strong evidence of gene conversion was taken when the fragment identified was supported by a *p*-value <0.05 from at least two different types of statistical test. Shaded results refer to those where significance could only be derived from a single statistical test. *g-scale*, mismatch penalty where 0 forces no mismatches and larger integers allow mismatches; *Begin* and *End*, first and last nucleotide of potential converted region, respectively; *L*, length of the converted region; *Diff.*, number of nucleotide mismatches in the potential converted region; *Total diff.*, total number of nucleotide mismatches between the two sequences.

² Only those partial gene conversion events that were not identified in intra-species analysis are presented for the total data set.