Genomic Organization of Mouse and Human 65 kDa FK506-Binding Protein Genes and Evolution of the FKBP Multigene Family

ARTICLE *in* GENOMICS · JULY 2002

Impact Factor: 2.28 · DOI: 10.1006/geno.2002.6777 · Source: PubMed

4 AUTHORS, INCLUDING:

Jimin Gao
Wenzhou Medical College
122 PUBLICATIONS **1,213** CITATIONS

SEE PROFILE

doi:10.1006/geno.2002.6777, available online at http://www.idealibrary.com on IDEAL **Article**

Genomic Organization of Mouse and Human 65 kDa FK506-Binding Protein Genes and Evolution of the FKBP Multigene Family

Charles E. Patterson,[1] Jimin Gao,[1] Alejandro P. Rooney,[2] and Elaine C. Davis[1,*]

[1]*Department of Cell Biology, University of Texas Southwestern Medical Center at Dallas, Texas 75390, USA*
[2]*Department of Biological Sciences, Mississippi State University, Starkville, Mississippi 39762, USA*

**To whom correspondence and reprint requests should be addressed. Fax: (214) 648-8694. E-mail: elaine.davis@utsouthwestern.edu.*

FK506-binding proteins (FKBPs) are peptidyl–prolyl *cis/trans* isomerases (PPIases) that bind the immunosuppressive drug FK506. Of the many eukaryotic FKBPs that have been identified, FKBP65 is an endoplasmic reticulum-localized protein that associates with tropoelastin in the secretory pathway. Unlike any other FKBP characterized so far, FKBP65 is developmentally regulated and may be intimately involved in organogenesis. Here, we report the isolation, sequencing, and genomic organization of the mouse FKBP65 gene (*Fkbp10*) and provide a comparison with the human ortholog. Mouse *Fkbp10* contains 10 exons and 9 introns encompassing 8.5 kb. The exon–intron organization of *Fkbp10* displays a pattern of repetition that reflects the coding sequence of the four PPIase, or FK506-binding, domains present in the mature protein. The exon organization of the PPIase domains differs from that of the other FKBP family members. The evolution of the FKBP65 gene and other members of the FKBP multigene family were therefore investigated from a taxonomically diverse array of prokaryotic and eukaryotic taxa. These analyses suggest that the FKBP multigene family emerged early in the evolutionary history of eukaryotes, and during that time some members, including the FKBP65 gene, have experienced gene elongation by means of PPIase domain duplication.

Key Words: FKBP, FKBP65, PPIase, prolyl isomerase, tropoelastin, immunophilin, chaperone, endoplasmic reticulum, evolution, domain shuffling

INTRODUCTION

Immunophilins were initially discovered as intracellular targets for the immunosuppressive drugs cyclosporin A (CsA) and FK506, giving rise to the cyclophilins (CyPs) and FK506-binding proteins (FKBPs), respectively. Although CyPs and FKBPs are structurally unrelated, all immunophilins possess the potential for peptidyl–prolyl *cis/trans* isomerization (PPIase) activity through their ligand-binding domains. The ability of the binding domain to catalyze *cis/trans* isomerization of a peptidyl–prolyl bond is separate and distinct from immunosuppressive activity, and binding of the respective immunosuppressive ligands can inhibit the PPIase activity of the active site. FKBPs are encoded by a multigene family. Some family members consist of a single PPIase (or FK506-binding) domain, but other members have multiple PPIase domains [reviewed in 1]. FKBP12 is considered to be the major FKBP responsible for mediating the immunosuppressive effects of FK506 by controlling T-cell and interleukin-2 signaling [2]. FKBP12 and its ortholog FKBP12.6 have also been implicated as key components of the skeletal and cardiac ryanodine receptor, respectively [reviewed in 3]. Given the diversity of the FKBP family members (with single to tetrameric FK506-binding domains, various additional functional domains, variations in amino acid composition, and wide subcellular localization), it is likely that many of these proteins possess alternative functions apart from immunosuppression. Indeed, many immunophilins have been implicated in events such as protein folding, assembly, and trafficking; coregulation of molecular complexes; transcriptional and translational regulation; and cell–cell interactions [4].

0888-7543/02 $35.00

doi:10.1006/geno.2002.6777, available online at http://www.idealibrary.com on IDEAL

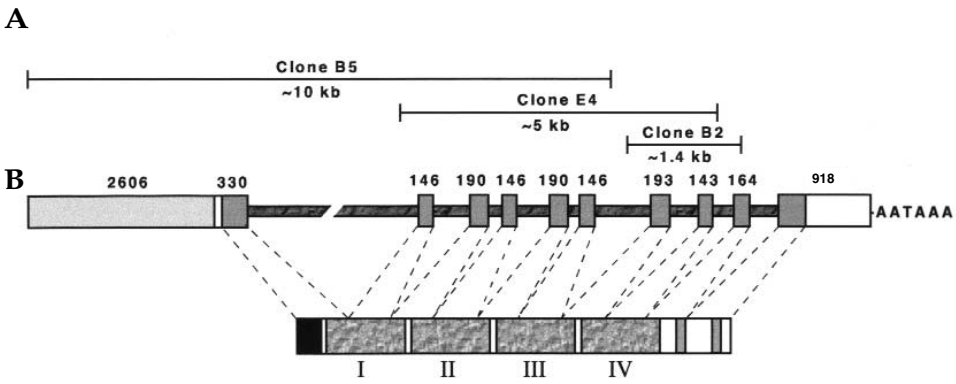

FIG. 1. Genomic organization of mouse *Fkbp10*. (A) Schematic exon–intron organization of *Fkbp10* as revealed by direct sequencing and PCR analysis of the three overlapping genomic clones (clones B5, E4, and B2) and the 3′ region. The promoter region, exons, and introns are represented by a light gray box, gray boxes, and filled bars, respectively. The numbers on top indicate exon sizes. (B) The repetitively alternating exon sizes correspond to the four FK506-binding domains of the mature protein and are shown schematically. The signal sequence, four FK506-binding domains, and putative EF-hand calcium-binding domains are represented by black, gray textured, and light gray boxes, respectively.

Proline isomerization has been identified as a relatively slow step in protein folding [5]. It is not surprising, therefore, that several FKBPs have been identified in the endoplasmic reticulum (ER), where they have been hypothesized to function as molecular chaperones and/or foldases [reviewed in 6]. For the most part, specific ligands for the ER-localized FKBPs are unknown. One exception is FKBP65, which has been shown to associate with the extracellular matrix protein tropoelastin in the ER [7]. Because tropoelastin is very proline-rich, FKBP65 may assist in the folding of peptidyl–proline bonds present in the protein before its transport from the ER to the cell surface. FKBP65 shows a developmentally regulated expression pattern that is similar to tropoelastin. The strong expression of FKBP65 in tissues that do not contain elastin, however, suggests that it may have other ligands [8].

The consensus sequence for the FK506-binding domain consists of approximately 108 amino acids, which share high homology with the 108-amino-acid archetypal FKBP12 [9]. The three-dimensional structure of FKBP12 has been resolved by X-ray crystallography [10] and NMR [11] and consists of five antiparallel β-strands wrapping around a single α-helix. Although this motif structure is found in all FKBP family members, many family members share low homology to FKBP12 and it is clear that significant evolutionary divergence must have occurred. Additionally, many FKBP family members are often mosaic, containing unrelated domains such as carboxy-terminal tetratricopeptide repeat (TPR) motifs, EF-hand calcium binding domains, and, in some cases, calmodulin-binding domains and leucine zipper repeats [1,12]. Most FKBPs contain only one FK506-binding domain, but others may contain as many as four consecutive domains, as seen for FKBP65.

The four consecutive PPIase domains of FKBP65 make it an interesting candidate for investigating the evolutionary and phylogenetic relationship of the FKBP multigene family. Apart from FKBP65, only its paralog, FKBP60, has four PPIase domains. Furthermore, FKBP65 seems to be restricted to vertebrates (unpublished data), whereas many smaller FKBP family members have a wider phylogenetic distribution [1,13]. It is well established that although intron size is subject to high variation during evolution, exon–intron boundaries between related genes and their functional domains are reasonably conserved [14] and therefore may provide insight into the relationship between genes encoding FK506-binding domains.

To investigate the evolutionary association of FKBP65 with the other members of this diverse family of proteins, we determined its genomic organization and compared it with that of other FK506-binding domains. In addition, we investigated the evolutionary relationships among 100 FK506-binding domains from a taxonomically diverse array of prokaryotes and eukaryotes. Our results suggest that FKBP gene family members have evolved by an interesting system of gene duplication and gene elongation over the course of eukaryotic evolutionary history.

RESULTS

Nomenclature of the FKBP Multigene Family.

The nomenclature used to distinguish different FKBP protein family members has been a source of confusion. In addition, there are currently no strict guidelines for assigning gene nomenclature to the genes encoding these family members. The genes discussed below have been approved by the HUGO Gene Nomenclature Committee and consist of the following proteins and their corresponding genes: FKBP12 (*FKBP1A*), FKBP12.6 (*FKBP1B*), FKBP13 (*FKBP2*), FKBP9 (*FKBP11*), FKBP22 (*FKBP14*), FKBP23 (*FKBP7*), FKBP25 (*FKBP3*), FKBP36 (*FKBP6*), FKBP37 (*AIP*), FKBP38 (*FKBP8*), FKBP51 (*FKBP5*), FKBP52 (*FKBP4*), FKBP60 (*FKBP9*), and FKBP65 (*FKBP10*).

Genomic Organization of the Mouse FKBP65 Gene

To determine the structural organization of the mouse FKBP65 gene (*Fkbp10*), three overlapping subclones were obtained through *Eco*RI and *Bam*HI digestions of mouse genomic DNA and used for sequencing the exon–intron boundaries. The boundary positions were deduced from sequence comparisons between the genomic DNA and that previously published for the FKBP65 cDNA [15]. We found

TABLE 1: Exon-intron organization of the mouse and human FKBP10 genes

Mouse *Fkbp10*		Sequences at exon-intron junctions		
Exon	bp	5′ donor sequences 3′ acceptor sequences	Intron	bp
1[a]	242	5′ UTR... ^{2639}*ATG*TTCCTTG...TTGACTCCAG^{2880}**gt**aagagctt	1	>3900
2[b]	146	cctcctcc**ag** ^{1}CTATGACCGT...ATCGGTGTGG^{146} **gt**aaggaggc	2	345
3	190	ttccccac**ag** ^{492}CGGGCCTCAT...TTGACAACAG^{681} **gt**agaagcta	3	113
4	146	ttctctgc**ag** ^{795}CTACAGTAGG...AAAGGCTATG^{940} **gt**aagggaat	4	304
5	190	cttccccc**ag**^{1245}GGACTGTGAT...TTGATTCCAG^{1434}**gt**caggggg	5	110
6	146	tgtattac**ag**^{1545}CTACTCCCGA...AATGGGACAG^{1690}**gt**agggtttg	6	535
7	193	ctggcctc**ag**^{2226}GAGACAAGAT...TCTTCTCGTC^{2418}**gt**gggttctt	7	266
8	143	ctccccac**ag**^{2685}CCACGACTAT...GAGAATGGAG^{2827}**gt**gagagaca	8	207
9	164	atctctcc**ag**^{3035}CCCGGGGTGT...CCCAGAAGAG^{3198}**gt**gggtcagg	9	288
10	918	cctcccccc**ag**^{3487}TTCTCTTCCT...To the polyA signal		

Human *FKBP10*		Sequences at exon-intron junctions		
Exon	bp	5′ donor sequences 3′ acceptor sequences	Intron	bp
1[c]	245	5′ UTR... ^{87}*ATG*TTCCCCG...TTGATTCAAG^{331} **gt**aaccccgg	1	>3600
2	146	cccccccc**ag** ^{332}CTATGATCGC...ATCGGCCTGG^{477} **gt**gagaaggg	2	887
3	190	gctctcac**ag** ^{478}CGGGGCTCAT...TCGACACCAG^{667} **gt**gaggggct	3	103
4	146	ttttctgc**ag** ^{668}CTACAGTAAG...AAAGGCTATG^{813} **gt**gagggtgg	4	685
5	190	ccttcccc**ag** ^{814}GGACAGTGAT...TCGATTCCAG^{1003}**gt**caggaggg	5	130
6	146	tgtattgc**ag**^{1004}CTACTCCCGC...AATGGAACTG^{1149}**gt**aggggcgt	6	593
7	193	ctggcctc**ag**^{1150}GAGACAAGAT...TGTTCACCTC^{1342}**gt**gggtccgg	7	485
8	143	tctccccc**ag**^{1343}GCATGACTAC...GAGAGTGGAG^{1485}**gt**gaggggct	8	562
9	164	ctccctcc**ag**^{1486}CCCGGGGAGT...TCCGGAGGAG^{1649}**gt**gggtgaag	9	405
10	960	cctgcccc**ag**^{1650}TTCTCCACCT...To the polyA signal		

Exon and flanking intron sequences are shown in uppercase and lowercase, respectively. The ATG start sites are italicized and 5′ donor and 3′ acceptor sites are indicated in bold.

[a]Nucleotide positions for the 5′ UTR and exon 1 of mouse *Fkbp10* (superscript numbers) are according to the submitted genomic sequence (GenBank acc. no. AF456412).

[b]Nucleotide positions for exons 2-10 of mouse *Fkbp10* (superscript numbers) are according to the submitted genomic sequence (GenBank acc. no. AF456413).

[c]Nucleotide positions for human *FKBP10* (superscript numbers) are according to the submitted cDNA (GenBank acc. no. AB045981).

Fkbp10 to consist of 10 exons and 9 introns spanning approximately 8.5 kb of genomic DNA (Fig. 1A). With the exception of introns 1 and 9, the actual size of each intron was determined by sequencing the genomic clones. Overall, the intron sizes were small, ranging from 110 bp for intron 5 to 535 bp for intron 6 (Table 1). The size of intron 1 was determined to be approximately 3.9 kb by subtraction of the compiled sequences from clone B5 and PCR amplification of genomic DNA. All donor and acceptor sites at the exon–intron boundaries contained the expected, invariant GT-AG intronic dinucleotides [16].

The exons displayed a striking pattern of repetition (Fig. 1A), alternating between 146 -bp and 190 bp starting with exon 2 and continuing through exon 6. The pattern is altered only slightly (by 3 nt) for exons 7 and 8, perhaps as a result of intron sliding. Upon further examination, the repetitive pattern could be attributed to the four FK506-binding domains that are present in the mature protein (Fig. 1B). The first three domains each consist of a 146-bp exon flanked at the 5′ end by approximately 122 bp of the preceding exon and at the 3′ end by 59 bp of the following exon. The pattern for the fourth FK506-binding domain differs only by the 3-bp shift between exons 7 and 8. Exon 9, which consists of 164 bp, completes the last 59 bp of the fourth domain. Following the final FK506-binding domain, two putative EF-hand calcium-binding motifs are located in the C terminus of the protein, one in each of exons 9 and 10.

While sequencing *Fkbp10*, several nucleotide substitutions were discovered upon comparison with the published cDNA [15]. Within the coding region, these substitutions included a dinucleotide CG→GC substitution of nucleotide 132 and nucleotide 133, resulting in a Val to Leu substitution of amino

doi:10.1006/geno.2002.6777, available online at http://www.idealibrary.com on IDEAL

FIG. 2. Nucleotide sequence of the proximal promoter region of *Fkbp10*. The DNA sequence is numbered from the ATG start codon (bold), which continues to the end of exon 1. A TATA-like box is indicated in bold and underlined and the GT dinucleotide repeats are italicized. DNA consensus binding sites for possible transcription factors were identified using MatInspector 5.1. The first letter of the abbreviation of the transcription factor is directly above the 5′ most base of the potential core binding site. Multiple transcription factors capable of binding to the same core sequence are separated by a slash (/).

acid 14; and a G→C substitution of nucleotide 659, resulting in a Gly to Ala substitution of amino acid 189. Two additional substitutions within the 3′ UTR included a G→C substitution at nucleotide 2148 and a dinucleotide TG→GT substitution at nucleotide 2567 and nucleotide 2568.

Characterization of the Putative Promoter of Mouse *Fkbp10*

A total of 2.6 kb of 5′ flanking region was sequenced in order to characterize the putative promoter region of mouse *Fkbp10*. Using the computer program MatInspector 5.1, we analyzed 2.3 kb of this sequence and found numerous consensus transcription factor binding sites typical of eukaryotic genes (Fig. 2). Although attempts to determine the transcription initiation site for *Fkbp10* were unsuccessful, a TATA-like motif sequence, CATAAAA, was found at position –176 to –170 bp from the ATG translational start codon (labeled +1 in Fig. 2). Furthermore, a binding site for upstream stimulating factor-2, a basic/helix loop helix/leucine zipper domain-containing protein known to activate or enhance expression of some genes [17], was found just downstream of the TATA-like motif. Other potential transcription factors flanking the TATA-like motif include an SP1 site (–108 to –105), two VDR/RXR sites (–93 to –90 and –63 to –60), and a single SMAD3 site (–11 to –8). This region of *Fkbp10* also contains consensus sequences for many MZF1 sites (GGGG), possibly due to the numerous stretches of high GC content (61% overall). Whether the presence of these putative transcription factor binding sites within the 5′ flanking region of the *FKBP65* has any role in the regulation of FKBP65 expression remains unknown.

Genomic Organization of Human *FKBP10*

The organization of human *FKBP10* was determined using genomic DNA sequences obtained from the human Celera database. Introns were identified where human genomic *FKBP10* DNA sequence diverged from that of the submitted cDNA (GenBank acc. no. BAB20974). The structural organization of *FKBP10* was found to be nearly identical in mouse and human. Excluding introns, mouse *Fkbp10* was 84% similar to human *FKBP10* at the nucleotide level and 89% similar to human FKBP65 at the amino acid level. More specifically, human FKBP65 was found to differ from mouse FKBP65 by only a single amino acid addition at Val24 in exon 1 and a slight expansion of the 3′ untranslated region of exon 10 (Table 1). The single amino acid addition within human FKBP65 is positioned within the cleavable amino-terminal signal sequence, outside of the first FK506-binding domain, and therefore would not change the size of the protein with respect to the mature mouse FKBP65.

Genomic Organization of the FK506-Binding Domains of the Human FKBP Family

The genomic organization of a few FKBPs containing single PPIase domains has been previously determined. These include *FKBP6* (which is one of many genes located in the 1-Mb region commonly deleted in Williams syndrome [18]), *FKBP1A*, and *FKBP2* [19]. To determine if other FKBP family members shared a similar genomic structure to that of *FKBP10*, the organization of the FK506-binding domains of known human FKBP genes was determined using sequences obtained from the Celera database or through available genomic sequences present in GenBank. Resulting alignments were analyzed for intronic sequence where genomic DNA sequence diverged from that of the cDNA. In all cases, introns conformed to the GT-AG rule [16]. The overall genomic organization of *FKBP10* was found to be very similar to its

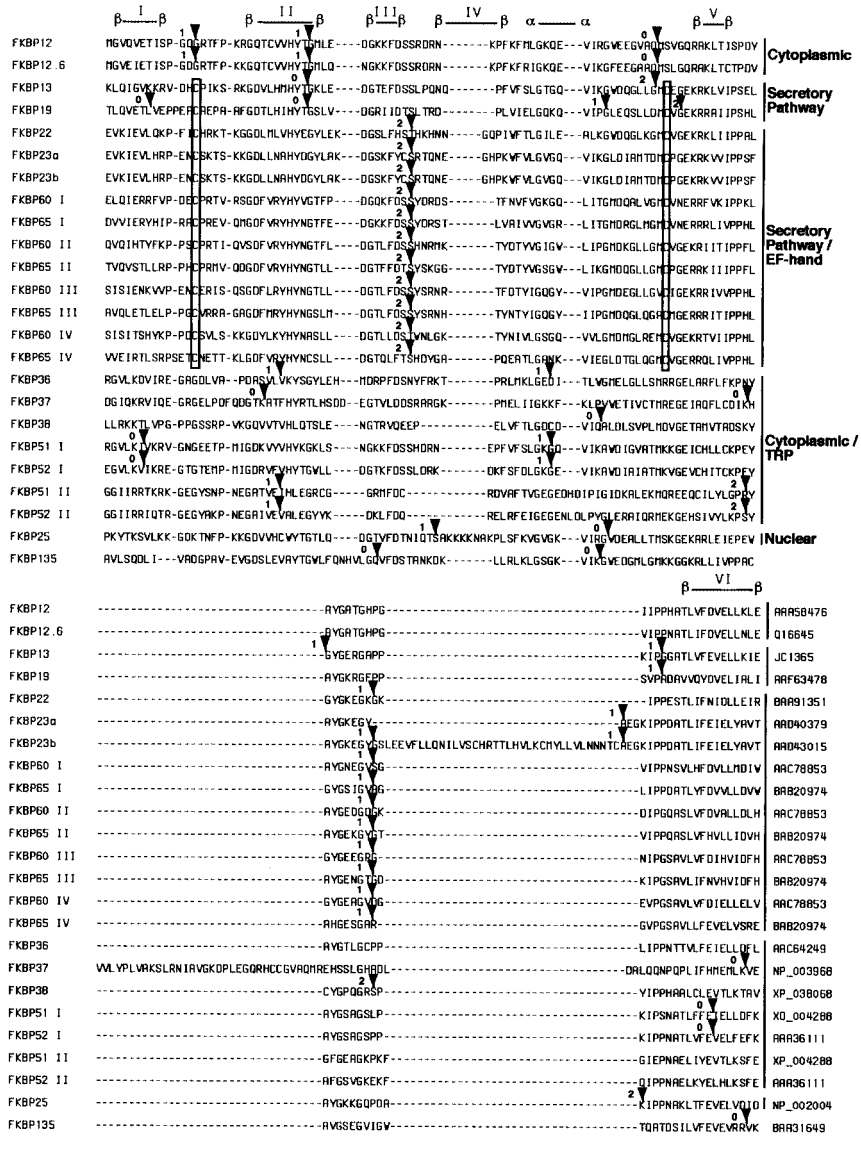

FIG. 3. Genomic structure of the FK506-binding domains of the human FKBPs. Domains defined by the FK506-binding domain consensus sequence [9] were aligned and grouped by cellular localization. Arrows indicate intron positions and adjacent numbers indicate the phase of the intron. Sequences that relate to secondary structures (β-strand; α-helix) are shown above the amino acid sequences. Conserved cysteine residues present in the ER-localized FKBPs are boxed. Cellular localization of each FKBP, the presence of additional domains, and the GenBank accession numbers are denoted on the right side. FKBP23b is a recently identified isoform of FKBP23a. The aligned FKBP family members are encoded by the following genes: FKBP12 (*FKBP1A*), FKBP12.6 (*FKBP1B*), FKBP13 (*FKBP2*), FKBP19 (*FKBP11*), FKBP22 (*FKBP14*), FKBP23 (*FKBP7*), FKBP25 (*FKBP3*), FKBP36 (*FKBP6*), FKBP37 (*AIP*), FKBP38 (*FKBP8*), FKBP51 (*FKBP5*), FKBP52 (*FKBP4*), FKBP60 (*FKBP9*), and FKBP65 (*FKBP10*).

paralog *FKBP9*, both of which contain two C-terminal EF-hand calcium-binding domains and are localized to the ER. With the exception of exons 1, 7, and 10, which are slightly smaller at 221 bp (from the start methionine), 187 bp, and 474 bp, respectively, the genomic organization of *FKBP9* was identical to that of *FKBP10* (data not shown). This is also reflected in the close evolutionary relationship seen between these two genes as shown in the phylogenetic tree (Fig. 4) and supported statistically by high bootstrap values (greater than 90%).

As expected, the genomic organization within each of the four FK506-binding domains of *FKBP10* and *FKBP9* was strongly conserved (Fig. 3). The organization also closely resembles that of two other ER-localized FKBP genes known to contain C-terminal EF-hand domains, *FKBP14* and *FKBP7*. This suggests that each of the ER-localized FKBPs is closely related. This is also reflected in the close evolutionary relationship between these genes seen in the phylogenetic tree (Figs. 4 and 5), although low statistical support was obtained for this grouping. Human *FKBP8* also appears to share a similar single intron position at the C-terminal portion of the FK506-binding domain, despite the presence of TPR domains in place of EF-hand domains and low overall amino acid similarity to the other FKBPs. However, it does not appear to be closely related to *FKBP14* and *FKBP7* on the basis of where it clusters in the tree (Fig. 4).

Human *FKBP5* and its paralog *FKBP4* each contain two FK506-binding domains and C-terminal TPR domains, and are highly similar in terms of their genomic organization. Unlike the four very similar FK506-binding domains of *FKBP10* and *FKBP9*, the exon–intron boundaries of the two successive domains of *FKBP5* and *FKBP4* are divergent (Fig. 3). This is further reflected in the phylogeny (Fig. 4), in which the two domains are more closely related to other FKBPs than to each other and are supported by high bootstrap values (98% or more). The individual domains of *FKBP5* or *FKBP4* do, however, demonstrate conservation when linked by the intron positions found within the single FK506-binding domain of *FKBP6*. Similarly *AIP*, which contains C-terminal TPR domains and a single FK506-binding domain interrupted by non-homologous sequence, shares some similarity with the second domain of *FKBP5* and *FKBP4* despite an additional intron located at the C terminus of the domain. The results of our phylogenetic analysis suggest that these genes are closely related, although

doi:10.1006/geno.2002.6777, available online at http://www.idealibrary.com on IDEAL

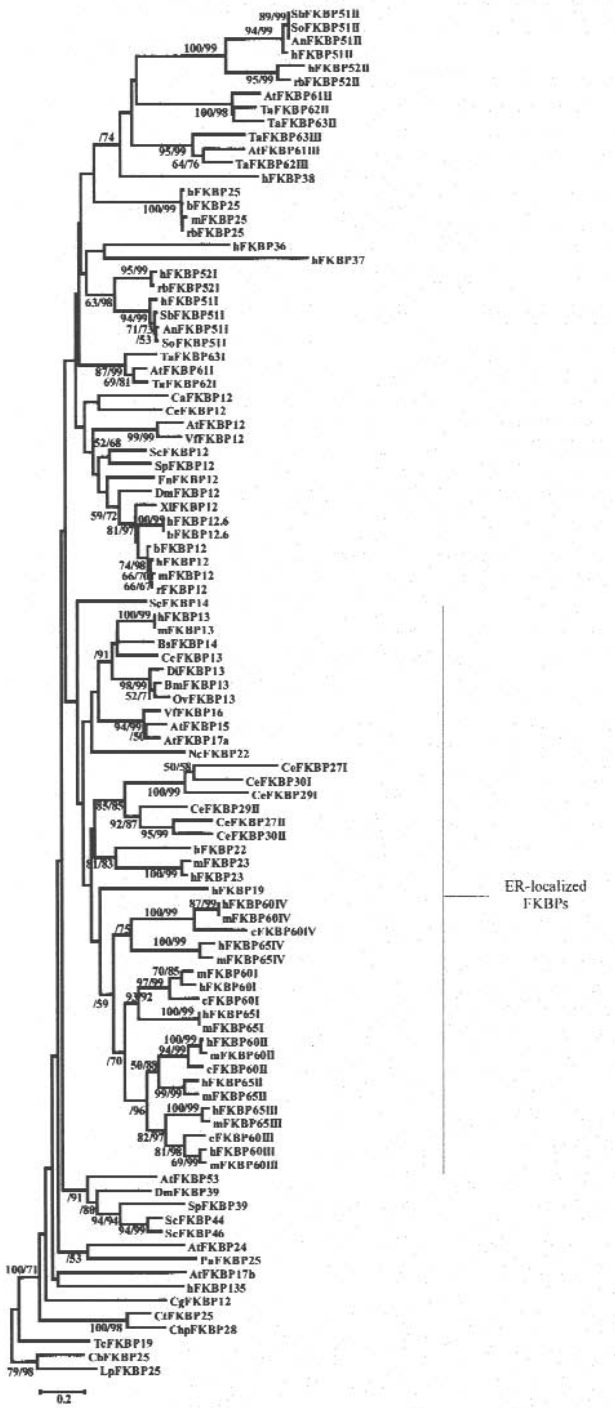

FIG. 4. Phylogenetic relationships among 100 FKBP domains. The domains of multi-domain proteins were analyzed separately. Branches are drawn in proportion to Poisson distances. Numbers along branches represented bootstrap values greater than 50%. The first number represents the value generated using the maximum likelihood method; the second number is the value generated using the neighbor-joining method. A blank on either side of the slanted line indicates a value lower than 50% for the respective method. The monophyletic grouping of FKBP domains containing unique cysteine residues specific to ER-localized FKBPs is shown. The tree is unrooted. Common names of some organisms are cited as follows: b, bovine (*Bos taurus*); c, chicken (*Gallus gallus*); h, human (Homo sapiens); m, mouse (*Mus musculus*); and rb, rabbit (*Oryctolagus caniculus*). Species names are abbreviated as follows: An, *Aotus nancymae*; At, *Arabidopsis thaliana*; Bm, *Brugia malayi*; Bs, *Botryllus schlosseri*; Ca, *Candida albicans*; Cb, *Coxiella burnetii*; Ce, *Caenorhabditis elegans*; Cg, *Corynebacterium glutamicum*; Chp, *Chlamydophila pneumoniae*; Ct, *Chlamydia trachomatis*; Di, *Dirofilaria immitus*; Dm, *Drosophila melanogaster*; Fn, *Filobasidiella neoformans*; Lp, *Legionella pneumophila*; Nc, *Neurospora crassa*; Ov, *Onchocerca volvulus*; Pa, *Pseudomona aureoginosa*; Sb, *Saimiri boliviensis*; Sc, *Saccharomyces cerevisiae*; So, *Saguinus oedipus*; Sp, *Schizosaccharomyces pombe*; Ta, *Triticum aestivum*; Tc, *Thermococcus* sp. KS-1; Vf, *Vicia faba*; Xl, *Xenopus laevis*.

these groupings are not statistically reliable as reflected by low bootstrap support (Fig. 4).

FKBP1A and *FKBP2* have previously been shown to share analogous exon–intron organization, with the exception of a few intron insertions or deletions [19]. The position of these introns coincides only with *FKBP1B* and *FKBP11* (Fig. 3). Similarly, the single domain of *FKBP3*, another TPR-domain containing FKBP, contains one intron that is positioned closely to an intron found within *FKBP8* and the gene encoding FKBP135. Other intron positions within *FKBP3*, however, fail to coincide with other FKBP family members. Human FKBP135, a high-molecular-weight, hypothetical FKBP that contains only a single FK506-binding domain [20], also contains a single intron within its gene that coincides with *AIP*. Of the two remaining introns located in the gene encoding FKBP135, only one seems to be aligned near the intron positioned closely between *FKBP3* and *FKBP8*. The promiscuity and complexity of intron positions seen for the FKBP multigene family demonstrate the poor evolutionary conservation of intron positions between family members. This complexity of intron positions also correlates, in some cases, with the phylogeny shown in Fig. 4. In such cases where there is a correlation, the statistical support may be low.

Emergence of the FKBP Multigene Family

Although the genomic organization of the FKBP gene family members yields some insight into their evolutionary relationships, the introns within the genes of many FKBP family members seem to be unstable and the genes may be subject to multiple intron insertions and deletions. Because this limits the utility of intron position as a phylogenetic marker, we conducted a phylogenetic analysis of 100 distinct FK506-binding domains from diverse organisms.

On the basis of clustering patterns shown in Fig. 4, we conclude that the FKBP multigene family arose early in eukaryotic evolution. For example, the *Arabidopsis thaliana* FKBP61 gene is more closely related to the vertebrate FKBP51/52 genes than it is to other *A. thaliana* FKBP genes. Similarly, the *A. thaliana* FKBP53 gene is more closely related to the FKBP39 genes of *Drosophila* and yeast than it is to other *A. thaliana* FKBP genes. Another example is the human and mouse FKBP13 gene, which is more closely related to nematode FKBP13 and FKBP14 genes than to other human and mouse genes. Trandinh and colleagues [21] suggested

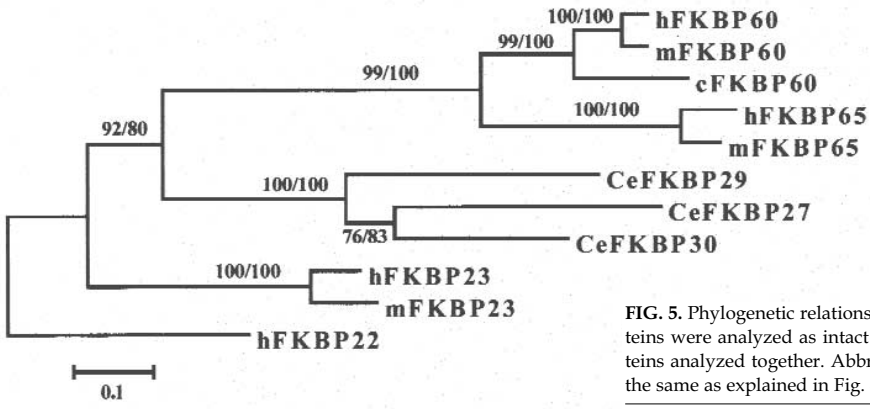

FIG. 5. Phylogenetic relationships among a subgroup of FKBP proteins. The proteins were analyzed as intact sequences with the domains of multi-domain proteins analyzed together. Abbreviations and numerical values along branches are the same as explained in Fig. 4. The tree is unrooted.

that the FKBP multigene family emerged before the divergence of prokaryotes and eukaryotes. We found evidence supported by relatively high bootstrap values that some prokaryotic FKBPs (for example, the FKBP25 gene of *Pseudomonas aeruginosa*) cluster among eukaryotes to the exclusion of other prokaryotes (Fig. 4), which is consistent with their interpretations. However, if we assume that the early emergence hypothesis [21] explains this pattern, we must assume that lateral gene transfer has not occurred, which may not necessarily be true. There is evidence to suggest that lateral gene transfer between eukaryotes and prokaryotes was a relatively frequent event during the early history of their divergence [22,23], and we cannot rule out this possibility for FKBPs. Therefore, it is currently unclear whether the early emergence hypothesis or lateral gene transfer hypothesis is correct.

Other, less ancient, gene duplication events also occurred during the evolutionary history of the FKBP multigene family. For example, *FKBP9* and *FKBP10*, which seem to be restricted to vertebrates, seem to have duplicated before the divergence of birds and mammals about 310 million years ago [24] based on the clustering patterns shown in Fig. 4. The exact timing of this event, however, is unclear because lower vertebrates, such as sharks, fish, and amphibians, have not yet been surveyed to determine if they possess these genes.

FKBP Gene Elongation: Intragenic Domain Duplication or Domain Shuffling?

As mentioned previously, some FKBPs possess multiple FK506-binding domains. To investigate hypotheses of gene elongation due to domain duplication or exon shuffling, the domains were analyzed as distinct sequences. If the sequences arose through intragenic gene duplication, and not exon shuffling, this will yield the same information as analyzing multiple-domain genes intact. This is best shown in the analysis of *FKBP9* and *FKBP10*. The relationships shown in the phylogeny reconstructed from the analysis of separate domains (Fig. 4) mirror those reconstructed from complete protein sequences containing all four FK506-binding domains (Fig. 5). This suggests that *FKBP9* and *FKBP10* were elongated due to intragenic domain duplications. If this were not the case and the domains were acquired from other genes through some other mechanism, the phylogenies in Figs. 4 and 5 would not have been congruent. The tree in Fig. 4 also provides information concerning the order of domain elongation. In this case, the first duplication event involved domain IV versus the ancestor of domains I–III of *FKBP9* and *FKBP10*. The next duplication involved domain I and the ancestor of domains II and III, and finally domains II and III split. The bootstrap support for the latter is high (Fig. 4), but the support for domain I clustering with domains II and III is not very high. Caution should thus be used when interpreting the order of domain duplications. Nevertheless, the results suggest that the domains duplicated before *FKBP9* and *FKBP10* split, as the domain from one gene clusters more closely with the corresponding domain from the other gene than to other domains from the same gene. This is supported well statistically (Fig. 4).

Similar patterns can be seen in the genes encoding vertebrate FKBP51/52 and plant FKBP61/62/63 genes, although they are more problematic. In particular, domain II from the vertebrate FKBP51/52 genes may be more closely related to the genes encoding plant FKBP61/62/63. However, the statistical support is low (Fig. 4), so homoplasy (that is, convergences, reversals, and/or parallelisms) may have confounded the groupings. If the groupings are correct, then a possible explanation may be provided by domain duplication accompanied by domain loss. Statistical support for this explanation is similarly low, and it is therefore difficult to discriminate between these two hypotheses with the current data.

doi:10.1006/geno.2002.6777, available online at http://www.idealibrary.com on IDEAL

DISCUSSION

The function of certain multigene families may be altered or enhanced through different mechanisms, including gene duplication or the addition/deletion of functional domains. Gene duplication may allow for the development of unique family members that may acquire different ligand specificities during the course of evolution, whereas domain shuffling or other methods of module exchange allows for the addition of functional domains and the creation of family members that possess distinct functions. The FKBP multigene family is a good candidate to study such processes. For example, the N-terminal portion of the FK506-binding domain is quite divergent between many family members and may have involved the recruitment of unrelated exons [19]. Additionally, the presence of extraneous domains, such as the more common EF-hand and TPR domains, have likely given many family members diverse functions over the course of evolution.

FKBP10 seems to have obtained four successive FK506-binding domains by gene elongation through a process of domain duplication. Such a model explains the strong conservation of domain structure and phylogenetic relationship within and between *FKBP9* and *FKBP10*. Additionally, divergence events found within the fourth and last domain, such as intron sliding and amino acid addition and deletion, as well as the phylogenetic distance from the other three domains, suggest that it is the parental domain of the remaining three FK506-binding domains. This is also supported by the phylogenetic relationships seen in Fig. 4 and discussed in the previous section. Similar mechanisms may also explain the evolution of the genes encoding vertebrate FKBP51/52 and plant FKBP61/62/63, although the evolution of these genes is more difficult to explain at the present time. Further analyses will be required to understand the evolutionary processes that have influenced these genes.

Previously, Pahl and Keller [25] demonstrated that *Streptomyces chrysomallus fkbB*, which encodes a 33-kDa protein containing two successive FK506-binding domains, is transcriptionally coupled to, and lies directly upstream of, *fkbA*, which encodes FKBP12. The authors concluded that this gene cluster of three consecutive FKBP domains likely originated from a double gene duplication of *fkbA*. Thus, domain duplication through gene elongation appears to be a fairly common mechanism through which many FKBP family members have obtained multiple PPIase domains.

Himukai and colleagues [26] suggested that there is a relationship between the subcellular localization of the FKBPs and the primary structure of the catalytic domain. The results of our phylogenetic analysis support this conclusion in that the FKBP domains containing unique cysteine residues specific to ER-localized FKBPs form a monophyletic group distinct from other FKBPs. The statistical reliability behind this grouping, however, is quite low (data not shown). If these relationships hold true, they suggest that subcellular compartmentalization evolved early in eukaryotic evolution, at least as early as the fungi-animal-plant divergence, about 1.6 billion years ago [27]. An interesting avenue for future study would be to examine the FKBPs of protists, particularly *Giardia* and other diplomads, as the split between these from rest of eukaryotes is believed to be the oldest in the eukaryotic lineage [28], and was recently estimated to have occurred over 2.2 billion years ago [29].

MATERIALS AND METHODS

Genomic screening of mouse Fkbp10. A genomic library (Stratagene 946313) in the λFixII vector derived from male 129/SvJ mouse DNA was screened (1.0×10^6 clones) using a random, ^{32}P-labeled, full-length, 1.74-kb FKBP65 cDNA [8]. Filters were prehybridized for 1 hour in Rapid-hyb buffer (Amersham, Piscataway, NJ) and hybridized in fresh Rapid-hyb buffer for 3 hours. Hybridizations and wash steps were carried out at 65°C. A single clone was identified, amplified, and digested with *Bam*HI or *Eco*RI after purifying phage DNA using standard methods [30]. Restriction fragments were ligated into pBluescript KS(-) and exon-containing clones were further identified by colony hybridization. Three overlapping clones were obtained that contained a combined total of 2.6 kb of the promoter and coding region up to nucleotide 1592 (within exon 8) of the published cDNA [15].

Exon–intron organization of mouse Fkbp10. Oligonucleotides were chosen based on the published FKBP65 cDNA sequence [15] and used to confirm all exon–intron boundaries. Due to their overall small size, complete intronic sequence for introns 2–8 was determined by direct sequencing of genomic clones. For determining the size of intron 1, compiled sequences obtained from clone B5 (Fig. 1A) were subtracted from its total length. The size of intron 9 was determined by sequencing cloned PCR products. Both the final size of intron 1, approximately 3.9 kb, and the absence of any introns after intron 9 were confirmed by PCR of mouse genomic DNA.

Computer analysis of the proximal promoter. The sequence of the 5′ region of mouse *Fkbp10* was examined for possible transcription factor binding sites against published matrices and the TRANSFAC profile database [31] using the MatInspector 5.1 program (http://genomatix.gsf.de). A 2.3-kb fragment continuing to the end of exon 1 was used for analysis. Transcription factors were chosen based on a rigid core similarity of 1.0, signifying the highest conserved bases of a matrix match exactly in the sequence, and a more stringent matrix similarity score of > 0.93, where 1.0 corresponds to a perfect match to the matrix [32].

Exon–intron organization of the human FKBP gene family. To determine the genomic organization of the FK506-binding domains of the human FKBP family, full-length human cDNAs submitted to GenBank (see Fig. 3 for accession numbers) were used to perform BLAST searches of the human Celera database (http://www.celera.com). Full-length or partial contiguous sequences of human genomic DNA were obtained for each FKBP. The resulting genomic sequences were then aligned with their respective cDNAs. Regions of genomic sequence containing FK506-binding domains were determined using an FK506-binding domain consensus sequence [9] and exon–intron organization of these domains was derived from analysis of non-overlapping sequences.

Computational analysis. The amino acid sequences of FK506-binding domains from various organisms were obtained from GenBank searches and aligned by visual inspection according to the consensus FK506-binding domain sequence [9]. Poisson distances were used in order to measure the extent of protein sequence divergence. Phylogenetic trees were reconstructed from these distances using the neighbor-joining method [33] and performed with the computer program MEGA2 [34]. Statistical reliability was assessed with 1000 bootstrap replications. The computer program PROTML in the PHYLIP version 3.6a2 [35] computer package was also used to construct maximum likelihood trees. In this case, JTT distances [36] were used to construct phylogenetic trees. Due to the computational intensity of the maximum likelihood method, only 100 bootstrap replicates were carried out to assess statistical reliability using this method.

ACKNOWLEDGMENTS

We thank Kevin Kennon for administrative assistance. This work was supported by National Institutes of Health grant HL60394 to E.C.D.

RECEIVED FOR PUBLICATION DECEMBER 17, 2001;
ACCEPTED MARCH 22, 2002.

REFERENCES

1. Galat, A. (2000). Sequence diversification of the FK506-binding proteins in several different genomes. *Eur. J. Biochem.* **267:** 4945–4959.
2. Schreiber, S. L., Albers, M. W., and Brown, E. J. (1993). The cell cycle, signal transduction, and immunophilins-ligand complexes. *Acc. Chem. Res.* **26:** 412–420.
3. Marks, A. R. (1996). Cellular functions of immunophilins. *Physiol. Rev.* **76:** 631–649.
4. Hamilton, G. S., and Steiner, J. P. (1998). Immunophilins: beyond immunosuppression. *J. Med. Chem.* **41:** 5119–5143.
5. Kiefhaber, T., Quaas, R., Hahn, U., and Schmid, F. X. (1990). Folding of ribonuclease T1. 2. Kinetic models for the folding and unfolding reactions. *Biochemistry* **29:** 3061–3070.
6. Davis, E. C. (2000). FK506-binding proteins in the secretory pathway. In *Immunophilins in the Brain* (B. G. Gold, G. Fischer, and T. Herdegen, Ed.), pp. 59–66. Prous Science, Barcelona, Spain.
7. Davis, E. C., Broekelmann, T. J., Ozawa, Y., and Mecham, R. P. (1998). Identification of tropoelastin as a ligand for the 65-kD FK506-binding protein, FKBP65, in the secretory pathway. *J. Cell Biol.* **140:** 295–303.
8. Patterson, C. E., Schaub, T., Coleman, E. J., and Davis, E. C. (2000). Developmental regulation of FKBP65: an ER-localized extracellular matrix binding-protein. *Mol. Biol. Cell* **11:** 3925–3935.
9. Kay, J. E. (1996). Structure-function relationships in the FK506-binding protein (FKBP) family of peptidyl-prolyl *cis-trans* isomerases. *Biochem. J.* **314:** 361–385.
10. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1991). Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science* **252:** 839–842.
11. Lepre, C. A., Thomson, J. A., and Moore, J. M. (1992). Solution structure of FK506 bound to FKBP-12. *FEBS Lett.* **302:** 89–96.
12. Lam, E., Martin, M., and Wiederrecht, G. (1995). Isolation of a cDNA encoding a novel human FK506-binding protein homolog containing leucine zipper and tetrarticopeptide repeat motifs. *Gene* **160:** 297–302.
13. Maruyama, T., and Furutani, M. (2000). Archaeal peptidyl prolyl cis-trans isomerases (PPIases). *Front. Biosci.* **5:** D821–836.
14. Betts, M. J., Guigo, R., Agarwal, P., and Russell, R. B. (2001). Exon structure conservation despite low sequence similarity: a relic of dramatic events in evolution? *EMBO J.* **20:** 5354–5360.
15. Coss, M. C., Winterstein, D., Sowder, R. C., and Simek, S. L. (1995). Molecular cloning, DNA sequence analysis, and biochemical characterization of a novel 65-kDa FK506-binding protein (FKBP65). *J. Biol. Chem.* **270:** 29336–29341.
16. Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acid Res.* **10:** 459–472.
17. Henrion, A. A., Martinez, A., Mattei, M. G., Kahn, A., and Raymondjean, M. (1995). Structure, sequence, and chromosomal location of the gene for USF2 transcription factors in mouse. *Genomics* **25:** 36–43.
18. Meng, X., Lu, X., Morris, C. A., and Keating, M. T. (1998). A novel human gene FKBP6 is deleted in Williams syndrome. *Genomics* **52:** 130–137.
19. Hendrickson, B. A., *et al.* (1993). Structural organization of the genes encoding human and murine FK506-binding protein (FKBP)13 and comparison to *FKBP1*. *Gene* **134:** 271–275.
20. Ishikawa, K., *et al.* (1998). Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. *DNA Res.* **5:** 169–176.
21. Trandinh, C. C., Pao, G. M., and Saier, M. H. (1992). Structural and evolutionary relationships among the immunophilins: two ubiquitous families of peptidyl-prolyl *cis-trans* isomerases. *FASEB J.* **6:** 3410–3420.
22. Doolittle, W. F. (2000). The nature of the universal ancestor and the evolution of the proteome. *Curr. Opin. Struct. Biol.* **10:** 355–358.
23. Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. USA* **97:** 8392–8396.
24. Kumar, S., and Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. *Nature* **392:** 917–919.
25. Pahl, A., and Keller, U. (1994). *Streptomyces chrysomallus* FKBP-33 is a novel immunophilin consisting of two FK506 binding domains; its gene is transcriptionally coupled to the FKBP-12 gene. *EMBO J.* **13:** 3472–3480.
26. Himukai, R., Kuzuhara, T., and Horikoshi, M. (1999). Relationship between the subcellular localization and structures of catalytic domains of FKBP-type PPIases. *J. Biochem. (Tokyo)* **126:** 879–888.
27. Wang, D. Y., Kumar, S., and Hedges, S. B. (1999). Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc. R. Soc. Lond. B* **266:** 163–171.
28. Roger, A. J., *et al.* (1998). A mitochondrial-like chaperonin 60 gene in Giardia lamblia: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc. Natl. Acad. Sci. USA* **95:** 229–234.
29. Hedges, S. B., *et al.* (2001). A genomic timescale for the origin of eukaryotes. *BMC Evol. Biol.* **1:** 4.
30. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Bacteriophage lambda vectors. In *Molecular Cloning: A Laboratory Manual* (C. Nolan, Ed.), pp. 2.2–2.125. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
31. Wingender, E., *et al.* (2001). The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.* **29:** 281–283.
32. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23:** 4878–4884.
33. Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406–425.
34. Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001). MEGA2. Institute for Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, PA, http://www.megasoftware.net.
35. Felsenstein, J. (2001). PHYLIP: Phylogeny Inference Package. Version 3.6a2. Department of Genetics, University of Washington, Seattle, http://evolution.genetics.washington.edu/phylip.html.
36. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8:** 275–282.

Sequence data from this article have been deposited with the DDBJ/EMBL/GenBank Data Libraries under accession numbers AF456412 and AF456413 (Fkbp10).