Available online at www.sciencedirect.com
ScienceDirect
Gene 383 (2006) 64–70
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Origin and evolution of gene for prolactin-induced protein

Takashi Kitano [a,*], Wei Tian [a], Kazuo Umetsu [a], Isao Yuasa [b], Kentaro Yamazaki [a], Naruya Saitou [c], Motoki Osawa [d]

[a] *Department of Experimental and Forensic Pathology, Yamagata University School of Medicine, 2-2-2 Iidanishi, Yamagata 990-9585, Japan*
[b] *Division of Legal Medicine, Faculty of Medicine, Tottori University, Yonago 683-8503, Japan*
[c] *Division of Population Genetics, National Institute of Genetics, Mishima 411-8540, Japan*
[d] *Department of Forensic Medicine, Tokai University School of Medicine, Isehara, 259-1193, Japan*

Received 29 March 2006; received in revised form 10 July 2006; accepted 14 July 2006
Available online 28 July 2006
Received by T. Gojobori

Abstract

Prolactin-induced protein (PIP) is a small protein secreted into the fluid in several glands. We determined the *PIP* coding sequences of 5 hominoid species and estimated the numbers of synonymous and nonsynonymous substitutions for each branch of the mammalian *PIP* gene tree. The branch connecting hominoids and Old World monkeys showed significantly higher nonsynonymous than synonymous substitutions. These changes tended to be accumulated in the fibronectin-binding domain. Many other primate branches also showed higher nonsynonymous than synonymous substitutions, thus suggesting that the *PIP* genes of primates have experienced some kind of positive selection. We also considered the phylogenetic relationship of the *PIP* gene with the alpha-2-macroglobulin gene family. The results indicate that the *PIP* gene arose by partial gene duplication from a member of the alpha-2-macroglobulin gene family after the divergence between amphibians and other tetrapods.

Keywords: Phylogenetic tree; Nonsynonymous substitution; Positive selection; Mammal; Gene duplication

1. Introduction

The hormonally responsive prolactin-induced protein (PIP) is expressed in various normal exocrine organs such as sweat, salivary, and lacrimal glands, as well as in seminal plasma (Haagensen and Mazoujian, 1986). Although its functions in secretory fluids are not well understood, the protein is reported to be a potent inhibitor of T lymphocyte programmed cell death induced by cross-linking of CD4 and T cell receptor (Gaubin et al., 1999). It was also reported that the protein is an aspartyl proteinase with specific fibronectin-degrading ability (Caputo et al., 2000). The CD4-binding domains and fibronectin-binding domains on the protein have been identified (Basmaciogullari et al., 2000; Caputo et al., 2003).

The human *PIP* gene is located on chromosome 7q32–36 (Myal et al., 1989), and the entire gene is about 7 kb long. The gene has 4 exons (Myal et al., 1991) comprising 146 amino acids (Murphy et al., 1987). PIP shares homology with another secretory protein, the seminal vesicle autoantigen (SVA). SVA, a secretory protein present in mouse seminal plasma, exerts an inhibitory effect on the motility of spermatozoa by interacting with the cell membrane phospholipid (Huang et al., 1999).

Evolutionary studies on the *PIP* genes have yielded limited data (Osawa et al., 2004; Clark and Swanson, 2005), with the results suggesting the operation of positive selection on the gene. Thus, in this study, we sequenced the *PIP* coding sequences (CDS) of 5 primate species, and carried out a more detailed evolutionary analysis of this gene. We also examined the phylogenetic relationship of the *PIP* gene with other genes having homologous sequences, such as those of alpha-2-macroglobulin (A2M), pregnancy-zone protein (PZP), alpha-2-macroglobulin-like 1 (A2ML1), and ovostatin 2 (OVOS2), and considered the origin of the mammalian *PIP* gene.

Abbreviations: PIP, prolactin-induced protein; A2M, alpha-2-macroglobulin; FN, fibronectin.

* Corresponding author. Tel.: +81 23 628 5271; fax: +81 23 628 5273.
E-mail address: tkitano@med.id.yamagata-u.ac.jp (T. Kitano).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.07.014

2. Materials and methods

2.1. DNA sequencing of primate PIP genes

Genomic DNAs from 2 common chimpanzees (*Pan troglodytes*), 3 gorillas (*Gorilla gorilla*), 1 orangutan (*Pongo pygmaeus*), 1 agile gibbon (*Hylobates agilis*), and 1 siamang (*Symphalangus syndactylus*) were used for DNA sequencing. PCR amplification was performed in a volume of 20 μl containing 50 ng of DNA, 1× PCR buffer, 0.2 mM concentration of each dNTP, 0.5 μM concentration of each oligonucleotide primer, 1.5 mM $MgCl_2$, and 1 U of AmpliTaq Gold polymerase (Applied Biosystems). All PCR primers were designed based on intron regions to amplify full CDSs, based on the human (NC_000007) and chimpanzee (AB098482) sequences in the DDBJ/EMBL/GenBank international nucleotide sequence database. A list of primers used in this study is available from one of the authors (TK). The typical PCR conditions used in this study consisted of 40 cycles of 30-s denaturation at 95 °C, followed by 15-s primer annealing at 55 °C and 1-min extension at 72 °C. Immediately proceeding these cycles, a 10-min hot-start step at 95 °C was included. The PCR products were confirmed by 1.5% agarose gel electrophoresis and purified by using a QIAquick PCR purification kit (Qiagen). The purified PCR products were sequenced with a BigDye terminator cycle sequencing kit version 1.1 and an ABI Prism 3100 genetic analyzer (Applied Biosystems). Both strands were sequenced, and 2 overlapping peaks on a site were regarded to indicate heterozygosity, and the site was designated by following Nomenclature Committee of the International Union of Biochemistry (NC-IUB).

2.2. Analysis of PIP coding sequences

The CDSs of the *PIP* gene for the following 8 mammals were retrieved from the nucleotide sequence database: human (*Homo sapiens*, J03460), chimpanzee (AB098482), Japanese macaque (*Macaca fuscata*, AB098481), cow (*Bos taurus*, AB098480), rabbit (*Oryctolagus cuniculus*, AB098478), guinea pig (*Cavia porcellus*, AB098479), mouse (*Mus musculus*, S73282), and rat (*Rattus norvegicus*, AF054270). The CLUSTALW program (Thompson et al., 1994) was used for the multiple alignment. Aligned regions without gaps were used for subsequent analyses. A phylogenetic tree was constructed by using the neighbor-joining method (Saitou and Nei, 1987) along with the two-parameter method (Kimura, 1980). MEGA3 software (Kumar et al., 2004) was used for the tree construction. The ancestral sequences on the tree were reconstructed by the likelihood-based Bayesian method (Yang et al., 1995) using the BASEML program in the PAML package (Yang, 1997). The numbers of synonymous and nonsynonymous substitutions for each branch were estimated by using the modified Nei and Gojobori's method (Nei and Gojobori, 1986; Zhang et al., 1998). The ratio (R) of transitional changes to transversional changes for the data set was obtained by counting the total numbers of transitional and transversional changes that were observed in the entire phylogenetic tree. This method yielded $R = 296/254 = 1.17$; therefore, we assumed $R = 1$ in subsequent computations. Fisher's exact tests for synonymous and nonsynonymous substitutions were performed according to Zhang et al. (1998).

2.3. Polymorphism data for the human PIP gene

We used genomic DNAs from 30 Japanese and 30 Germans to sequence 2 regions (I and II) of the human *PIP* gene. The regions I (615 bp) and II (664 bp) contained exon 2 and exons 3–4, respectively (Supplementary Fig. 1). PCR and sequencing were done as described above. Chimpanzee (NW_109346) and Japanese macaque (AB098481) sequences were used in order to determine ancestral states for each SNP (single nucleotide polymorphism). Two measures of nucleotide variability, π (Nei, 1987) and θ_W (Watterson, 1975), were calculated. Tajima's (1989) D and Fay and Wu's (2000) H were calculated to test for deviations from the standard neutral model by using the allele frequency spectrum. A program for performing these statistics, kindly provided by J.C. Fay, was used.

2.4. Analysis of the relationship between PIP and its related genes

We performed PSI-BLAST (Altschul et al., 1997) with 10 passes against the NCBI non-redundant protein database. The human PIP (NP_002643) was used as a query for the homology search. From the PSI-BLAST result, we collected amino acid sequences that gave hit length of more than 100 amino acid residues and higher than 30% positive (meaning identical or similar) residues without alignment gaps. Human, cow, dog, and mouse were used as representative mammals, with the chicken (*Gallus gallus*) for birds, African clawed frog (*Xenopus laevis*) for amphibians, and pufferfish (*Tetraodon nigroviridis*) and zebrafish (*Danio rerio*) for fish. We also obtained homologous sequences from the Japanese lamprey (*Lethenteron japonicum*) and 7 invertebrate species, i.e., purple sea urchin (*Strongylocentrotus purpuratus*), sea scallop (*Chlamys farreri*), horseshoe crab (*Limulus sp.*), red flour beetle (*Tribolium castaneum*), honey bee (*Apis mellifera*), soft tick (*Ornithodoros moubata*), and kuruma shrimp (*Marsupenaeus japonicus*). The list of these sequences is given in Supplementary Table 1. Redundant sequences and sequences that contained an alignment gap of more than 5 contiguous amino acid residues were not used for the analysis (see Supplementary Table 1). In total, 92 amino acid sequences were used for the tree reconstruction of the *PIP* genes and their related genes. Furthermore we carried out tBLASTn searches against gss (genome survey sequence), wgs (whole genome shotgun sequence), and htgs (unfinished high throughput genomic sequences) databases for chicken, African clawed frog, zebrafish, and pufferfish (*Takifugu rubripes*) by using the human *PIP* (NP_002643) as a query. However, any other additional sequences were not observed. The T-Coffee program (Notredame et al., 2000) was used for the multiple alignment. For maximum likelihood phylogenetic inference, first, a neighbor-joining tree was constructed with the JTT model (Jones et al., 1992), followed by the branch swapping (nearest neighbor interchange) method,

with the neighbor-joining tree used as the initial tree. The PROTML program in the MOLPHY package (Adachi and Hasegawa, 1996) was used.

3. Results and discussion

3.1. Synonymous and nonsynonymous substitutions in mammalian PIP genes

We sequenced the *PIP* genes from 2 common chimpanzees, 3 gorillas, 1 orangutan, 1 siamang and 1 agile gibbon (DDBJ/EMBL/GenBank International Nucleotide Sequence Database accession numbers are AB251899–AB251906). Fig. 1 shows a multiple alignment of translated amino acid sequences of the mammalian PIP proteins. The chimpanzee sequences determined in this study were identical to the sequence encoded by the chimpanzee *PIP* gene deposited in the database except for 1 synonymous change (T342C). In the gorillas, 2 polymorphic sites were observed (G343A and A359G). Both substitutions could change the type of amino acid and were located in the neighborhood of each other (Fig. 1). We used 343=G and 359=A for the following analyses as a representative of gorillas.

Fig. 2 shows the phylogenetic relationship among mammalian *PIP* genes and the numbers of synonymous and nonsynonymous substitutions on each branch. The d*N*/d*S* ratio of the number of nonsynonymous substitutions per nonsynonymous site (d*N*) to the number of synonymous substitutions per synonymous site (d*S*) was estimated for each branch. Nine branches showed a d*N*/d*S* value higher than 1, and 8 of them were in the primate lineages. Application of Fisher's exact test showed that the number of nonsynonymous substitutions was significantly higher than that of the synonymous ones ($p=0.02$) in the branch that connected hominoids and an Old World monkey, where 18 nonsynonymous substitutions (*n*), 1 synonymous substitutions (*s*), 112.7 potential synonymous sites (*S*), and 310.3 potential nonsynonymous sites (*N*) were estimated. Table 1 shows the numbers of amino acid changes for each domain among the primate PIP proteins. On the branch of Node *b*–Node *c*, 5 out of 18 amino acid changes were clustered in the FN-1 domain. The rate of amino acid changes in this domain was 2 to 3 times higher than that for the other regions. This domain also accumulated amino acid changes in the branch that connected to an Old World monkey (Node *b*–Macaque). In contrast, amino acid changes in hominoids tended to accumulate in the FN-2 domain. The NH_2-terminal region, which is the first CD4-binding domain, was conserved among the primate PIP proteins. Furthermore, 4 cysteine residues for 2 pairs of disulfide bonds (Schaller et al., 1991) were conserved among the mammals examined (Fig. 1). By using site-specific mutagenesis, Caputo et al. (2000) demonstrated that an aspartate residue (they called it Asp^{22}) was required for proteinase activity. This aspartate residue is located in the FN-1 domain (Fig. 1), and is shared by human and

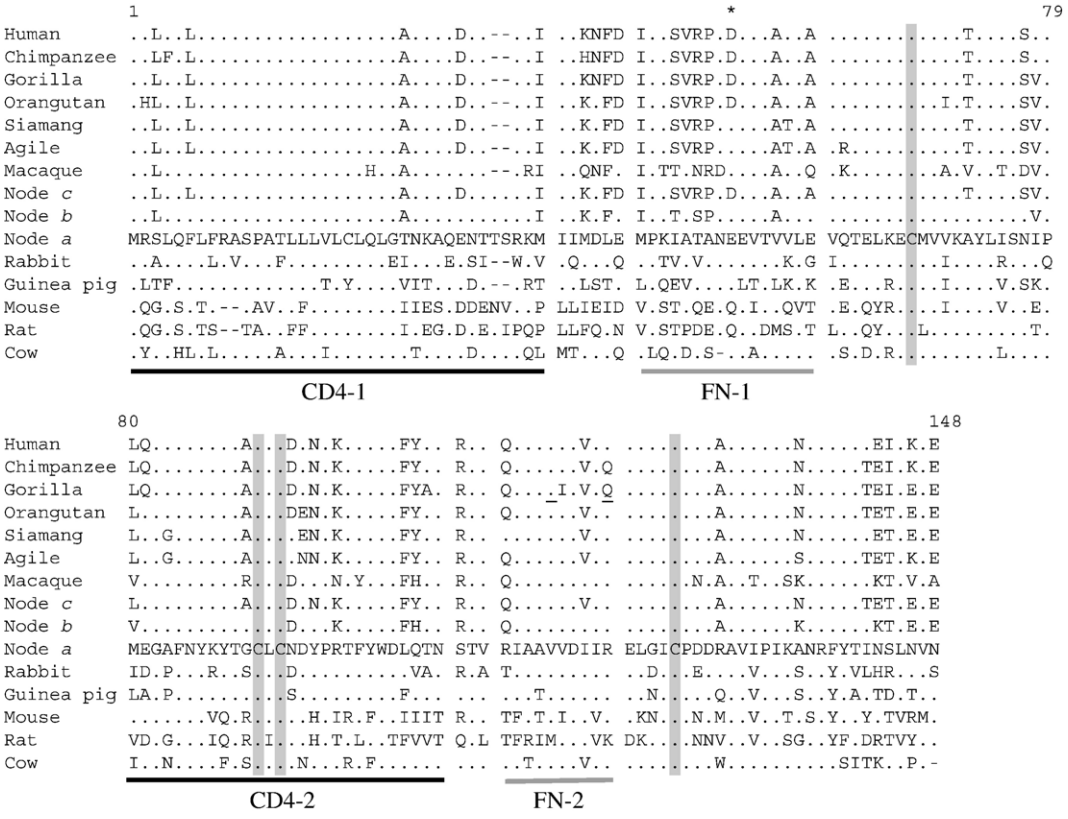

Fig. 1. Amino acid sequences for present-day and ancestral PIP proteins. Dots show the same amino acids as those of the sequence at node *a*. Four cysteine residues (67, 91, 93, and 125) for 2 pairs of disulfide bonds are shown on a grey background. CD4-1 and CD4-2 indicate CD4-binding domains; and FN-1 and FN-2, fibronectin-binding domains. The 2 underlined residues were polymorphic among 3 gorillas (V115I and Q120R). The aspartate residue (Asp^{22}) is indicated by the asterisk.

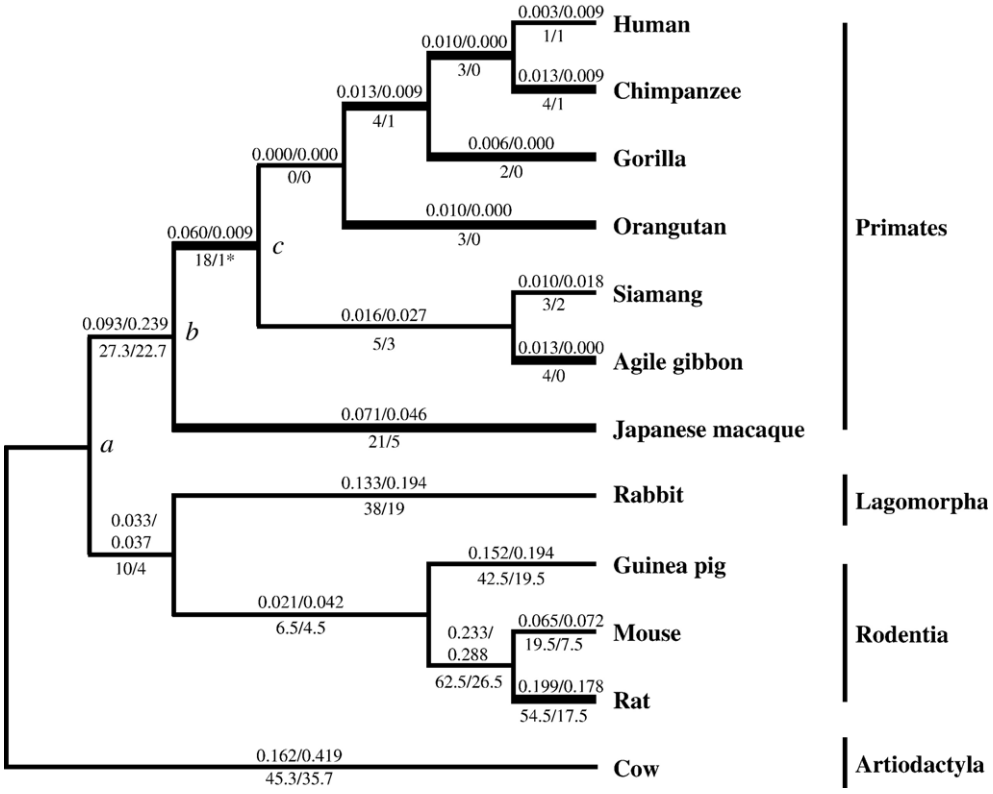

Fig. 2. A phylogenetic tree of mammalian *PIP* genes. d*N*/d*S* are indicated above each branch, and a branch that has a d*N*/d*S* value higher than 1 is shown by a thick line. The numbers of nonsynonymous (*n*) and synonymous (*s*) substitutions are indicated as *n*/*s* below each branch. Three ancestral nodes are shown by *a* (common ancestor of primates, rodents, and rabbit), *b* (common ancestor of primates), and *c* (common ancestor of hominoids). The statistical significance was tested by Fisher's exact test (*, less than 5%).

great apes. If so, the proteinase activity of PIP was acquired during the evolution of the great apes. Further study is needed to elucidate the time of acquisition of proteinase activity by PIP proteins.

Although the function of the PIP protein is not well understood, it has been suggested that PIP is a binding factor playing a role in host defense against infections (Schenkels et al., 1997; Gaubin et al., 1999). The majority of genes are considered to be evolving under neutral mutation pressure. However, some genes are evolving through positive selection. For example, several genes are considered to be positively selected by interaction between host defense systems and pathogens (Hughes and Nei, 1988, 1989; Ina and Gojobori, 1994; Kitano et al., 1998; Kitano and Saitou, 1999; Schaner et al., 2001; Baum et al., 2002; Osada et al., 2002; Sumiyama et al., 2002; Filip and Mundy, 2004; etc.). The high rate of nonsynonymous substitutions for the primate *PIP* genes also suggests the existence of positive selection on this gene, which selection might be caused by some kind of interaction between the PIP protein and pathogens.

To clarify whether the higher number of nonsynonymous substitutions of the *PIP* gene in the primate lineages was a result of positive selection or relaxation of selection, we sequenced and analyzed polymorphisms in the genome sequence of the human *PIP* gene (DDBJ/EMBL/GenBank International Nucleotide Sequence Database accession numbers are AB267097–AB267216). In total, 3 polymorphic sites in Japanese and 5 in Germans were observed in the 2 sequenced regions (I and II). Table 2 summarizes the statistics on the polymorphisms of the human *PIP* gene (see also Supplementary Fig. 1). The nucleotide diversity (π) determined by the SNP Consortium (Sachidanandam et al., 2001) for the entire human genome is 0.075%, and the corresponding values for the human *PIP* genes reported here (Region I+II) were 0.022% for Japanese and

Table 1
Number of amino acid changes for each domain of primate PIP proteins

	CD4-1	IT-1	FN-1	IT-2	CD4-2	IT-3	FN-2	C-terminal	Total
No. of residues	35	6	16	20	28	3	10	28	146
Node *b*–Node *c*	2 (0.06)	1 (0.17)	5 (0.31)	2 (0.10)	4 (0.14)	0 (0.00)	1 (0.10)	3 (0.11)	18
Node *b*–Macaque	2 (0.06)	2 (0.33)	5 (0.31)	5 (0.25)	3 (0.11)	0 (0.00)	0 (0.00)	5 (0.18)	22
In hominoids	2 (0.06)	2 (0.33)	2 (0.13)	3 (0.15)	7 (0.25)	0 (0.00)	5 (0.50)	6 (0.21)	27

The number of amino acid changes per residue for each domain is shown in parentheses. CD4-1 and CD4-2: 2 CD4-binding domains, FN-1 and FN-2: 2 fibronectin-binding domains, IT-1, IT-2, and IT-3: 3 interval regions, C-terminal: C-terminal region.

Table 2
Polymorphisms in humans at the *PIP* locus

	Region I (613 bp)		Region II (664 bp)		Region I+II (1277 bp)	
	Japanese	Germans	Japanese	Germans	Japanese	Germans
S	2	1	1	4	3	5
π (%)	0.026	0.034	0.019	0.123	0.022	0.080
θ_W (%)	0.070	0.035	0.032	0.129	0.050	0.084
D_{Tajima}	−1.072	−0.029	−0.526	−0.105	−1.101	−0.101
$H_{Fay\ \&\ Wu}$	−1.612**	−1.377**	−1.645**	−0.232	−3.258***	−1.609*

S: number of segregating sites; π: average number of nucleotide differences per site between 2 sequences; θ_W: nucleotide diversity based on the proportion of segregating sites in a sample; D_{Tajima}: Tajima's D (Tajima, 1989); $H_{Fay\ \&\ Wu}$: Fay and Wu's H (Fay and Wu, 2000). Asterisks indicate the significance level (*10%, **5%, ***1%).

0.080% for Germans. Although the π in region II of Germans was slightly high, these values are comparable with that value for the entire human genome. Tajima's D, which examines the difference (D) between π and θ_W, was negative for all estimates. However, none of the D values differed significantly from the neutral expectation of zero. In contrast, Fay and Wu's H-test showed significance values except for the region II of Germans. The results revealed an excess of derived variants at high frequency that were likely to have been influenced by hitchhiking of the derived changes under positive selection on the human *PIP* gene. Thus, our data suggest that the higher number of nonsynonymous substitutions of the *PIP* gene in the primate lineages was caused by positive selection rather than by relaxation of selection.

3.2. Phylogenetic relationship between the PIP gene and the A2M gene family, and origin of the PIP gene

We searched the NCBI protein database by PSI-BLAST using the human PIP amino acid sequence as a query, and observed weak homologies with the C-terminal region of members of the *A2M* gene families including *PZP*, *OVOS2*, and *A2ML1* (Supplementary Table 1). The average percent of amino acid identities between PIP and A2M, PZP, A2ML1, OVOS2 in mammalians were 16.1%, 16.3%, 17.1%, and 16.7%, respectively. Fig. 3A shows a phylogenetic tree among the *A2M* gene family with the *PIP* and *SVA* genes. The amino acid length

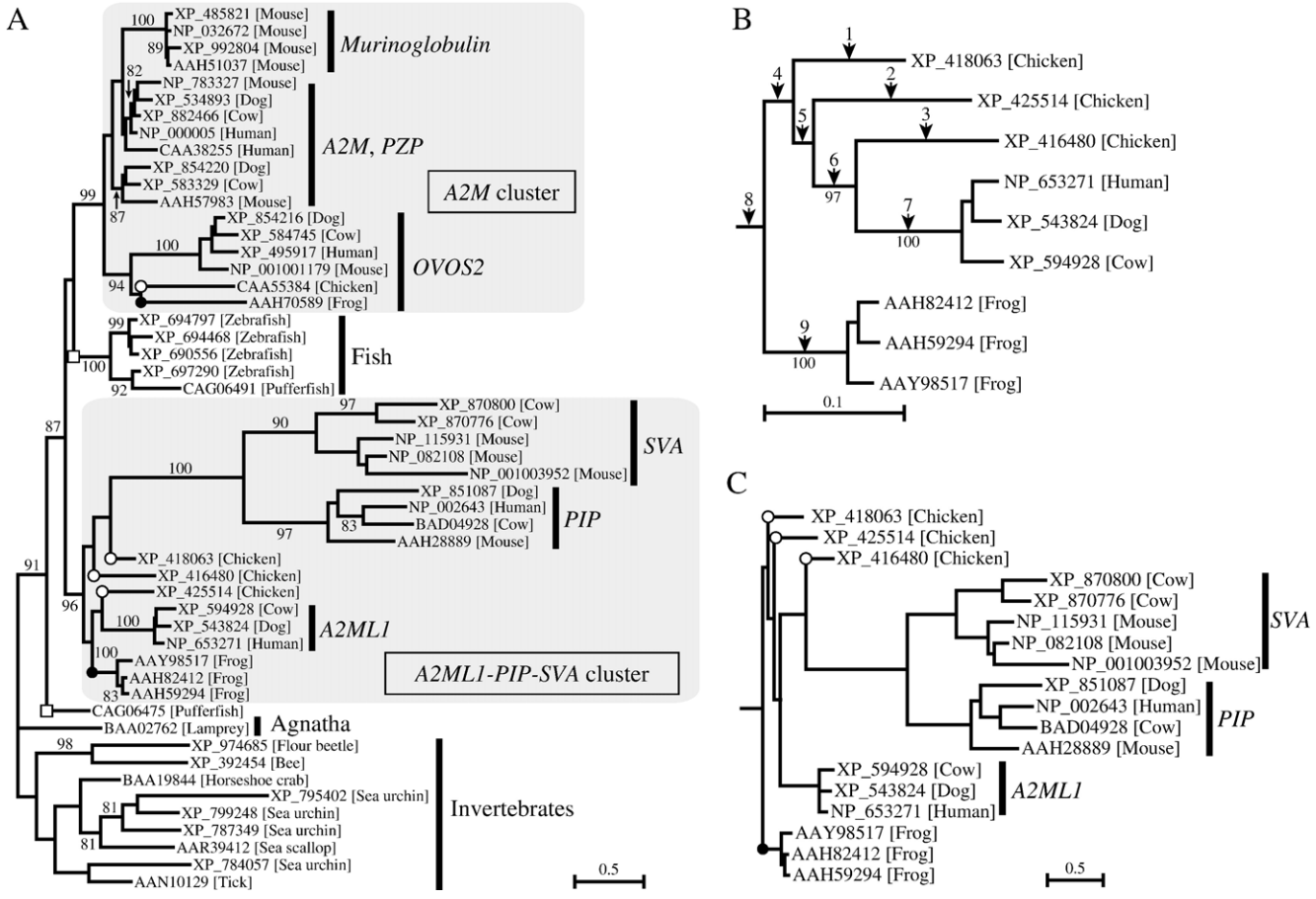

Fig. 3. A phylogenetic tree of the *A2M* gene family constructed by the maximum likelihood method (A). Bootstrap values greater than 80% are shown on each branch. White squares, black circles, and white circles indicate internal nodes connecting fish, frog, and chicken, respectively. A phylogenetic tree of the *A2ML1* cluster excluding the *PIP* and *SVA* genes constructed by the maximum likelihood method (B). Bootstrap values greater than 80% are shown on each branch. Arrows with numbers indicate candidate branches connecting the *PIP–SVA* cluster. A phylogenetic tree of the *A2ML1–PIP–SVA* cluster constructed by the maximum likelihood method (C). Black circles and white circles indicate internal nodes connecting fish, frog, and chicken, respectively.

compared was 92. The phylogenetic tree was constructed by using the region of exon 2 to exon 4 for the human *PIP* gene which corresponds to a part of exon 19 to exon 21 for the human *A2M* gene. In fact, *A2M* gene family members shared the location of exon–intron boundaries (Supplementary Fig. 2) excluding invertebrate sequences. All invertebrate sequences formed a cluster, and a lamprey sequence was located on the middle between the vertebrate and invertebrate sequences. Thus, the lamprey and invertebrates sequences can be used as out-groups of the *A2M* gene family for vertebrates. Two large clusters were observed in vertebrates; the *A2M* cluster (including *A2M*, *PZP*, *OVOS2*, and *Murinoglobulin*) and the *A2ML1–PIP–SVA* cluster. Although the branching position of the pufferfish sequence (CAG06475) did not form a cluster with other fish sequences, the fish cluster formed a cluster with the *A2M* cluster. There are 2 possibilities for this result; one is that fish genes homologous to the *A2ML1–PIP–SVA* cluster genes were deleted or undetected, and the other is that the branching point of the fish homologous genes is not correct. It is also possible to assume that the gene duplication between the *A2M* cluster and the *A2ML1–PIP–SVA* cluster occurred after the divergence between fish and other vertebrates. In any case, our data suggest that the *A2ML1–PIP–SVA* cluster emerged after the divergence between fish and other vertebrates. Frog homologous genes were observed in the both clusters.

To obtain more reliable phylogenetic relationships in the *A2ML1–PIP–SVA* cluster, we used A2M domain sequences and constructed a phylogenetic tree excluding the *PIP* and *SVA* genes (B). The amino acid length compared was 339, and the pufferfish (CAG06475) was used as an out-group. Three frog sequences formed a cluster. It is not clear whether these 3 genes were allelic or represented different loci. Three chicken sequences did not form a cluster and their genes were located at different chromosomal positions. By using this topology (Fig. 3B), we added the *PIP–SVA* cluster for each branch, indicated by the arrow, with number and estimated likelihood values for each tree. Fig. 3C shows the maximum likelihood tree for the *A2ML1–PIP–SVA* cluster. The amino acid length compared was 104. Although the likelihood values of 8 trees excluding tree number 7 were not significantly different, the *PIP–SVA* genes formed a cluster with the predicted chicken gene (XP_416480), which is not probably an orthologue of the *PIP–SVA* genes. This predicted chicken gene was located on chromosome 1 (chr.1) in the 68,370K–68,416K region. In contrast, human and mouse synteny around the *PIP* gene was conserved; and the corresponding region of the chicken gene was located on chr.1 in the 71,750K–72,130K region, although the location of the probable chicken *PIP* orthologous gene was in the genome sequence gap. Thus it is not clear whether a chicken *PIP* orthologous gene exists or not. We also BLAST-searched against the chicken EST database using human *PIP* gene as a query, but obtained no *PIP* homologous sequence.

Although the exact branching point of the *PIP* gene is not clear, it is possible to predict that after the divergence between amphibians and other tetrapods, the *PIP* gene have emerged by a partial gene duplication, since mammalian *PIP* genes are shorter than other *A2M* gene members. After the gene duplication, the ancestral *PIP* gene might have acquired a new specific function and might have become expressed in sweat, salivary, and lacrimal glands, and in the seminal plasma in the mammalian lineage. The gene duplications that produced *SVA* genes occurred in the mammalian lineage as described elsewhere (Yoshida et al., 2001; Osawa et al., 2004). Based on analysis of the chicken genome sequence (International Chicken Genome Sequencing Consortium, 2004), some genes such as those for casein milk proteins and salivary-associated proteins (statherin and histatins) arose in the mammalian lineage after the divergence between mammalians and birds. It is likely that the *PIP* gene is also a mammalian-specific gene and appeared during mammalian evolution.

Acknowledgements

We are grateful to the Great Ape Information Network (GAIN) for providing us chimpanzee, gorilla, and orangutan samples, to Ueno Zoo, Tokyo, for providing us with the gorilla samples, and to Dr. Takenaka (deceased) for providing us the gibbon DNA samples.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2006.07.014.

References

Adachi, J., Hasegawa, M., 1996. MOLPHY: Program for Molecular Phylogenetics, Ver. 2.2. Institute of Statistical Mathematics, Tokyo.

Altschul, S.F., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Basmaciogullari, S., et al., 2000. Mapping the CD4 binding domain of gp17, a glycoprotein secreted from seminal vesicles and breast carcinomas. Biochemistry 39, 5332–5340.

Baum, J., Ward, R.H., Conway, D.J., 2002. Natural selection on the erythrocyte surface. Mol. Biol. Evol. 19, 223–229.

Caputo, E., Manco, G., Mandrich, L., Guardiola, J., 2000. A novel aspartyl proteinase from apocrine epithelia and breast tumors. J. Biol. Chem. 275, 7935–7941.

Caputo, E., et al., 2003. Structural study of GCDFP-15/gp17 in disease versus physiological conditions using a proteomic approach. Biochemistry 42, 6169–6178.

Clark, N.L., Swanson, W.J., 2005. Pervasive adaptive evolution in primate seminal proteins. PLoS Genet. 1, 335–342.

Fay, J.C., Wu, C.I., 2000. Hitchhiking under positive Darwinian selection. Genetics 155, 1405–1413.

Filip, L.C., Mundy, N.I., 2004. Rapid evolution by positive Darwinian selection in the extracellular domain of the abundant lymphocyte protein CD45 in primates. Mol. Biol. Evol. 21, 1504–1511.

Gaubin, M., et al., 1999. Potent inhibition of CD4/TCR-mediated T cell apoptosis by a CD4-binding glycoprotein secreted from breast tumor and seminal vesicle cells. J. Immunol. 162, 2631–2638.

Haagensen, D.E., Mazoujian, G., 1986. Biochemistry and immunohistochemistry of fluid proteins of the breast in gross cystic disease. In: Haagensen, D.E. (Ed.), Diseases of the Breast. W.B. Saunders Co, Philadelphia, pp. 474–500.

Huang, Y.H., Chu, S.T., Chen, Y.H., 1999. Seminal vesicle autoantigen, a novel phospholipid-binding protein secreted from luminal epithelium of mouse seminal vesicle, exhibits the ability to suppress mouse sperm motility. Biochem. J. 343, 241–248.

Hughes, A.L., Nei, M., 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335, 167–170.

Hughes, A.L., Nei, M., 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. U. S. A. 86, 958–962.

Ina, Y., Gojobori, T., 1994. Statistical analysis of nucleotide sequences of the hemagglutinin gene of human influenza A viruses. Proc. Natl. Acad. Sci. U. S. A. 91, 8388–8392.

International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Kitano, T., Saitou, N., 1999. Evolution of Rh blood group genes have experienced gene conversions and positive selection. J. Mol. Evol. 49, 615–626.

Kitano, T., Sumiyama, K., Shiroishi, T., Saitou, N., 1998. Conserved evolution of the Rh50 gene compared to its homologous Rh blood group gene. Biochem. Biophys. Res. Commun. 249, 78–85.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Murphy, L.C., Tsuyuki, D., Myal, Y., Shiu, R.P., 1987. Isolation and sequencing of a cDNA clone for a prolactin-inducible protein (PIP). Regulation of PIP gene expression in the human breast cancer cell line, T-47D. J. Biol. Chem. 262, 15236–15241.

Myal, Y., Gregory, C., Wang, H., Hamerton, J.L., Shiu, R.P., 1989. The gene for prolactin-inducible protein (PIP), uniquely expressed in exocrine organs, maps to chromosome 7. Somat. Cell Mol. Genet. 15, 265–270.

Myal, Y., Robinson, D.B., Iwasiow, B., Tsuyuki, D., Wong, P., Shiu, R.P., 1991. The prolactin-inducible protein (PIP/GCDFP-15) gene: cloning, structure and regulation. Mol. Cell. Endocrinol. 80, 165–175.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217.

Osada, N., et al., 2002. Search for genes positively selected during primate evolution by 5′-end-sequence screening of cynomolgus monkey cDNAs. Genomics 79, 657–662.

Osawa, M., Horiuchi, H., Tian, W., Kaneko, M., 2004. Divergent evolution of the prolactin-inducible protein gene and related genes in the mouse genome. Gene 325, 179–186.

Sachidanandam, R., et al., 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409, 928–933.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Schaller, J., Akiyama, K., Kimura, H., Hess, D., Affolter, M., Rickli, E.E., 1991. Primary structure of a new actin-binding protein from human seminal plasma. Eur. J. Biochem. 196, 743–750.

Schaner, P., et al., 2001. Episodic evolution of pyrin in primates: human mutations recapitulate ancestral amino acid states. Nat. Genet. 27, 318–321.

Schenkels, L.C., Walgreen-Weterings, E., Oomen, L.C., Bolscher, J.G., Veerman, E.C., Nieuw Amerongen, A.V., 1997. In vivo binding of the salivary glycoprotein EP-GP (identical to GCDFP-15) to oral and non-oral bacteria detection and identification of EP-GP binding species. Biol. Chem. 378, 83–88.

Sumiyama, K., Saitou, N., Ueda, S., 2002. Adaptive evolution of the IgA hinge region in primates. Mol. Biol. Evol. 19, 1093–1099.

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

Thompson, J.D., Gibson, T.J., Higgins, D.G., 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7, 256–276.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yang, Z., Kumar, S., Nei, M., 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141, 1641–1650.

Yoshida, M., Kaneko, M., Kurachi, H., Osawa, M., 2001. Identification of two rodent genes encoding homologues to seminal vesicle autoantigen: a gene family including the gene for prolactin-inducible protein. Biochem. Biophys. Res. Comm. 281, 94–100.

Zhang, J., Rosenberg, H.F., Nei, M., 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. U. S. A. 95, 3708–3713.