Phylogenetic Analyses Uncover a Novel Clade of Transferrin in Nonmammalian Vertebrates

Hirzahida Mohd-Padil,[1] Adura Mohd-Adnan,*[,1] and Toni Gabaldón*[,2,3]

[1]School of Biosciences & Biotechnology, Faculty of Science Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia

[2]Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Barcelona, Spain

[3]Universitat Pompeu Fabra (UPF), Barcelona, Spain

***Corresponding authors:** E-mail: tgabaldon@crg.es; adura@ukm.my.

Associate editor: Claudia Russo

Abstract

Transferrin is a protein super-family involved in iron transport, a central process in cellular homeostasis. Throughout the evolution of vertebrates, transferrin members have diversified into distinct subfamilies including serotransferrin, ovotransferrin, lactoferrin, melanotransferrin, the inhibitor of carbonic anhydrase, pacifastin, and the major yolk protein in sea urchin. Previous phylogenetic analyses have established the branching order of the diverse transferrin subfamilies but were mostly focused on the transferrin repertoire present in mammals. Here, we conduct a comprehensive phylogenetic analysis of transferrin protein sequences in sequenced vertebrates, placing a special focus on the less-studied nonmammalian vertebrates. Our analyses uncover a novel transferrin clade present across fish, sauropsid, and amphibian genomes but strikingly absent from mammals. Our reconstructed scenario implies that this novel class emerged through a duplication event at the vertebrate ancestor, and that it was subsequently lost in the lineage leading to mammals. We detect footprints of accelerated evolution following the duplication event, which suggest positive selection and early functional divergence of this novel clade. Interestingly, the loss of this novel class of transferrin in mammals coincided with the divergence by duplication of lactoferrin and serotransferrin in this lineage. Altogether, our results provide novel insights on the evolution of iron-binding proteins in the various vertebrate groups.

Key words: **transferrin, *Lates calcarifer*, phylogenetics, positive selection, functional divergence.**

Article

Introduction

Iron plays an essential role in many biochemical processes including electron transfer reactions, gene regulation, oxygen binding, and transport, as well as regulation of cell growth and differentiation (Beard 2001). Iron is highly toxic in its free ferrous ionic form (Fe^{2+}), which can lead to the formation of hydroxyl radicals; a major source of oxidative damage to proteins, nucleic acids, and lipids. Hence, iron must be transported as free ferric ion (Fe^{3+}). However, this form is insoluble and thus cells must store and transport ferric ion in a bound state by using specialized iron-binding proteins (Cheng et al. 2004). In addition to their potential toxic effects, iron levels must be regulated tightly to avoid pathogens taking advantage of this resource. In metazoans, efficient transport of iron is mediated by its tight but reversible binding to members of the transferrin family (Aisen and Leibman 1972).

Transferrin carries two specific high affinity Fe^{3+} binding sites arranged in two homologous lobes named N-lobe and C-lobe. Both N-lobe and C-lobe contain two subdomains each named N1 and N2 subdomains, and C1 and C2 subdomains, respectively (Gomme et al. 2005). Each pair of subdomains is connected by a flexible hinge (Cheng et al. 2004). In each lobe, these two subdomains interact to form a deep cleft capable of binding a single iron ion and a synergistic anion, which is octahedrally coordinated by four residues: an aspartic acid, two tyrosines, and a histidine, coupled to four synergistic anion-binding residues namely a threonine, an arginine, an alanine, and a glycine (MacGillivray et al. 1998; Lambert, Perri, Halbrooks, et al. 2005). Transferrin genes are generally structured in 17 exons, which are similar in size across most taxa (Schaeffer et al. 1987; Cunningham et al. 1992). In contrast to exons, introns can vary greatly in size as shown in the teleost fish *Oryzias latipes* (Mikawa et al. 1996).

Members of the transferrin family have been found in both vertebrates and invertebrates, and have been classified based on their amino acid sequences, putative functions, and locations. Transferrin subfamilies previously reported include serotransferrin, ovotransferrin, lactoferrin, and melanotransferrin. Other members of less widespread distribution include an inhibitor of carbonic anhydrase, the monolobal otolith matrix protein-1, the major yolk protein, toposome in sea urchins, the crayfish protein, and pacifastin (Lambert, Perri, Halbrooks, et al. 2005; Lambert, Perri, Meehan 2005; Lambert 2012). Earlier evolutionary studies have suggested that the current bilobal structure of transferrin has arisen by duplication of a monolobal form (Greene and Feeney 1968). In addition, melanotransferrin is thought to be the earliest diverging transferrin in vertebrates, originating right after the divergence of vertebrates and other chordates

(Nakamasu et al. 1999; Lambert, Perri, Meehan 2005). The serotransferrin and melanotransferrin split may have occurred after the lobe duplication (Lambert, Perri, Meehan 2005; Liu et al. 2010) and later duplication events may have created more recent transferrin groups, such as lactoferrin and ovotransferrin (Lambert, Perri, Meehan 2005).

To date, a large number of transferrin protein sequences in mammals and birds have been described (Ciuraszkiewicz et al. 2007). Serotransferrin encoded in mammals, and ovotransferrin encoded in birds have been reported in many studies (Williams 1968; Saleh et al. 2003). Lactoferrin and melanotransferrin in mammals have also been identified (Nakamasu et al. 1999; MacGillivray and Mason, 2002). With respect to fish, a single serotransferrin gene has been described (Lambert, Perri, Halbrooks, et al. 2005), but the recent availability of genome sequences has rapidly increased the number of putative transferrin fish sequences predicted by homology in public databases. This includes the otolith matrix protein-1 in *Danio rerio* (Murayama et al. 2005) and several melanotransferrin-like genes in teleost fish (Lambert, Perri, Halbrooks, et al. 2005). All such sequences were identified in the sequenced fish genomes included in Ensembl database, namely *D. rerio* (zebrafish), *Tetraodon nigroviridis* (green spotted pufferfish), *O. latipes* (medaka), *Takifugu rubripes* (fugu), *Gasterosteus aculeatus* (stickleback), and the newest, *Gadus morhua* (Atlantic cod) (Flicek et al. 2012). Expressed sequence tags (ESTs) data are also available in fish species including, among others, *Hippoglossus stenolepis* (Pacific halibut) (Douglas et al. 2007), *Dicentrarchus labrax* (European seabass) (Sarropoulou et al. 2009), *Ictalurus punctatus* (channel catfish) (Wang et al. 2010), *Paralichthys olivaceus* (Japanese flounder) (Aoki et al. 2000), and *Lates calcarifer* (Asian seabass) (Tan et al. 2008). Despite the fact that transferrin proteins have been isolated and are well characterized in higher vertebrates, its functions and its evolutionary relationships across fish species remain poorly understood.

Several studies have shown that transferrin contributes to the innate immune system in fishes. For instance, high concentrations of free iron ion have been shown to promote the growth of pathogens (Teehan et al. 2004). Thus, the ability of transferrin to bind iron and to control the propagation of invading pathogens enables this protein to act as an antimicrobial agent, and to play a frontier role in innate immune system mechanism in fish (Farnaud and Evans 2003; Ong et al. 2006). Consistent with this role, Stafford et al. (2001) found that the cleaved goldfish transferrin elicited nitric oxide response in macrophages. Transferrin was also found as one of the immune components in the skin mucus of the olive flounder (Palaksha et al. 2008). Moreover, the upregulation of transferrin expression has been observed in infected channel catfish (Peatman et al. 2008; Liu et al. 2010) and vaccinated Atlantic cod (Caipang et al. 2008).

In addition, previous evolutionary analyses have suggested the action of positive selection in different transferrin proteins, which was interpreted as important for facilitating adaptation to the environment and pathogenic infections (Ellis 2001; Lambert, Perri, Meehan 2005; Andersen et al. 2011). These include the identification of footprints of positive selection in mammalian lactoferrin, which was hypothesized to result from a genetic arm race (Liang and Jiang 2010). In fish, positive selection has been suggested to occur in transferrin members of salmonid species (Ford et al. 1999; Ford 2001). More recently, a study by Andersen et al. (2011) showed a highly divergent allele frequency among *Gad. morhua* populations, which suggest the action of positive selection. Positive selection in fish transferrin has been suggested to be driven by its competition for iron with proteins of pathogenic bacteria (Ford et al. 1999). Consistent with the fact that the proteins involved in defence mechanisms have a high probability to undergo positive selection, lactoferrin in mammals and fish transferrin genes were expected to function specifically in defence mechanisms (Ford 2001; Yang 2006; Liang and Jiang 2010; Andersen et al. 2011).

In one of our previous studies, the screening of EST data from *L. calcarifer*—one of the most important aquaculture species in the Asia Pacific region—led us to the identification of a transferrin gene (Choong HC, unpublished MA thesis). Our initial attempts to classify this protein into one of the existing transferrin subfamilies were challenged by a mammalian-centred classification system, which failed to account for the observed diversity of fish transferrins. We thus undertook a comprehensive comparative analysis of vertebrate transferrin genes, which unearthed the existence of a novel clade of transferrin absent from mammals, and an intricate story of duplications, losses, and sweeps of accelerated evolution. Our analyses include assessment of domain sequences and gene structures, phylogenetic inference, and detection of positive selection and functional divergence. Altogether, our results establish a novel lineage in the evolution of transferrin, which evolved by duplication, followed by functional divergence and, subsequently, gene loss in the mammalian lineage. Although only experimental characterization can provide compelling evidence of functional diversification, our analyses uncover patterns compatible with important functional shifts occurred at the origin of this novel class of transferrins.

Results

Analysis of Transferrin Sequences

The complete transferrin protein in *L. calcarifer* was deduced from the complementary DNA (cDNA) sequence, based on a 2,398 bases (690 amino acids) long open reading frame (ORF), as predicted by ORF finder. Searches in protein domain databases (Prosite, Simple Modular Architecture Research Tools [SMART], and Pfam) revealed the presence of two transferrin domains (TR_FER) located in a region spanning residue 24 to residue 680. From Prosite database, five motifs with three distinct patterns that specifically located in the domain regions, namely transferrin-like domain signature-1 (PS00205), transferrin-like domain signature-2 (PS00206), and transferrin-like domain signature-3 (PS00207) were identified. Six disulfide bridges, four iron-binding residues (Asp, Tyr, Tyr, and His), and four anion-binding residues (Thr, Lys, Ala, and Gly) were found in each lobe of this protein sequence.

Using a PSI-BLAST analysis, we retrieved a total of 35 vertebrate homologous protein sequences from the Swiss-Prot database Release 2011_12 (Uniprot Consortium 2012). Of these, we selected 23 sequences for subsequent analyses, after discarding truncated sequences, or those that were isoforms of a protein already present in the data set. In addition, a total of 89 transferrin homologous proteins were found by searching against 13 fully sequenced genomes (*O. latipes, Tak. rubripes, T. nigroviridis, G. aculeatus, D. rerio, Gad. morhua, Ciona intestinalis, Xenopus tropicalis, Anolis carolinensis, Gallus gallus, Ornithorhynchus anatinus, Mus musculus*, and *Homo sapiens*) downloaded from Ensembl database. A final data set containing 53 transferrin homologs was built by merging the Ensembl and Swiss-Prot data sets, and removing redundant proteins and isoforms (supplementary table S2, Supplementary Material online).

Our computational prediction of signal peptide cleavage indicates that all mammal transferrins have a predicted signal peptide, which is in line with transferrin's function in iron transportation (supplementary table S2, Supplementary Material online). This contrast with the situation in fishes for which some transferrins scattered across all transferrin subfamilies did not show a predicted signal peptide. Most of the identified signal peptide cleavage sites occur between residue 19 and residue 20. Similar to previous studies, lactoferrin is distinguished by its high pI = ~8–9 (Pakdaman et al. 1998; McGillivray and Mason 2002) when compared with other transferrins (supplementary table S2, Supplementary Material online). Interestingly, this was also the case for some of the novel transferrins included in this analysis. Additionally, all transferrin proteins studied showed potential phosphorylation sites on Ser, Thr, and Tyr residues (supplementary table S2, Supplementary Material online).

Phylogenetic Analyses

Phylogenetic analyses based on Bayesian and maximum likelihood (ML) methods largely resolved the members of transferrin sequences into known clades (fig. 1). With the exception of the node separating the fish serotransferrin clade (IIId) with clade containing amphibian serotransferrin and saxiphilin (IIIc) in ML phylogenetic tree, which has bootstrap value (52%), all the main clades received a high support with >95% posterior probability in Bayesian and >74% bootstrap value in ML phylogenetic tree, respectively. However, several internal branches in the ML analysis received a rather weak bootstrap support.

Our reconstructed ML and Bayesian phylogenetic trees consistently reveal the presence of four major transferrin clades, which diverged at the vertebrate ancestor. This is consistent with the two rounds of whole genome duplication (2R) hypothesis for the vertebrate ancestor (Dehal and Boore 2005). Among these four groups, two include members of all major vertebrate groups, namely the melanotransferrin clade (II) and a clade including, among other types, serotransferrins, lactoferrins, and ovotransferrins (clade III). These results are compatible with an ancient origin of melanotransferrin (Nakamasu et al. 1999; Lambert, Perri, Meehan 2005), but suggest that the ancestral vertebrate presented four major transferrin clades. In addition, our topology indicates that mammalian lactoferrin, serotransferrin, and inhibitors of carbonic anhydrase evolved through more recent duplications within mammals. Thus, all these transferrins should be considered co-orthologous to serotransferrin in fish and amphibian, saxiphilin in amphibian, ovotransferrin-like in nonavian sauropsids, and ovotransferrins in birds, because they all diverged through speciation events (Gabaldón 2008). Interestingly, the otolith proteins (clade I) seem an ancient transferrin class, lost in other lineages but fish, although further analyses are needed to clarify the correct phylogenetic placement of this clade, which represents the only type of monolobal ferritin present in vertebrates. Finally, we detected a fourth clade emerging from a duplication at the vertebrate ancestor, which has been so far undescribed but which was highly supported by bootstrap and posterior probability values in both of our reconstructed ML and Bayesian trees. This clade (IV) is present across amphibians, sauropsids, and fishes, but is strikingly absent from mammals. To confirm this result, we performed exhaustive searches in Ensembl and NCBI sequence databases using several members of this novel group as a query, but we were unable to find any mammalian protein clustering within this clade. Notably, clade IV presented a relatively long branch as compared with the basal branches of the two other clades, which suggested the presence of accelerated evolution, possibly driven by positive selection and neo- or subfunctionalization. To investigate this further, we inspected the alignments and performed specific tests to assess the presence of positive selection and functional divergence.

Conservation of Key Functional Residues

The multiple sequence alignment indicates the presence of 12 conserved disulfide bridges, of which 6 are located in each lobe. These disulfide bridges were linked by two Cys located at Cys_{10}–Cys_{53}, Cys_{20}–Cys_{44}, Cys_{270}–Cys_{366}, Cys_{316}–Cys_{333}, Cys_{330}–Cys_{347}, and Cys_{408}–Cys_{422} in the first lobe; and Cys_{558}–Cys_{595}, Cys_{568}–Cys_{586}, Cys_{687}–Cys_{791}, Cys_{734}–Cys_{757}, Cys_{754}–Cys_{772}, and Cys_{840}–Cys_{855}. Residues are numbered according to their position in the alignment and using *L. calcarifer* sequence as a reference. With the exception of otolith matrix protein-1 (*D. rerio1*, *Gad. morhua1*, *G. aculeatus1*, *Tak. rubripes1*, *O. latipes1*, *T. nigroviridis1*, and *Oncorhynchus mykiss*), all nonavian sauropsid transferrins (*A. carolinensis1*, *2*, and *3*), which have gaps insertion in their N-lobe, and *Gal. gallus4*, *X. tropicalis2*, *Rana catasbeiana*, and *Tak. rubripes3* that have some Cys substitutions to other residues, all Cys responsible for forming the disulfide bridges were found to be completely conserved.

With respect to the conservation of key residues involved in iron and anion binding, the newly identified group of transferrin sequences showed the largest number of substitutions. For instance, this clade shows 52 out of 101 observed substitutions (51%) of residues involved in iron binding (table 1). Similarly, 40 out of 99 (40%) substitutions observed for the anion-binding sites occur in this novel clade (table 2). As a

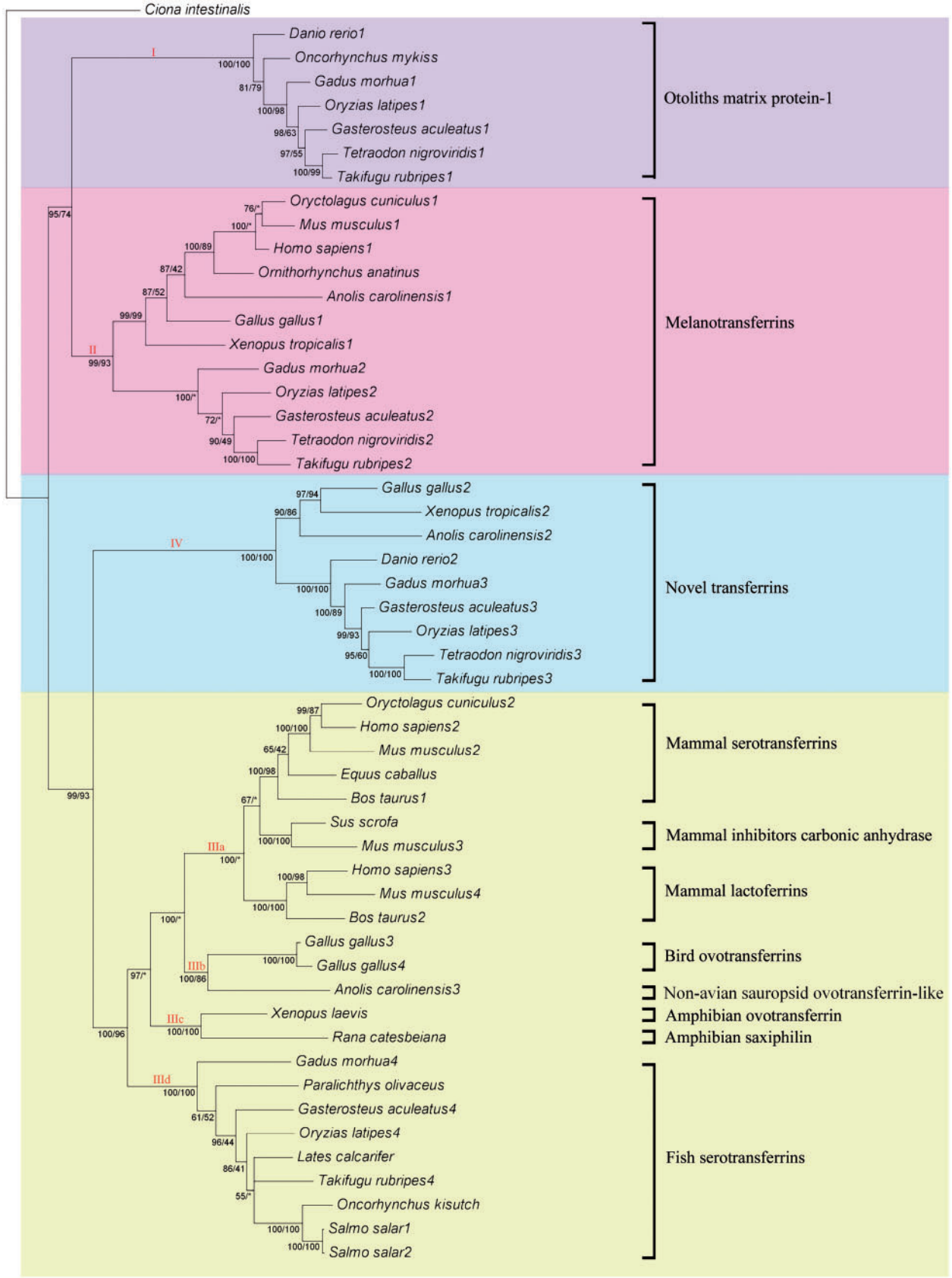

FIG. 1. Phylogenetic tree of vertebrate transferrins based on Bayesian and ML analyses. Major groups of vertebrate transferrins are resolved, including the newly identified clade (clade IV). Clade III is divided into several subfamilies, some of which emerged from more recent duplications. Numbers on the branches indicate Bayesian posterior probabilities followed by bootstrap support in the ML analyses. If the partition does not appear in the ML tree, this is indicated with an asterisk. The final log likelihood of the ML tree was −40,597.96204 and the gamma shape parameter 1.336.

Table 1. Iron-Binding Site Observation for Transferrin Protein among Taxa.

Members	Organisms	Active Sites (Iron-Binding Residues)							
		N-Lobe				C-Lobe			
		D_{64}	Y_{244}	Y_{356}	H_{430}	D_{610}	Y_{659}	Y_{783}	H_{863}
Outgroup	*Ciona intestinalis*	/	/	/	/	/	/	/	/
Otolith matrix protein-1	*Danio rerio1*	—	—	—	—	Y	/	E	N
	Oncorhynchus mykiss	—	—	—	—	F	/	E	N
	Gadus morhua1	—	—	—	—	F	/	E	P
	Oryzias latipes1	—	—	—	—	F	/	E	N
	Gasterosteus aculeatus1	—	—	—	—	F	/	E	N
	Tetraodon nigroviridis1	—	—	—	—	S	/	E	N
	Takifugu rubripes1	—	—	—	—	P	/	E	N
Melanotransferrin	*Oryctolagus cuniculus1*	/	/	/	/	N	/	D	/
	Homo sapiens1	/	/	/	/	S	/	/	/
	Mus musculus1	/	/	/	/	R	/	/	/
	Gallus gallus1	/	/	/	R	G	/	/	Q
	Ornithorhynchus anatinus	/	/	/	/	G	/	/	Q
	Anolis carolinensis1	—	—	—	/	S	/	/	R
	Xenopus tropicalis1	/	/	/	/	/	/	/	/
	Gad. morhua2	/	/	/	R	/	/	/	/
	O. latipes2	/	/	/	R	/	/	/	/
	G. aculeatus2	/	/	/	R	/	/	/	/
	T. nigroviridis2	/	/	/	R	/	/	/	/
	Tak. rubripes2	/	/	/	R	/	/	/	/
Novel transferrin	*D. rerio2*	/	I	F	G	/	V	N	N
	O. latipes3	/	V	F	G	/	L	N	N
	G. aculeatus3	/	I	F	G	/	V	N	N
	Tak. rubripes3	/	H	F	G	/	V	N	N
	T. nigroviridis3	/	L	F	A	/	V	N	N
	Gad. morhua3	/	I	/	G	/	V	N	N
	X. tropicalis2	/	V	F	K	/	L	N	N
	A. carolinensis2	—	—	F	R	/	T	N	G
	Gal. gallus2	/	V	F	R	/	V	N	S
Serotransferrin (Fish)	*Paralichthys olivaceus*	/	/	/	/	/	/	/	/
	Gad. morhua3	/	/	/	/	/	/	/	/
	G. aculeatus4	/	/	/	/	/	/	/	/
	O. latipes4	/	/	/	/	/	/	/	/
	Lates calcarifer	/	/	/	/	/	/	/	/
	Tak. rubripes4	/	/	/	/	/	/	/	/
	O. kisutch	/	/	/	/	/	/	/	/
	Salmo salar1	/	/	/	/	/	/	/	/
	S. salar2	/	/	/	/	/	/	/	/
Serotransferrin (Amphibian)	*X. laevis*	/	/	/	/	/	/	/	/
Saxiphilin (Amphibians)	*Rana catesbeiana*	/	H	/	G	E	L	N	P
Ovotransferrin-like (nonavian sauropsids)	*A. carolinensis3*	—	/	/	/	/	/	/	/
Ovotransferrin (Birds)	*Gal. gallus3*	/	/	/	/	/	/	/	/
	Gal. gallus4	/	/	A	/	/	/	/	/
Lactoferrin (Mammals)	*Bos taurus2*	/	/	/	/	/	/	/	/
	M. musculus4	/	/	/	/	/	/	/	/
	H. sapiens3	/	/	/	/	/	/	/	/
ICA (Mammals)	*M. musculus3*	/	/	/	/	/	/	S	R
	Sus scrofa	/	/	/	/	/	/	F	/
Serotransferrin (Mammals)	*Equus caballus*	/	/	/	/	/	/	/	/
	B. taurus1	/	/	/	/	/	/	/	/
	M. musculus2	/	/	F	/	/	/	/	/
	Ory. cuniculus2	/	/	/	/	/	/	/	/
	H. sapiens2	/	/	/	/	/	/	/	/

NOTE.—/ and — symbols indicate "as in reference" and "missing," respectively.

Table 2. Anion-Binding Site Observation for Transferrin Protein among Taxa.

Members	Organisms	Active Sites (Anion-Binding Residues)							
		N-Lobe				C-Lobe			
		T_{272}	R_{276}	A_{278}	G_{279}	T_{689}	R_{693}	A_{695}	G_{696}
Outgroup	*Ciona intestinalis*	/	K	/	/	/	/	/	/
Otolith matrix protein-1	*Danio rerio1*	—	—	—	—	P	T	V	/
	Oncorhynchus mykiss	—	—	—	—	P	T	V	/
	Gadus morhua1	—	—	—	—	P	T	V	/
	Oryzias latipes1	—	—	—	—	P	T	V	/
	Gasterosteus aculeatus1	—	—	—	—	P	T	V	/
	Tetraodon nigroviridis1	—	—	—	—	P	T	V	/
	Takifugu rubripes1	—	—	—	—	P	T	V	/
Melanotransferrin	*Oryctolagus cuniculus1*	/	/	V	/	P	S	/	/
	Homo sapiens1	/	/	V	/	A	S	/	/
	Mus musculus1	/	/	V	/	P	S	/	/
	Gallus gallus1	/	/	/	/	/	/	/	/
	Ornithorhynchus anatinus	/	/	V	/	S	/	S	/
	Anolis carolinensis1	—	—	P	S	/	/	/	/
	Xenopus tropicalis1	/	/	/	/	/	/	/	/
	O. latipes2	/	/	/	/	/	/	/	/
	G. aculeatus2	/	/	/	/	/	/	/	/
	T. nigroviridis2	/	/	/	/	/	/	/	/
	Tak. rubripes2	/	/	/	/	/	/	/	/
	Gad. morhua2	/	/	V	/	/	/	/	/
Novel transferrin	*D. rerio2*	S	W	/	/	S	S	/	/
	O. latipes3	S	W	/	/	S	S	/	/
	G. aculeatus3	S	W	/	/	G	S	/	/
	Tak. rubripes3	S	W	/	/	S	S	/	/
	Gad. morhua3	S	W	/	/	G	S	/	/
	T. nigroviridis3	S	W	/	/	S	S	/	/
	X. tropicalis2	N	W	S	/	G	S	/	/
	A. carolinensis2	S	W	S	/	G	S	/	/
	Gal. gallus2	N	W	S	/	S	S	G	/
Serotransferrin (fish)	*Paralichthys olivaceus*	/	K	/	/	/	/	/	/
	Gad. morhua4	/	K	/	/	/	/	/	/
	G. aculeatus4	/	K	/	/	/	/	/	/
	O. latipes4	/	K	/	/	/	/	/	/
	Lates calcarifer	/	K	/	/	/	/	/	/
	Tak. rubripes4	/	K	/	/	/	/	/	/
	O. kisutch	/	K	/	/	/	/	/	/
	Salmo salar1	/	K	/	/	/	/	/	/
	S. salar2	/	K	/	/	/	/	/	/
Serotransferrin (amphibian)	*X. laevis*	/	K	/	/	/	/	/	/
Saxiphilin (amphibian)	*Rana catesbeiana*	/	K	D	/	/	D	/	/
Ovotransferrin-like (nonavian sauropsids)	*A. carolinensis3*	/	R	/	/	/	/	/	/
Ovotransferrin (birds)	*Gal. gallus3*	/	/	/	/	/	/	/	/
	Gal. gallus4	/	/	/	/	/	/	/	/
Lactoferrin (mammals)	*Bos taurus2*	/	/	/	/	/	/	/	/
	M. musculus4	/	/	/	/	/	/	/	/
	H. sapiens3	/	/	/	/	/	/	/	/
ICA (mammals)	*M. musculus3*	/	W	/	/	/	T	E	/
	Sus scrofa	/	W	/	/	I	T	/	/
Serotransferrin (mammals)	*Equus caballus*	/	/	/	/	/	/	/	/
	B. taurus1	/	/	/	/	/	/	/	/
	M. musculus2	/	/	/	/	/	S	/	/
	Ory. cuniculus2	/	/	/	/	/	S	/	/
	H. sapiens2	/	/	/	/	/	S	/	/

NOTE.— / and — symbols indicate "as in reference" and "missing," respectively.

result, both N- and C-lobe in the novel group lack all the functional residues for iron binding except for the conserved Asp. For the anion-binding sites, both lobes of the novel transferrin group have accumulated a similar number of substitutions, with two or three key residues being altered in all fish, amphibian, and sauropsid sequences. This lack of conservation contrasts with that of most other transferrin groups, especially serotransferrin, which was the group with a larger fraction of conserved key residues. An intermediate degree of substitution at key residues was observed in melanotransferrin and saxiphilin. These observations may suggest that the novel group may not be able to bind iron or that its function may be different in this respect from the rest of the known transferrins. None of the proteins belonging to this clade has been experimentally characterized so far. However, although only a functional characterization of some of the members of the newly identified family will resolve whether they perform a clearly distinct function, sequence analysis may serve to detect potential footprints of functional variation.

Detection of Positive Selection in Novel Transferrin Lineages

Proteins undergoing distinct selective pressures can be identified by observing changes in their evolutionary rates (Liang and Jiang 2010). The long branch subtending the newly identified lineage suggests that the genes in this cluster have experienced a faster evolutionary rate as compared with the other clades, which can be the result of positive selection (Tsantes and Steiper 2009). To assess this possibility, we used a branch-specific model for sequence evolution and compared the likelihood of two scenarios: one implying neutral evolution and similar rates in all branches (M0), and one involving positive selection in the branch subtending the novel group (M2). The latter model was highly supported ($P \ll 0.01$, likelihood ratio test [LRT]), which is consistent with an accelerated rate of substitutions in the lineage leading to the novel group (fig. 2).

Functional Divergence of Novel Transferrin and Fish Serotransferrin Genes

The observed accelerated rate of substitutions in the lineage leading to the novel group suggests the possibility of positive selection for a new function. Unfortunately, none of the members of the novel group has been functionally characterized so far, which prevents us from directly analyzing functional shifts. We can, nevertheless, investigate the possibility of functional shifts by assessing the type of amino acid changes occurred within this lineage. For this, we assessed the existence of radical shifts of rates or functional properties of residues using the program DIVERGE2 (see Materials and Methods). Our results indicate a significant amount of functionally relevant shifts between the novel transferrin clade and its sister-group, serotransferrins. This was true for the two types of functional divergence contemplated in the analyses. Type I functional divergence showed a $\theta I = 0.528 \pm 0.041167$ between the two groups and was significantly greater than 0 ($P < 0.01$, Fisher transformation test), indicating specific rate shifts in following the duplication of these two subfamilies. By using a cutoff value of $P > 0.6$, 159 columns in the alignment were observed to have very conserved

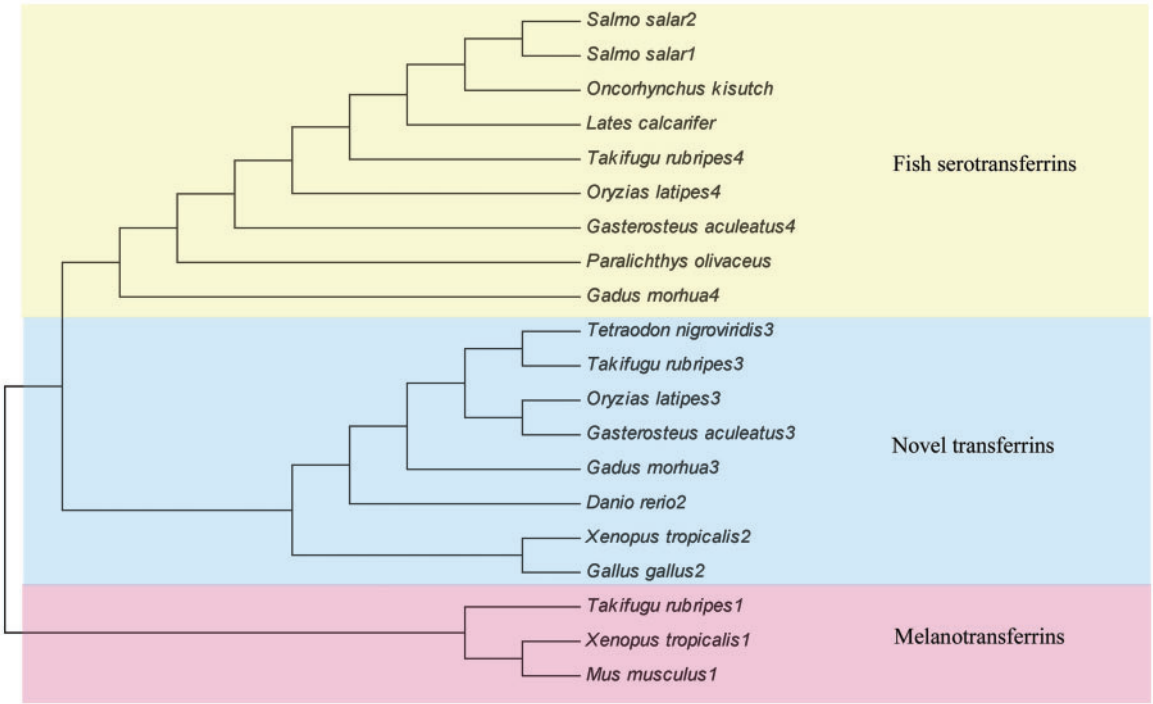

FIG. 2. Topology for a subset of transferrin proteins used for the branch-specific model. Branches subtending the newly identified group and serotransferrins were assumed to have different rate from a background. The branch of novel transferrin indicates $\omega > 0$ shows that the genes have undergone position selection. # indicates the branches that have different rate of substitution ($dN/dS = \omega$).

residues in one transferrin gene cluster whereas the residues in the other cluster were varied, suggesting that those sites were responsible in evolutionary rate's shift between both clusters. Similarly, type II functional divergence was highly significant with a value of θII = 0.195380 ± 0.067770, showing that some amino acid residues in both novel transferrin and fish serotransferrin clusters may have radical change in amino acid properties. Thirteen residues passed a stringent threshold

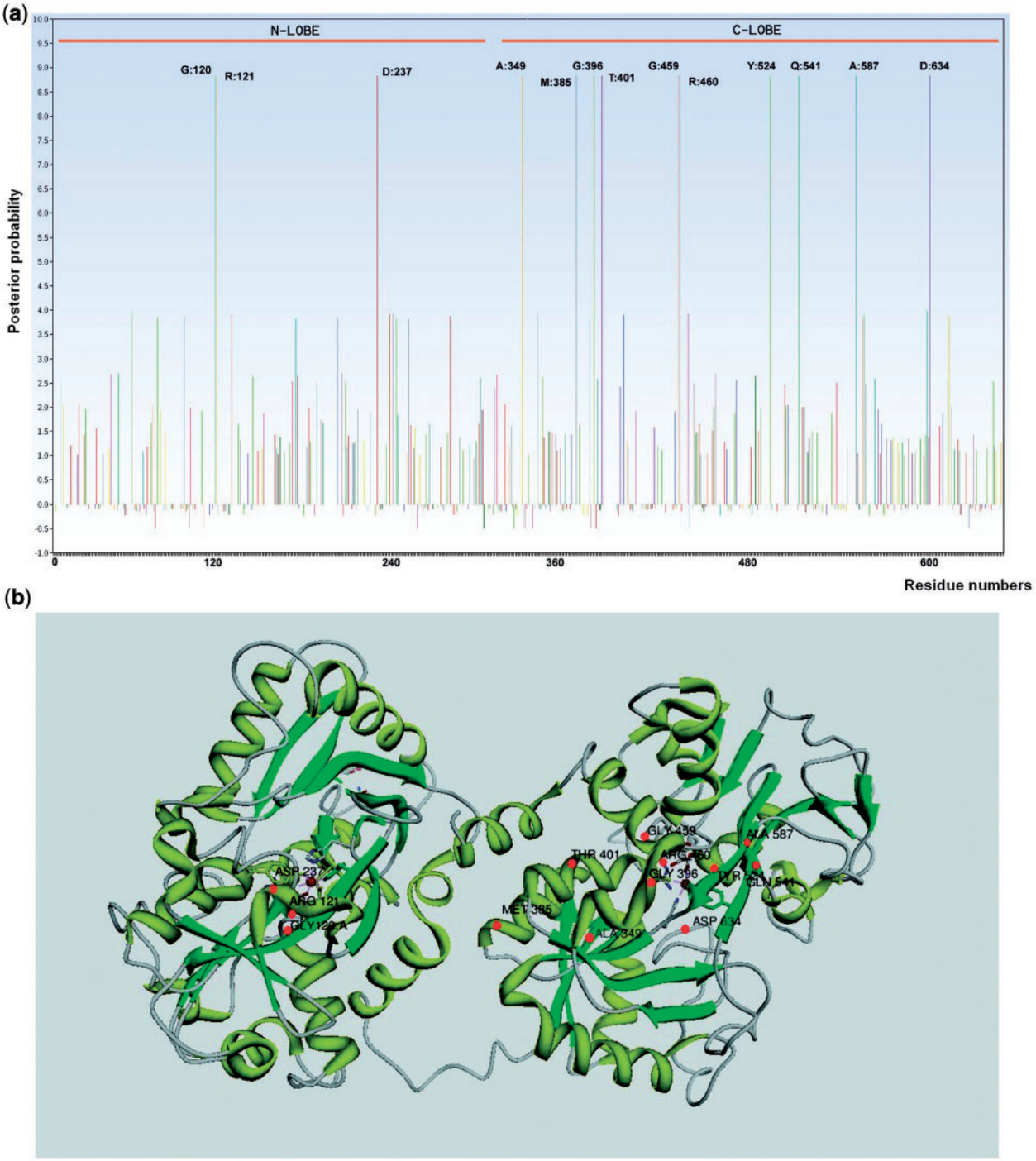

Fig. 3. Type 2 functional divergence. (*a*) Thirteen amino acids residues with posterior ratio more than 8 have been observed from type 2 functional divergence. The observation of amino acids location that have high posterior ratio ($R > 8$) shows that each cluster has a conserved radical change between each other. (*b*) The 3D protein for transferrin protein in *Lates calcarifer* was modeled to show the location of residues affected by radical changes. Five of the residues that have high posterior ratio found to form loop, five forming α-helix, whereas another three forming the β-strand. Tyr_{524} is the functional residue for iron binding, which has been substituted to Asp in novel sequences. Arg_{121} and Arg_{460} involve in anion binding, which has been substituted to Trp and Ser in the novel sequences, respectively. The substitutions of important residues might suggest that those sites may responsible to perform a different function after the duplication event. The locations of the residues are calculated based on the *Gallus gallus* ovotransferrin sequence (2D3I) obtained from PDB database (last accessed November 1, 2012).

of a posterior ratio higher than eight ($R > 8$). These residues are conserved within each cluster, but are different between the two clusters, having specific amino acids that show radically different functional properties.

To gain a more detailed insight in the possible functional changes triggered by these radical changes, we mapped these residues onto the 3D structure of a transferrin protein. Five amino acid residues were found to be involved in the formation of α-helix structure, three in β-strand, and the other five residues form the loop. Interestingly, all the residues identified in site property shift were located in the transferrin domain regions. Tyr_{524} is the functional residue for iron binding, which has been substituted to Asp in novel sequences. Arg_{121} and Arg_{460} involve in anion binding, which have been substituted to Trp and Ser in the novel sequences, respectively (The locations of the residues are calculated based on *Gal. gallus* ovotransferrin sequence [2D3I] taken from PDB database) (fig. 3). Altogether, our analyses of positive selection and sequence-based functional shifts suggest that these two clusters may have differed functionally after the duplication event.

Discussion

The functions of characterized members of the transferrin family such as serotransferrin, lactotransferrin, and ovotransferrin are clearly related to iron binding and transport (Octave et al. 1983; Anderson et al. 1987; Lambert, Perri, Halbrooks, et al. 2005). Serotransferrin specifically transports iron in physiological fluid and delivers it to the cells via receptor-mediated endocytosis (Octave et al. 1983), whereas lactoferrin and ovotransferrin serve an antimicrobial function to prevent the proliferation of invading microbes in mammals and oviparous organisms such as sauropsids, respectively (Graham and Williams 1975; Anderson et al. 1987). The function of melanotransferrin, in contrast, remains mostly unknown, although a possible involvement in the formation of cartilage nodules has been reported (Nakamasu et al. 1999). In contrast to the other transferrin proteins that have two iron-binding sites, melanotransferrin seems to contain a single one, and it is as yet unclear whether it transports iron (Sekyere and Richardson 2000). Otolith matrix protein-1, which has only been found in fish so far, represents the only monolobal transferrin in vertebrate (Andersen et al. 2011).

Our results have uncovered yet another class of transferrin, which is absent from mammals probably due to a single specific gene loss event at the base of this lineage. As there is no experimental data for any of the members of this novel clade, one can only speculate on its possible functional role. Our results suggest an extremely high level of sequence variation, which mostly occurred at the base of the clade, directly following the duplication event. This is compatible with a high degree of functional variation, which considering the high level of variation at key residues, may even involve the loss of the ability to bind iron. Our results also clarify the evolution within the other clades of transferrin genes and identify two duplications in mammals giving rise to the functionally divergent groups of mammalian lactoferrins, serotransferrins, and the inhibitor of carbonic anhydrase, and thus these duplications in mammals may have been concomitant with the loss of the newly identified clade. Unfortunately, our search for gene expression data in databases did not provide any significant information for the novel group, and therefore it remains unclear whether the novel group plays a tissue-specific role. Finally, an analysis of chromosomal regions around the genes coding for the transferrins in the novel group showed no conservation of synteny with other transferrin groups.

Altogether, our results support a burst of accelerated evolution directly following the duplication event and specifically affecting the lineage leading to the newly identified transferrin group. The substitutions that accumulated along this lineage involve some radical residue changes that have likely affected the overall function of these molecules. In particular, the substitution of the second Tyr to Phe in the N-lobe is likely to prevent iron binding (He et al. 1997), raising the possibility that the novel group does not possess this otherwise conserved function of transferrins. This is consistent with another observation by Lambert, Perri, Halbrooks, et al. (2005), in which a substitution of His to Arg in the N-lobe has been hypothesized to prevent iron binding. In addition, we identified an insertion of {[SNTE]-G-[KR]-S-[YF]} in the second motif (PS00206) of C-lobe. The analysis of pattern in Prosite database indicates that {[SNTE]GK} motif is involved in protein kinase C phosphorylation, suggesting a possible regulation by posttranslational modification in the novel group of proteins.

Altogether, our observations suggest that the novel group of transferrins may have a radically different function as compared with the other groups of transferrins. Characterized groups of transferrins already show that this family of proteins can adapt in many ways to fit the unique physiological demands in different organisms (Yang and Gui 2004). Considering the divergence times involved and the burst of sequenced divergence following the duplication that originated this newly identified group, we propose that this novel group of transferrins may perform a very distinct function, which will add to the already known functions of characterized transferrins. Clearly, experimental characterization of proteins from the newly identified group of transferrins will be needed to confirm our in silico predictions. Nevertheless, our study illustrates how the increasing amount of sequences deposited in databases, coupled with appropriate evolutionary analysis can guide the discovery of putative novel types of proteins.

Materials and Methods

Transferrin Sequence Analysis

The complete transferrin cDNA sequence was obtained from the analysis of ESTs in liver tissue of *L. calcarifer* (Choong HC, unpublished MA thesis). This cDNA sequence (Genbank Accession: JQ247193) was then translated into its potential ORF using ORF Finder (Rombel et al. 2002). PSI-BLAST was used to search for homologs of transferrin in Swiss-Prot database Release 56. Additional homologs were retrieved by searching the Pfam transferrin domain with *hmmsearch* program in HMMER 3 (Finn et al. 2011) in completely sequenced

genomes from the Ensembl database Release 65 (Flicek et al. 2012). These include six fish complete genomes namely—*D. rerio*, *O. latipes*, *T. nigroviridis*, *G. aculeatus*, *Tak. rubripes*, and *Gad. morhua*—and seven vertebrate genomes namely—*H. sapiens* (human), *M. musculus* (mouse), *Gal. gallus* (chicken), *Orn. anatinus* (platypus), *A. carolinensis* (anole lizard), *X. tropicalis* (western clawed frog), and *C. intestinalis* (tunicate). Identified sequences were analyzed using various tools. Domain analyses were carried out using several resources, including SMART (Schultz et al. 1998), Pfam 20.0 (Bateman et al. 2004), and Prosite 19.36 (Sigrist et al. 2010). In addition, SignalP 4.0 (Peterson et al. 2004) was used to predict the location of signal peptide cleavage for those transferrin homologs obtained. The isoelectric point (pI) of transferrin homologs was identified using compute pI/Mw tool (Bjellqviest et al. 1994). Finally, NetPhos 2.0 (Blom et al. 1999) was used to predict the phosphorylation sites, which may occur in transferrin proteins.

Phylogenetic Analyses

Sequences were aligned using M-coffee (Moretti et al. 2007), which allows to compute the alignments using several packages including T-coffee (Notredame et al. 2000), ClustalW (Larkin et al. 2007), MAFFT (Katoh 2008), Dialign-tx (Subramaniam et al. 2008), POA (Lee et al. 2002), ProbCons (Do et al. 2005), MUSCLE (Edgar 2004), PCMA (Pei et al. 2003), KAlign (Lasmann et al. 2009), PRODA (Phuong et al. 2006), and PRANK (Loytynoja and Goldman 2008). T-coffee then was used to combine all these alignments into one unique final alignment. The alignment provided by T-coffee showed a score = 94, indicating a good alignment. The average length of proteins included in the alignment was 659 amino acids, ranging from 333 (*Gad. morhua1*) to 843 (*R. catesbeiana*). The signal peptide regions in N-terminal and unconserved regions in C-terminals were removed to make a total alignment length of 978 residues, in which *R. catesbeiana* has formed a large insertion from residue 93 to residue 234. The multiple sequences alignment generated in clustal format (.aln) were converted to phylip (.phy) and nexus (.nxs) formats using ClustalW. ProtTest (Darriba et al. 2011) was used to choose the best-fitting model to build the phylogenetic tree. This was, in all cases, the WAG model. Phylogenetic analyses across all vertebrates were performed using Bayesian and ML approaches, as implemented in MrBayes (Ronquist and Huelsenbeck 2003) and PhyML 3.0 (Guindon et al. 2010), respectively. ML phylogenetic tree was constructed using 1,000 bootstrap samples. The Bayesian analysis was performed using a mixed-model approach, which allows this program to vary the parameters during the search (Ronquist and Huelsenbeck 2003). Ten thousand generations were run with a sampling frequency of 10, and at least 1,000 samples were collected per analysis run. Using the burn-in parameter, 25% of the first sampled trees were discarded. Finally, MEGA4 (Tamura et al. 2007) was used for viewing and editing of the resulting phylogenetic trees.

Detection of Positive Selection in Transferrin Genes

The data used for the detection of positive selection in transferrin genes were generated by selecting 20 out of 53 sequences used in the phylogenetic analyses. This subset focuses around the duplication event leading to the formation of the novel transferrin group and includes all fish serotransferrin and novel transferrin sequences with three additional melanotransferrin sequences. Sequences were aligned using M-coffee (alignment score = 96) and were back-translated with trimAl (Capella-Gutierrez et al. 2009) to obtain a codon alignment. Subsequently, the phylogenetic tree of this subset was built using *dnaml* program as implemented in PHYLIP 3.69 (Felsenstein 2005). The codon alignment and the phylogenetic tree were then used as input data to run codeml analysis implemented in PAML 4.5 program (Yang 2007). Codeml program allows the measurement of the relative rate of nonsynonymous (d*N*) to synonymous substitution (d*S*), or $\omega = dN/dS$, using Nei–Gojobori method (Nei and Gojobori 1986). To detect genes that have likely undergone positive selection, we used the branch model implemented in codeml, which assumes that the branches of interest have a different ratio as compared to the rest of the tree. In this analysis, two models were compared, namely M0, the null model in which all branches have equal rate, and M2, a model implying positive selection or acceleration in the postduplication branch subtending the novel group. The two models were compared using a LRT.

The Detection of Functional Divergence of Two Duplicated Transferrin Genes

DIVERGE2 (Gu and Velden 2002) analysis was used to estimate the functional divergence between the novel and serotransferrin clusters. For this, we used the same data set used in the detection of positive selection in this study. In this analysis, two types of functional divergence were computed, so-called type I (Gu 1999) and type II (Gu 2001). These serve to detect, respectively, site-specific rate shifts and site-specific property shifts after the gene duplication event. In this test, a value of theta (θ) significantly larger than 0 is taken as an indication of functional divergence between the two duplicated lineages. In addition, the 3D structure of *Gal. gallus* ovotransferrin protein (2D3I) (Mizutani et al. 2005) was downloaded from the PDB database (Bernstein et al. 1977). This 3D protein structure was viewed using Chimera version 1.6 (Petterson et al. 2004) and the changes in the novel family have been mapped from the alignment to this 3D structure to clearly show the residues that might be responsible of a functional shift between these two duplicated genes.

Supplementary Material

Supplementary figure S1 and tables S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by grants from the Malaysian Ministry of Science, Technology and Innovation

(07-05-MGI-GMB009), the Spanish Ministry of Science and Innovation (BFU2009-09168), and Universiti Kebangsaan Malaysia (UKM-DLP-2011-027). The authors thank Mohd Rashid Mohd Rani for providing the transferrin nucleotide sequence of *L. calcarifer*, Salvador Capella-Gutierrez for helping with coding translation, and Dr. Khairina Tajul Arifin, Sarahani Harun, Mohd-Faizal Abu-Bakar, and Mohd-Fairuz Mohd-Yusof for technical assistance.

References

Aisen P, Leibman A. 1972. Lactoferrin and transferrin: a comparative study. *Biochim Biophys Acta.* 257:314–323.

Andersen O, De Rosa MC, Pirolli D, Tooming-Klunderud A, Petersen PE, Andre C. 2011. Polymorphism, selection and tandem duplicationof transferrin genes in Atlantic cod (*Gadus morhua*)—conserved synteny between fish monolobal and tetrapod bilobal transferrin loci. *BMC Genet.* 12:51.

Anderson BF, Baker HM, Dodson EJ, Norris GE, Rumball SV, Waters JM, Baker EN. 1987. Structure of human lactoferrin at 3.2-A resolution. *Proc Natl Acad Sci U S A.* 84:1769–1773.

Aoki T, Hirono I, Tokuda Y, Toyohara H, Yamamoto E. 2000. Identification of viral induced genes in Ig+ leucocytes of Japanese flounder *Paralichthys olivaceus,* by differential hybridisation with subtracted and un-subtracted cDNA probes. *Fish Shellfish Immunol.* 10:623–630.

Bateman A, Coin L, Durbin R, et al. (13 co-authors). 2004. The Pfam protein families database. *Nucleic Acids Res.* 32:D138–D141.

Beard JL. 2001. Iron biology in immune function, muscle metabolism and neuronal functioning. *J Nutr.* 131:568S–580S.

Bernstein FC, Koetzle TF, Williams GJ, Meyer EE Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol.* 112:535.

Bjellqvist B, Basse B, Olsen E, Celis JE. 1994. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* 15: 529–539.

Blom N, Gammeltoft S, Brunak S. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.* 294:1351–1362.

Caipang CM, Hynes N, Puangkaew J, Brinchmann MF, Kiron V. 2008. Intraperitoneal vaccination of Atlantic cod, *Gadus morhua* with heat-killed *Listonella anguillarum* enhances serum antibacterial activity and expression of immune response genes. *Fish Shellfish Immunol.* 24:314–322.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.

Cheng Y, Zak O, Aisen P, Harrison SC, Walz T. 2004. Structure of the human transferrin receptor-transferrin complex. *Cell* 116: 565–576.

Ciuraszkiewicz J, Biczycki M, Maluta A, Martin S, Watorek W, Olczak M. 2007. Reptilian transferrins: evolution of disulphide bridges and conservation of iron-binding center. *Gene* 396:28–38.

Cunningham GA, Headon DR, Conneely OM. 1992. Structural organization of the mouse lactoferrin gene. *Biochem Biophys Res Commun.* 189:1725–1731.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164–1165.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3:e314.

Do CB, Mahabhashyam MS, Brudno M, Batzoglou S. 2005. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res.* 15:330–340.

Douglas SE, Knickle LC, Kimball J, Reith ME. 2007. Comprehensive EST analysis of Atlantic halibut (*Hippoglossus hippoglossus*), a commercially relevant aquaculture species. *BMC Genomics* 8:144.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res.* 32:1792–1797.

Ellis AE. 2001. Innate host defense mechanisms of fish against viruses and bacteria. *Dev Comp Immunol.* 25:827–839.

Farnaud S, Evans RW. 2003. Lactoferrin—a multifunctional protein with antimicrobial properties. *Mol Immunol.* 40:395–405.

Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Seattle (WA): Department of Genome Sciences, University of Washington. Distributed by the author.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39:W29–W37.

Flicek P, Amode MR, Barrell D, et al. (57 co-authors). 2012. Ensembl 2012. *Nucleic Acids Res.* 4:D84–D90.

Ford MJ. 2001. Molecular evolution of transferrin evidence for positive selection in salmonids. *Mol Biol Evol.* 18:639–647.

Ford MJ, Thornton PJ, Park LK. 1999. Natural selection promotes divergence of transferrin among salmonid species. *Mol Ecol.* 8: 1055–1061.

Gabaldón T. 2008. Large-scale assignment of orthology: back to phylogenetics? *Genome Biol.* 9:235.

Graham I, Williams J. 1975. A comparison of glycopeptides from the transferrins of several species. *Biochem J.* 145:263–279.

Gomme PT, McCann KB, Bertolini J. 2005. Transferrin: structure, function and potential therapeutic actions. *Drug Discov Today* 10: 267–273.

Greene FC, Feeney RE. 1968. Physical evidence for transferrins as single polypeptide chains. *Biochemistry* 7:1366–1371.

Gu X. 1999. Statistical methods for testing functional divergence after gene duplication. *Mol Biol Evol.* 16:1664–1674.

Gu X. 2001. Maximum likelihood approach for gene family evolution under functional divergence. *Mol Biol Evol.* 18:453–464.

Gu X, Velden VK. 2002. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics* 18:500–501.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.

He QY, Mason AB, Woodworth RC, Tam BM, MacGillivray RT, Grady JK, Chasteen ND. 1997. Inequivalence of the two tyrosine ligands in the N-lobe of human serum transferrin. *Biochemistry* 36: 14853–14860.

Katoh K. 2008. Recent developments in MAFFT multiple sequence alignment program. *Brief Bioinform.* 9:286–298.

Lambert LA. 2012. Molecular evolution of the transferrin family and associated receptors. *Biochim Biophys Acta.* 1820:244–255.

Lambert LA, Perri H, Halbrooks PJ, Mason AB. 2005. Evolution of the transferrin family: conservation of residues associated with iron and anion binding. *Comp Biochem Physiol B Biochem Mol Biol.* 142:129–141.

Lambert LA, Perri H, Meehan TJ. 2005. Evolution of duplications in the transferrin family of proteins. *Comp Biochem Physiol B Biochem Mol Biol.* 140:11–25.

Larkin MA, Blackshields G, Brown NP, et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.

Lasmann T, Frings O, Erik LL. 2009. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res.* 37:858–865.

Lee C, Grasso C, Sharlow M. 2002. Multiple sequence alignment using partial order graphs. *Bioinformatics* 18:452–464.

Liang GM, Jiang XP. 2010. Positive selection drives lactoferrin evolution in mammals. *Genetica* 138:757–762.

Liu H, Takano T, Abernathy J, Wang S, Sha Z, Jiang Y, Terhune J, Kucuktas H, Peatman E, Liu Z. 2010. Structure and expression of transferrin gene of channel catfish, *Ictalurus punctatus. Fish Shellfish Immunol.* 28:159–166.

Loytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320:1632–1635.
MacGillivray RTA, Mason AB. 2002. Transferrin. In: Templeton DM, editor. Cell and molecular biology of iron transport. New York: Dekker. p. 41–69.
MacGillivray RT, Moore SA, Chen J, et al. (14 co-authors). 1998. Two high-resolution crystal structures of the recombinant N-lobe of human transferrin reveal a structural change implicated in iron release. *Biochemistry* 37:7919–7928.
Mikawa N, Hirono I, Aoki T. 1996. Structure of medaka transferrin gene and its 5′-flanking region. *Mol Mar Biol Biotechnol.* 5:225–229.
Mizutani K, Mikami B, Aibara S, Hirose M. 2005. Structure of aluminium-bound ovotransferrin at 2.15 Angstroms resolution. *Acta Crystallogr D Biol Crystallogr.* 61(Pt 12):1636–1642.
Moretti S, Armougom F, Wallace IM, Higgins DG, Jongeneel CV, Notredame C. 2007. The M-coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Res.* 35:W645–W648.
Murayama E, Herbomel P, Kawakami A, Takeda H, Nagasawa H. 2005. Otolith matrix proteins OMP-1 and Otolin-1 are necessary for normal otolith growth and their correct anchoring onto the sensory maculae. *Mech Dev.* 122:791–803.
Nakamasu K, Kawamoto T, Shen M, Gotoh O, Teramoto M, Noshiro M, Kato Y. 1999. Membrane-bound transferrin-like protein (MTf): structure, evolution and selective expression during chondrogenic differentiation of mouse embryonic cells. *Biochim Biophys Acta.* 1447:258–264.
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 DG, Heringa J. 2000. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 302: 205–217.
Octave JN, Schneider YJ, Crichton RR, Trouet A. 1983. Iron mobilization from cultured hepatocytes: effect of desferrioxamine B. *Biochem Pharmacol.* 32:3413–3418.
Ong WY, He X, Chua LH, Ong CN. 2006. Increased uptake of divalent metals lead and cadmium into the brain after kainite-induced neuronal injury. *Exp Brain Res.* 173:468–474.
Pakdaman R, Petitjean M, El Hage Chahine JM. 2008. Transferrin—a mechanism for iron uptake by lactoferrin. *Eur J Biochem.* 254(1):144–153.
Palaksha KJ, Shin GW, Kim YR, Jung TS. 2008. Evaluation of non-specific immune components from the skin mucus of olive flounder (*Paralichthys olivaceus*). *Fish Shellfish Immunol.* 24:479–488.
Peatman E, Terhune J, Baoprasertkul P, Xu P, Nandi S, Wang S, Somridhivej B, Kucuktas H, Li P, Dunham R, Liu Z. 2008. Microarray analysis of gene expression in the blue catfish liver reveals early activation of the MHC class I pathway after infection with *Edwardsiella ictaluri*. *Mol Immunol.* 45:553–566.
Pei J, Sadreyev R, Grishin NV. 2003. PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* 19: 427–428.
Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem.* 25: 1605–1612.
Phuong TM, Do CB, Edgar RC, Batzoglou S. 2006. Multiple alignment of protein sequences with repeats and rearrangements. *Nucleic Acids Res.* 34:5932–5942.
Rombel IT, Sykes KF, Rayner S, Johnston SA. 2002. ORF-FINDER: a vector for high-throughput gene identification. *Gene* 282(1–2): D33–D41.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
Saleh MC, Los-Monteros AE, Zerpa GAD, et al. (19 co-authors). 2003. Myelination and motor coordination are increased in transferrin transgenic mice. *J Neurosci Res.* 72:587–594.
Sarropoulou E, Sepulcre P, Poisa-Beiro L, et al. (11 co-authors). 2009. Profiling of infection specific mRNA transcripts of the European seabass *Dicentrarchus* labrax. *BMC Genomics* 10:157.
Schaeffer E, Lucero MA, Jeltsch JM, Py MC, Levin MJ, Chambon P, Cohen GN, Zakin MM. 1987. Complete structure of the human transferrin gene. Comparison with analogous chicken gene and human pseudogene. *Gene* 56:109–116.
Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A.* 95:5857–5864.
Sekyere E, Richardson DR. 2000. The membrane-bound transferrin homologue melanotransferrin: roles other than iron transport? *FEBS Lett.* 483:11–16.
Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. 2010. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* 38: 161–166.
Stafford JL, Neumann NF, Belosevic M. 2001. Products of proteolytic cleavage of transferrin induce nitric oxide response of goldfish macrophages. *Dev Comp Immunol.* 25:101–115.
Subramaniam AR, Kaufmann M, Morgenstern B. 2008. DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. *Algorithm Mol Biol.* 3:6.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.
Tan SL, Mohd-Adnan A, Mohd-Yusof NY, Forstner MR, Wan KL. 2008. Identification and analysis of a prepro-chicken gonadotropin releasing hormone II (preprocGnRH-II) precursor in the Asian seabass, *Lates calcarifer*, based on an EST-based assessment of its brain transcriptome. *Gene* 411:77–86.
Teehan GS, Bahdouch D, Ruthazer R, Balakrishnan VS, Snydman DR, Jaber BL. 2004. Iron storage indices: novel predictors of bacteremia in hemodialysis patients initiating intravenous iron therapy. *Clin Infect Dis.* 38:1090–1094.
Tsantes C, Steiper ME. 2009. Age at first reproduction explains rate variation in the strepsirrhine molecular clock. *Proc Natl Acad Sci U S A.* 106:18165–18170.
Uniprot Consortium. 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 40:D71–D75.
Wang S, Peatman E, Abernathy J, et al. (41 co-authors). 2010. Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome Biol.* 11:R8.
Williams J. 1968. A comparison of glycopeptides from the ovotransferrin and serum transferrin of the hen. *Biochem J.* 108:57–67.
Yang L, Gui JF. 2004. Positive selection on multiple antique allelic lineages of transferrin in the polyploid *Carassius auratus*. *Mol Biol Evol.* 21:1264–1277.
Yang Z. 2006. Computational molecular evolution. New York: Oxford University Press.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.