

## Evolution of the ferritin family in vertebrates

Jun-Hoe Lee,<sup>1</sup> Kiew-Lian Wan,<sup>1,2</sup>  
Adura Mohd-Adnan,<sup>1,2</sup> Toni Gabaldón<sup>3</sup>

<sup>1</sup>School of Biosciences and Biotechnology,  
Faculty of Science and Technology,  
Universiti Kebangsaan Malaysia,  
<sup>2</sup>Malaysia Genome Institute, UKM-MTDC  
Technology Centre, Selangor, Malaysia;  
<sup>3</sup>Bioinformatics and Genomics  
Programme, Centre for Genomic  
Regulation (CRG), Barcelona, Spain

### Abstract

Ferritins are ubiquitous, highly-conserved proteins that constitute one of the most important components of the cellular machinery devoted to the management of iron levels. Various ferritins have been described in vertebrates, though their exact functions and phylogenetic relationships remain to be established. Our attempts to properly annotate two ferritin subunits isolated from the Asian sea bass *Lates calcarifer*, prompted us to investigate the evolutionary relationships among vertebrate ferritins and their relationships with non-vertebrate homologs. We carried out a detailed screening of mined ferritin sequences by examining the regulatory elements and gene structures. Subsequently, we performed comprehensive phylogenetic analyses involving the various metazoan and vertebrate ferritin chain types, respectively. Our analyses suggest that a single ferritin chain duplicated in the early vertebrates and that the various ferritin chain types in vertebrate and non-vertebrate species evolved independently through lineage-specific duplications. Notably, this includes the mitochondrial ferritin found only in insects and mammals that we show to result from two parallel lineage-specific duplications followed by convergent events of mitochondrial targeting. Regarding the various cytosolic ferritin chains in vertebrates, our results suggest a scenario of a duplication at the base of vertebrates followed by more recent duplications in teleosts and amphibians. This scenario implies that the light chain in mammals is orthologous to the middle chain in teleosts, in contrast to previous claims of a paralogous relationship coupled with differential gene loss. We hypothesise that the extensive differences in sequence and function between these two orthologous chains may have been driven by the adaptation of tetrapods to terrestrial environments, which involved changes in the dynamics of iron uptake and

storage. Altogether, our analyses clarify the evolutionary relationships among vertebrate ferritins and pave the way for the interpretation of functional adaptations within an evolutionary framework.

### Introduction

Iron is a vital trace element required by almost all organisms, in which it conjugates with various proteins to form metalloproteins, acting as an electron carrier or biocatalyst. These proteins, in turn, are involved in major biological processes such as energy metabolism, cell proliferation, immunity and DNA repair.<sup>1-3</sup> The high redox potential of iron that facilitates its biological role is also a double-edged sword, since free iron can trigger the formation of oxygen radicals that could damage cellular lipids, proteins and nucleic acid.<sup>4</sup> Consequently, organisms have developed a complex and tightly-regulated system to balance the dynamic needs of cellular iron and to prevent damage caused by excessive amounts of unbound iron.<sup>5</sup> Central to iron metabolism is ferritin, an ubiquitous protein that is able to sequester large amounts of iron and store it efficiently in a non-toxic and reversible form.<sup>6</sup>

The structure of the ferritin protein is highly conserved from bacteria to human, which underscores its biological importance.<sup>7</sup> A typical apoferritin molecule consists of a hollow globular protein polymer, made of 24 ferritin chains, with a cavity that can store up to 4000 iron atoms as a mineralised iron core.<sup>6</sup> In bacteria, three ferritin-like proteins have been isolated: one composed of 24 identical subunits (FTN), another that contains a covalent-bound haem group (BFR) and finally a smaller 12-mer DNA-binding protein (Dps).<sup>8,9</sup> The gene organisation of ferritin is highly conserved within the animal kingdom, displaying a 4-exon/3-intron structure with nearly-identical intron positions.<sup>10</sup> In mammals, two ferritin chains have been identified, which have been termed as heavy (H) and light (L) chains respectively.<sup>6</sup> While in non-vertebrate metazoans, two ferritin subunits that were claimed to be homologous to the mammalian H- and L-chains have been discovered in *Drosophila* and *Trichoplusia*.<sup>11,12</sup> Two subunits have also been identified in *Caenorhabditis elegans* (FTN-1 and FTN-2), but only one of them (FTN-1) is critical for iron regulation.<sup>13</sup> In addition, mitochondrial ferritin has been discovered in several mammalian species, *Drosophila* and plants.<sup>14,15</sup>

The regulation of ferritin in metazoans is mediated at both the transcriptional and translational levels.<sup>16</sup> However, only the translational control appears to be iron-dependent

Correspondence: Toni Gabaldón, Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Dr. Aiguader Street 88, Barcelona 08003, Spain.  
Tel. +34.93.316.01.00 - Fax: +34.93.316.00.99.  
E-mail: tgabaldon@crge.es

Key words: ferritin, vertebrate, evolution, mitochondria.

Acknowledgements: this work was supported by a grant from the Malaysian Ministry of Science, Technology and Innovation (07-05-MGI-GMB009) and a grant from the Spanish Ministry of Science and Innovation (BFU2009-09168). The authors wish to thank the Marine Fish Production and Research Centre, Terengganu, Malaysia for providing the fish samples and Salvador Capella-Gutiérrez for valuable help in the phylogenetic analysis.

Contributions: JHL generated the data, sequenced the clones, analysed and interpreted the data, and drafted the manuscript. KLW was involved in the generation of the clones. AMA and TG conceived and designed the study, participated in the interpretation of the data and drafting of the manuscript.

Received for publication: 25 November 2011.  
Revision received: 15 February 2012.  
Accepted for publication: 17 February 2012.

This work is licensed under a Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).

©Copyright J.H. Lee et al., 2012  
Licensee PAGEpress, Italy  
*Trends in Evolutionary Biology* 2012; 4:e3  
doi:10.4081/eb.2012.e3

where it is mediated through a conserved RNA structure in the 5'-untranslated region (UTR) known as the iron-responsive element (IRE).<sup>17</sup> IREs provide a binding site for the iron-regulatory proteins (IRPs), which repress ferritin translation in conditions of low iron.<sup>18</sup> The IRE consists of a stem-loop structure of 26-30 nucleotides in length, with an apical loop sequence of CAGUGN, and a UGC/C-bulge five nucleotides upstream of the apical loop.<sup>17</sup> On the other hand, the transcriptional control of ferritin is linked to oxidative stress, inflammation and immunity, which subsequently trigger the increased synthesis of a group of genes (including ferritin) known as phase II detoxifying enzymes.<sup>19,20</sup> The transcription of these genes are effected through the binding of the heterodimer nuclear factor E2-related factor 2 (Nrf2)/small Maf protein to a conserved DNA sequence known as the antioxidant-responsive element (ARE).<sup>20</sup>

The majority of the studies in eukaryotes

have been focused so far on the mammalian H- and L-chains.<sup>6</sup> The mammalian H-chain is a 21 kDa peptide that contains a ferroxidase center and functions primarily to rapidly oxidise Fe(II).<sup>21</sup> On the other hand, the L-chain is a 19 kDa peptide with a number of carboxyl ligands along its cavity surface that contribute to its role of iron nucleation to form ferrihydrite.<sup>22</sup> The L-chain also contains a *salt bridge* that provides enhanced protein stability in varying pH conditions.<sup>23</sup> The mechanism of iron uptake by ferritin can be summarised into 3 main phases: i) Fe<sup>2+</sup> binding, ii) Fe<sup>2+</sup> oxidation to Fe<sup>3+</sup> by the H-chain, and iii) nucleation and formation of an iron core by the L-chain.<sup>24,25</sup> Additional studies have also revealed that oxidation of iron by H-chains is a relatively quick step compared to the mineralisation phase carried out by the L-chains.<sup>26</sup> The overall iron uptake rate, however, is dependent on the ratio of H/L chains in a ferritin molecule. These complementary functions of both chains could explain the larger proportions of L-chains in isoferritins of liver and spleen tissues that are involved in iron storage, while H-chains are more predominant in other tissues such as heart and kidney that are involved in rapid iron exchange.<sup>27</sup>

In comparison to their mammalian counterparts, ferritins from other vertebrates have not been well-studied. A third ferritin subunit, the middle (M)-chain, was first isolated along with the H- and L-chain from bullfrog tadpoles.<sup>28</sup> Currently, only two studies have clearly identified M-chain peptide sequences from two Antarctic teleosts, *Trematomus bernacchii* and *Trematomus newnesi*, while other predicted M-chain sequences in the GenBank database are often identified as additional H- or L-chain isoforms.<sup>29,30</sup> The M-chain is a 19.9 kDa peptide capable of carrying out dual tasks in iron storage (*i.e.* iron oxidation and core-formation).<sup>29</sup> This unique ability was explained through ligands detected that are similar to the ferroxidase centres of the H-chain and carboxyl groups of typical L-chains. To date, there have been no reported M-chains isolated from any other amphibians or mammals. Additionally, there are no published studies of L-chains isolated from fish or other earlier divergent vertebrates. These findings have prompted speculations on the possible evolution of the ferritin family, in which an ancestral multi-functional M-chain, conserved in fish and amphibians and subsequently lost in mammals, was substituted with a highly-specialised L-chain that originated in the tetrapod lineage.<sup>30</sup> Thus, according to this hypothesis, amphibians would represent an intermediate state of the ferritin family evolution and the M- and L-chains would be paralogous proteins with complementary distributions in fish and mammals,

respectively. Nevertheless, these speculations have never been backed by proper phylogenetic analyses and the actual evolutionary relationships among ferritin chains remains to be clarified. This situation is in sharp contrast with that found in plant ferritins, for which evolutionary relationships have been recently investigated.<sup>31</sup>

Our studies on gene expression of infected *L. calcarifer* identified ferritin among the proteins found to be upregulated during the early stages of infection by *Cryptocaryons irritans* (not published). In order to study ferritin in detail, we generated two putative *L. calcarifer* full-length ferritin cDNA sequences. Although BLAST hits have identified one of the proteins as an M-chain, initial phylogenetic analyses indicated orthology relationships with the mammalian H-chain and several different amphibian chains. To further clarify the evolutionary relationships among ferritins in different vertebrates we conducted an exhaustive phylogenetic analysis of the protein sequences, as well as both the transcriptional and translational regulatory elements of their corresponding genes. Our results clarify both the orthology and paralogy relationships among the ferritin subunits and reveal a complex history of duplications and functional divergences. Altogether our results show that evolutionary relationships are best established through phylogenetic analysis and that initial annotation of proteins based on mere BLAST matches can mislead evolutionary interpretations.

## Materials and Methods

### RNA extraction and cDNA preparation

Liver tissues were harvested from adult *L. calcarifer* obtained from the Marine Fish Production and Research Centre, Tg. Demong, Terengganu, Malaysia. Tissue samples were immediately collected from anaesthetised fish and stored in liquid nitrogen at -80°C for further use. The tissues were then homogenised using mortar and pestle under liquid nitrogen, and total RNA was extracted using TRIReagent™ (Molecular Research Center, Inc., Cincinnati, OH, USA) according to the manufacturer's instructions. Subsequently, RNA clean-up was carried out using the Rneasy Minikit (Qiagen, Kuala Lumpur, Malaysia). Total RNA concentration was measured by spectrophotometry and its quality was assessed by agarose gel electrophoresis. Preparation of 5' and 3'-RACE-ready cDNA were then carried out with SuperScript III from the GeneRacer™ Kit (Invitrogen, Inc., Carlsbad, CA, USA) with random primers for

5'-cDNA and GeneRacer oligo-dT primers for 3'-cDNA.

### Primer designs and rapid amplification of cDNA-ends (RACE)

Two pairs of RACE primers (FerM-F and FerM-R; FerH-F and FerH-R) were designed based on incomplete open reading frames (ORF) of putative ferritin clones assembled from the Asian seabass EST library (not published) (Supplementary Table 1). 5' and 3'-RACE were carried out using the GeneRacer 5'/3'-primers and touchdown PCR as recommended by the manufacturer's protocols.<sup>32</sup> Touchdown PCR was used as follows: 94°C for 3 min, 5 cycles of 94°C for 30 sec, 72°C for 2 min; 5 cycles of 94°C for 30 sec, 70°C for 2 min; 25 cycles of 94°C for 30 sec, 63.3-67.9°C for 30 sec, 70°C for 2 min and final extension at 70°C for 3 min.

### Sequence analysis and data mining

DNA sequence data obtained were quality-clipped based on a Phred-value threshold of 30 and vector sequences were trimmed with CROSS\_MATCH.<sup>33</sup> Sequences from 5' and 3'-RACE-PCRs were combined using the Cap3 contig assembly tool in BioEdit (version 7.0.5.3).<sup>34</sup> Complete cDNA sequences were then screened for the presence of IREs using SIREs (<http://ccb.gimppc.org/sires/index.html>).<sup>35</sup> ORF Finder was then used to predict the open reading frames (ORFs) in the sequence data (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>).

An initial protein BLAST (BLASTP 2.2.26+) search was carried out against the NCBI GenBank. Subsequently, the amino acid and mRNA sequences from both PCR products respectively were used as seed to retrieve similar protein sequences from the Ensembl database of selected vertebrate genomes (release 60) (<http://www.ensembl.org/index.html>).<sup>36</sup> The cDNA sequence database was mined using the tBLASTN search algorithm with an E-value cutoff of <0.0001.<sup>37</sup> An additional similarity search was performed using the GenBank EST database to retrieve sequences from *Xenopus laevis* and other chordates (*Ciona intestinalis*, *Ciona savignyi*, *Branchiostoma floridae*, *Petromyzon marinus* and *Eptatretus burgeri*) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Finally, the annotated ferritin sequences of *Rana catesbiana*, *Drosophila melanogaster*, *Haliotis discus*, *Crassostrea gigas* and *Eriocheir sinensis* were manually retrieved from GenBank and added into the dataset (<http://www.ncbi.nlm.nih.gov/genbank/>). The ORFs for all retrieved sequences from Ensembl and GenBank were predicted using ORF Finder.

## Analysis of gene regulatory elements

To check for the possibility of pseudogenes, sequences were screened for the presence of IREs in the 5'-UTR using SIREs.<sup>35</sup> Sequences were also scanned for the presence of signal peptides using the SignalP 3.0 web server (<http://www.cbs.dtu.dk/services/SignalP/>).<sup>38</sup> The RNA secondary structures predictions were carried out using the Mfold web server (<http://mfold.rna.albany.edu/?q=mfold>),<sup>39</sup> with the arrangement of the apical loop nucleotides modified based on previous structural studies.<sup>40,41</sup> Subsequently, the exon/intron structure of retrieved sequences were examined based on gene information from Ensembl.<sup>36</sup>

The presence of AREs was predicted by screening the region of 7 kb upstream of the using the MEME Suite web server (version 4.6.1) (<http://meme.nbcrl.net/>).<sup>42</sup> Putative DNA motifs were predicted using the MEME algorithm and compared with known motifs in the Jaspur, Transfac and Uniprobe databases using the TOMTOM scanning algorithm.<sup>42</sup>

## Multiple sequence alignment and codon substitution analysis

Multiple sequence alignments were carried out using a similar strategy as implemented in the PhylomeDB (<http://phylomedb.org/>) pipeline, which combines several aligners and the *Head or Tails* approach.<sup>43,44</sup> Briefly, alignments were performed in both forward and reverse directions using Muscle (version 3.7),<sup>45</sup> MAFFT (version 6.717b),<sup>46</sup> and DIALIGN-TX (version 1.0.2).<sup>47</sup> A consensus output based on the 6 previous alignments was generated using M-Coffee.<sup>48</sup> trimAL was then used to remove large gap regions (with the default *gappout* option) and to generate a codon alignment using the reverse-translate feature.<sup>49</sup>

Codon alignment was then analysed using the *yn00* utility in PAML to assess its reliability for subsequent phylogeny reconstruction.<sup>50</sup> The synonymous and non-synonymous substitution rates between each sequence pair were estimated using the Yang and Nielsen method as implemented in PAML.<sup>51</sup>

## Phylogenetic analysis of vertebrate ferritin

Phylogenetic analyses were carried out using the Maximum Likelihood (ML) and Bayesian algorithms as implemented by PHYML (version 3.0)<sup>52</sup> and MrBayes (version 3.1.2),<sup>53</sup> respectively. To examine the broad evolutionary history of ferritin across metazoans, an unrooted phylogenetic reconstruction was performed using selected vertebrate and non-vertebrate amino acid sequences. This was followed by a second phylogenetic analysis to specifically investigate the phylogeny of vertebrate ferritin

genes. For the ML analyses, the LG model were used for the amino acid sequences with optimised gamma distributions and proportion of invariant sites.<sup>54</sup> Non-parametric bootstrap tests of 1000 replications were also performed. As for the Bayesian analysis, mixed models approach were applied for the amino acid sequences.<sup>53</sup> Four independent chains were started respectively from two runs simultaneously and the period of sufficient convergence was assessed using AWTY (<http://king2.scs.fsu.edu/CEBProjects/awty/awty.php>).<sup>55</sup> For both analyses on metazoan and vertebrate sequences respectively, 5 million generations were ran (sampling every 100th tree) and the first 12,500 trees were discarded.

## Results

### Identification of putative M-chain and H-chain ferritin in Asian seabass

The complete sequences for putative M- and H-chain ferritin transcripts were obtained through 5' and 3'-RACE. Final assembled transcript sequences for putative M-chain (Latca\_1) and H-chain (Latca\_2) of *L. calcarifer* are 1130 and 1148 nucleotides in length with predicted ORFs of 531 and 534 nucleotides, respectively. We detected IRE sequences 173 and 253 nucleotides upstream of the start codon in Latca\_1 (TTGCTTCAACAGTGCCTTGAACGG) and Latca\_2 (CTGCTTCAACAGTGCCTTGAACGG), respectively, which supports their identity as iron-regulated ferritin-coding genes (Figure 1). Sequences have been deposited in GenBank with accession numbers HQ424195 and HQ424196, respectively.

### Data mining

The initial BLAST search against the NCBI GenBank protein database returned sequences that were annotated with various or unspecified subunit names (Supplementary Figure 1). Additionally, we found that the sequence information of all functional ferritin subunits were not available for several species. Therefore we chose to proceed with an assumption-free BLAST search against the Ensembl database.

Using BLAST searches against the Ensembl database (release 60), we retrieved 87 similar sequences from 14 species (Supplementary Table 2). The retrieval of additional sequences from GenBank was necessary due to the presence of only a single amphibian species (*Xenopus tropicalis*) in the Ensembl database while the chordate sequences were to be used as an outgroup. Similarity searches against the *Xenopus laevis* EST database identified 5 individually-distinct ferritin sequences with

IREs. As for the chordates, similarity searches retrieved a unique IRE-bearing sequence in *Branchiostoma floridae*, *Petromyzon marinus* and *Eptatretus burgeri*. Additionally, sequences from the urochordates (*Ciona intestinalis* and *Ciona savignyi*) were retrieved from GenBank as the initial sequences retrieved from Ensembl show a high degree of divergence, but lacked information on the 5'-UTR and appeared truncated.

A major difficulty in carrying out a detailed analysis on ferritin is the presence of numerous pseudogenes and various proteins that share sequence similarity with the conserved ferritin domains yet likely do not play a significant role in iron regulation. Several of these ferritin-like sequences (that lack the IRE and resemble processed pseudogenes) have been shown to be expressed in *Rana catesbiana*,<sup>28</sup> while in human and mouse these genes are associated with the coding of tumour-specific antigens.<sup>56,57</sup> Therefore, we undertook many stringent measures to identify and exclude these spurious sequences from our analysis. The first consisted of checking for the presence of IREs, which were found in 52 Ensembl sequences and in all the additional sequences retrieved from GenBank. Subsequently, we examined the exon/intron structures and 22 sequences were found to be intronless or possessed structures that differed from the 4-exon/3-intron arrangement. Among the intronless sequences, the presence of a signal peptide was predicted in four sequences: *Homo sapiens* Mtf, *Mus musculus* Mtf, *Rattus norvegicus* Mtf and *Bos taurus* Mtf. Additionally, the selection of the representative species of a taxon, particularly for sequences mined from the GenBank EST database, was constrained by those that possessed all the relevant subunits to prevent any subsequent misleading inferences on gene duplications or losses. In total, 30 sequences retrieved from Ensembl passed all filters and were used for phylogenetic analysis.

### Comparative analysis of the regulatory elements

The IRE sequences of the various ferritin subunits show that the apical loop sequence of CAGUGC was conserved in all vertebrates H-chains, except for *Anolis carolinensis* (Anole lizard) and the amphibian sequences (Figure 1). For the M/L-chains, variation in the sixth residue of the apical loop sequences was observed across the vertebrates. The presence of a UGC bulge was predicted in *B. floridae* as well as most vertebrate sequences except for the agnathans (*P. marinus* and *E. burgeri*) and sauropsids (*A. carolinensis* 2 and *Taeniopygia guttata* 2) that possessed a C-bulge.

Initially, no IRE was detected in the *Ciona* sequences by the SIREs web server. However, a closer analysis of the sequences show the

presence of a putative IRE in the 5'-UTR of *C. intestinalis* and *C. savignyi*. Both putative IREs fulfil all of the features of canonical IREs with the exception of a guanine residue in place of a cytosine residue to form an unpaired bulge at the eighth nucleotide position. However, IRE sequences that have a guanine bulge at the eighth residue position and an apical loop of CAGUGU have been reported in several species from the Crustacea subphylum, which support the identity of the *Ciona* sequences as iron-regulated ferritins.<sup>58</sup>

The presence of AREs were predicted approximately 4.5 kb upstream of the coding regions of the mammalian H-chain sequences, and 1.5 kb upstream of the mammalian L-chain sequences (Figure 2). No ARE-similar motifs were predicted in the non-mammalian and mitochondrial sequences. The consensus motif predicted by MEME for the mammalian H- and L-chains was found to be a significant match to the core ARE sequence (P value: 3.96e-09). While the MEME motif does not include any additional nucleotides downstream of the core ARE sequence, the nucleotides upstream of the core ARE show similarities with the extended ARE.

### Multiple sequence alignment and codon substitution analysis

Comparison of the amino acid sequence alignment show that Latca\_2 possesses all the residues of the ferroxidase centres (Glu24, Glu58, Glu59, His62, Glu104, Gln138) conserved in mammalian H-chains (Supplementary Figure 2). On the other hand, Latca\_1 shares all the conserved ferroxidase center residues and contains two conserved residues found in the mammalian L-chain that are involved in nucleation (Glu54, Glu61). However, a third nucleation residue (Glu57) was found to be less conserved within the putative teleost M-chains. Only *Danio rerio* appear to share that residue whereas it has been substituted with Asp in Latca\_1, *Gasterosteus aculeatus*, and *Takifugu rubripes*, and His in *Oryzias latipes*.

Analysis of the codon substitution rates by PAML shows that approximately 85% of the sequence pairs has a dS (number of synonymous substitutions per synonymous sites) value  $\geq 2.00$  (data not shown), which suggest codon saturation at the nucleotide level and consequently, loss of reliable phylogenetic information.<sup>59</sup> Therefore, we did not proceed with the codon-based phylogenetic analysis.

### Phylogeny of vertebrate ferritin

An initial phylogenetic analysis of all sequences retrieved from Ensembl using Maximum Likelihood show a topology that lacks support in many branches and several long-branch attraction groups, such as the clustering of *R. catesibiana* and *C. intesti-*

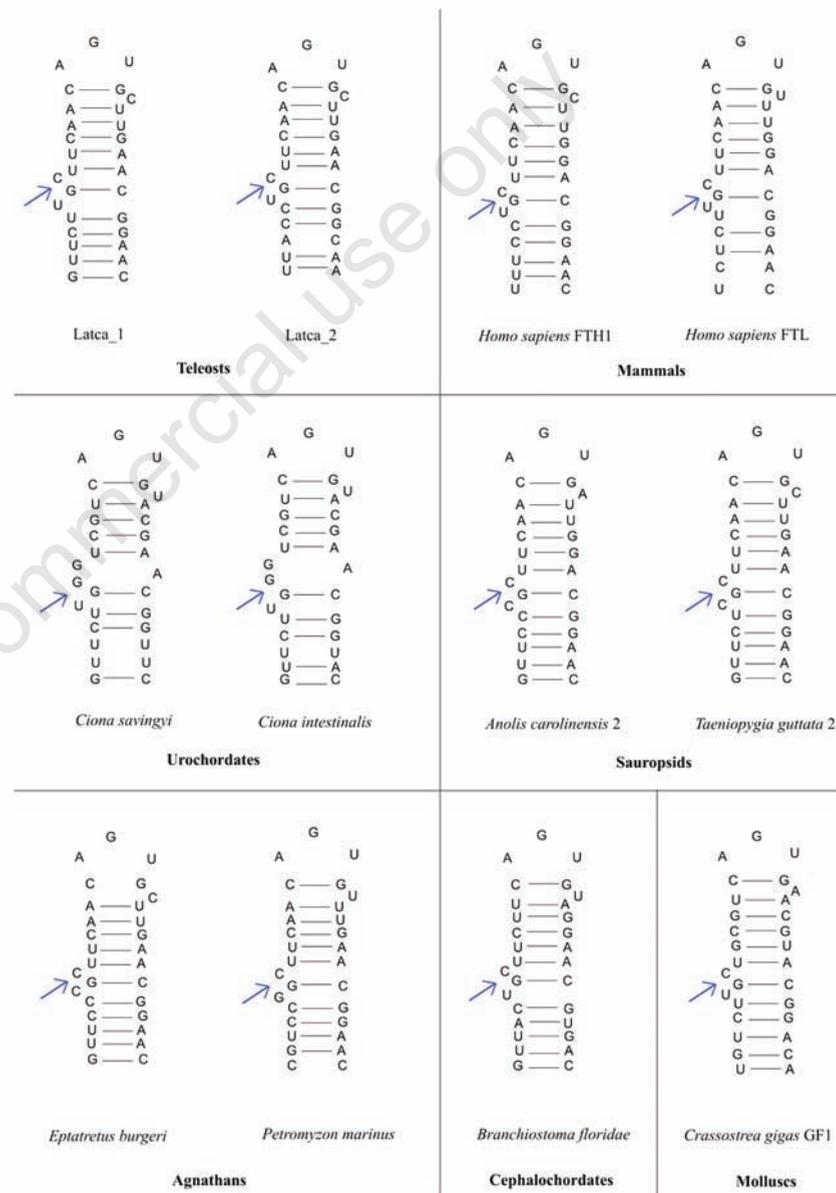
*nalis\_03967* (Supplementary Figure 3). Other than the clustering of the mammalian mitochondrial ferritin (Mtf), the grouping of the sequences do not appear to follow any clear patterns.

For the phylogenetic analysis on metazoan ferritins, 14 vertebrate (7 species, including *L. calcarifer*) and 10 non-vertebrate sequences (4 species) were used. The choice of such a set of vertebrate and non-vertebrate sequences was made to minimise the effect of long-branch attraction artefacts, as several sequences (*e.g.* amphibians, *Ciona*) were found to be highly-divergent (discussed below).

The phylogenetic trees constructed from the metazoan amino acid sequences using

Maximum Likelihood (Figure 3) and Bayesian analyses (Supplementary Figure 4) show a clear separation between the vertebrate and non-vertebrate sequences, indicating independent diversification of the various taxa chains in each lineage. Within the vertebrate taxa, the sequences appear to be grouped according to their chain types. However, the non-vertebrate ferritin sequences appeared to group with a less discernible pattern.

Subsequent phylogenetic analysis of the vertebrate ferritin subunits was based on sequences from 20 species (2 urochordates, 1 cephalochordate, 2 agnathans, 6 teleosts, 3 amphibians, 2 sauropsids and 4 mammals), including *L. calcarifer*. Although a broader tax-



**Figure 1. Secondary structure of ferritin iron-responsive elements (IREs) of selected metazoans. The IREs of teleost, mammalian, cephalochordate, ferritins as well as several mollusc species (*C. gigas*) possess a U-G-C bulge structure (indicated by blue arrows). However, variations in the bulge structures were predicted for the agnathans, urochordates and one of the ferritin subunits from the sauropsids.**

onomic sampling, especially around the basal lineages, would have been desirable, we are currently limited by the relatively scarce genomic sampling of vertebrates. The use of sequences from incomplete genomes was limited to those showing the IRE regulatory region and a minimum 90% coverage of the coding sequences. Attempts to increase the taxonomic coverage by relaxing the thresholds and including partial sequences resulted in clearly artefactual topologies (Supplementary Figure 3).

The resulting phylogenetic reconstructions using ML (Figure 4) and Bayesian (Supplementary Figure 5) analyses display generally consistent topologies in terms of the clustering of sequences according to chain types. In general, all trees show a similar pattern that suggest accelerated rates of evolution in L/M-chain sequences resulting in the presence of some inconsistencies with the known phylogeny of vertebrates.

## Discussion

While there have been extensive studies on ferritin of mammals, there remains many gaps in our current knowledge about ferritin in other groups of vertebrates and the evolutionary history of this gene family. Our characterisation of two ferritin subunits in *L. calcarifer*, which consist of the M- and H-chain supports previous studies on the existence of two distinct, iron-regulated ferritin subunits in teleosts.

### Variations in the regulatory mechanisms of ferritin

Our analysis of the secondary structure of IREs indicate that the cephalochordate *B. floridae* share the UGC-bulge structure found in most vertebrate ferritins, yet this was missing in the agnathans (*E. burgeri* and *P. marizon*), the predicted anole lizard H-chain (*A. carolinensis* 2) and zebra finch L-chain (*T. guttata* 2) (Figure 1). However, UGC-bulges have been reported in various non-vertebrates such as the echinoderm *Asterias forbesi*<sup>60</sup> and several mollusc species (*H. discus*, *C. gigas*).<sup>61,62</sup> Additionally, variations of CGC and CGG-bulges had been predicted in the echinoderm *Apostichopus japonicus* as well as the crustaceans *Pacifastus vannamei* and *Litopenaeus vannamei*.<sup>58</sup> A previous study on the evolution of IREs showed that variations in the bulge and apical loop are common for related species.<sup>58</sup> These observations, along with our analysis, suggest that the bulge structure, as well as other factors such as the apical loop, might have evolved independently several times, and converged on similar bulge/apical loop sequence structures accommodating specific

affinities and regulatory requirements.<sup>63,64</sup>

The predicted presence of AREs in both mammalian H- and L-chain sequences confirm and extend previous studies on mouse<sup>65,66</sup> and human ferritins.<sup>67,68</sup> The predicted absence of AREs in non-mammalian ferritin sequences may result from variations in the DNA-binding sequence from the core ARE sequence, complicating its detection.<sup>69</sup> Further investigations

into non-mammalian ARE (or the possibility of its absence) would be highly useful as homologs of Nrf-2 and small Maf proteins have been reported in chicken<sup>70,71</sup> and zebrafish.<sup>72</sup>

### A parallel duplication followed by mitochondrial targeting of ferritin

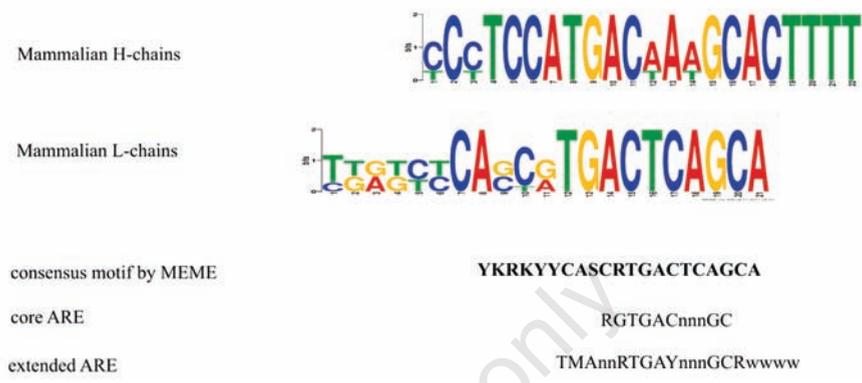


Figure 2. The motifs predicted in mammalian ferritin sequences show high similarity to the antioxidant-responsive element (ARE) sequence. The consensus motif predicted by MEME Suite web server (version 4.6.1) shows high similarity to the core ARE sequence and the upstream nucleotides in the extended ARE. No ARE-matching motifs were predicted in non-mammalian ferritins.

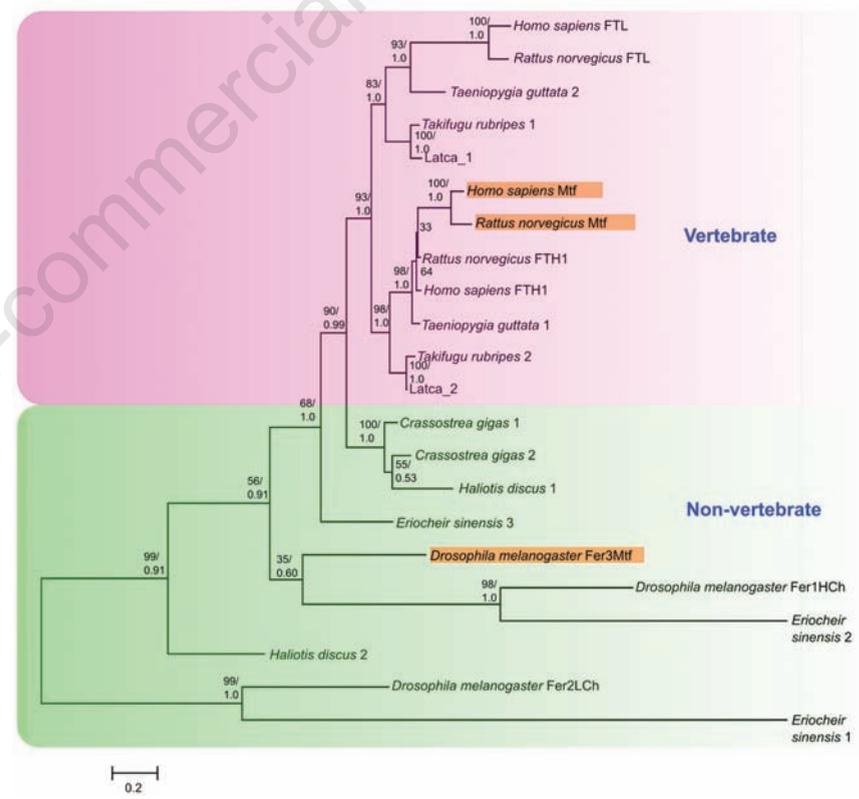


Figure 3. Maximum likelihood tree generated from amino acid sequences of vertebrates and non-vertebrates. A clear distinction is observed for the clustering of the various ferritin chain types within the vertebrate and non-vertebrate taxa. The mammalian (*H. sapiens*, *M. musculus*) and non-vertebrate (*D. melanogaster*) mitochondria sequences are highlighted in brown and clustered separately. Values at nodes indicate Maximum Likelihood percentages (1000 replications)/Bayesian posterior probabilities. The scale bar represents the estimated number of substitutions per site.

## in insects and mammals

Our phylogenetic analysis of the vertebrate and non-vertebrate ferritin subunits show a clear distinction between these two groups of lineages (Figure 3, Supplementary Figure 4). This suggest that the different subunits in vertebrate and non-vertebrate lineages evolved independently through lineage-specific duplications. An open question is whether the various chains converged on a similar function in the different lineages. Remarkably, we identify one case of apparent convergent evolution in the case of mammalian and insect mitochondrial ferritin (Mtf).

Mtf has been described in some mammals and in *Drosophila*, where it has been shown to localise to the mitochondria.<sup>6,15</sup> Mammalian Mtf gene is intronless, located on an autosome and is synthesised as a precursor with an N-terminal signal peptide that guides the protein towards the mitochondria.<sup>6</sup> In contrast, the *Drosophila* Mtf is located on the X-chromosome, consists of 2 exons and possesses a truncated signal peptide.<sup>15</sup> Interestingly, many ferritin pseudogenes have been found on the human and rat X-chromosome and are believed to have undergone gene conversion events.<sup>73</sup>

Both *Drosophila* and mammalian Mtf's are primarily expressed in the testis, though mammalian Mtf is also expressed in several other high-metabolic cell types.<sup>6</sup> While the complete array of functions of Mtf in both groups of organisms is still being studied, it is believed that one of the roles Mtf is affording protection against oxidative damage.<sup>74</sup> A recent study has suggested that the mitochondrial iron metabolism in spermatogenesis is conserved from insects to mammals due to the high expression of several proteins such as Mtf in the testis of *Drosophila* and mammals.<sup>75</sup>

Nonetheless, despite notable similarities in function and sites of expression, it is believed that both ferritins evolved separately due to differences in the gene organisation and protein structure.<sup>6,15</sup> Our analysis of the vertebrate/non-vertebrate ferritin phylogeny supports this view. In addition, the Mtf sequences are not clustered according to their chain types, but rather clearly distinguished between the vertebrate and non-vertebrate taxa. The mammalian Mtf are positioned closely to other mammalian H-chains and our data mining analysis did not uncover any Mtf in other vertebrate groups. This suggests that the vertebrate Mtf evolved specifically in mammals, likely from a retro-transposition event where a spliced H-chain mRNA was reinserted into the genome (as suggested by the lack of introns). While the origin of insect Mtf remains unknown, our analysis strongly support the poliphylly of insect and mammalian mitochondrial ferritins. Therefore, this could represent an example of convergent evolution in which two events of duplication fol-

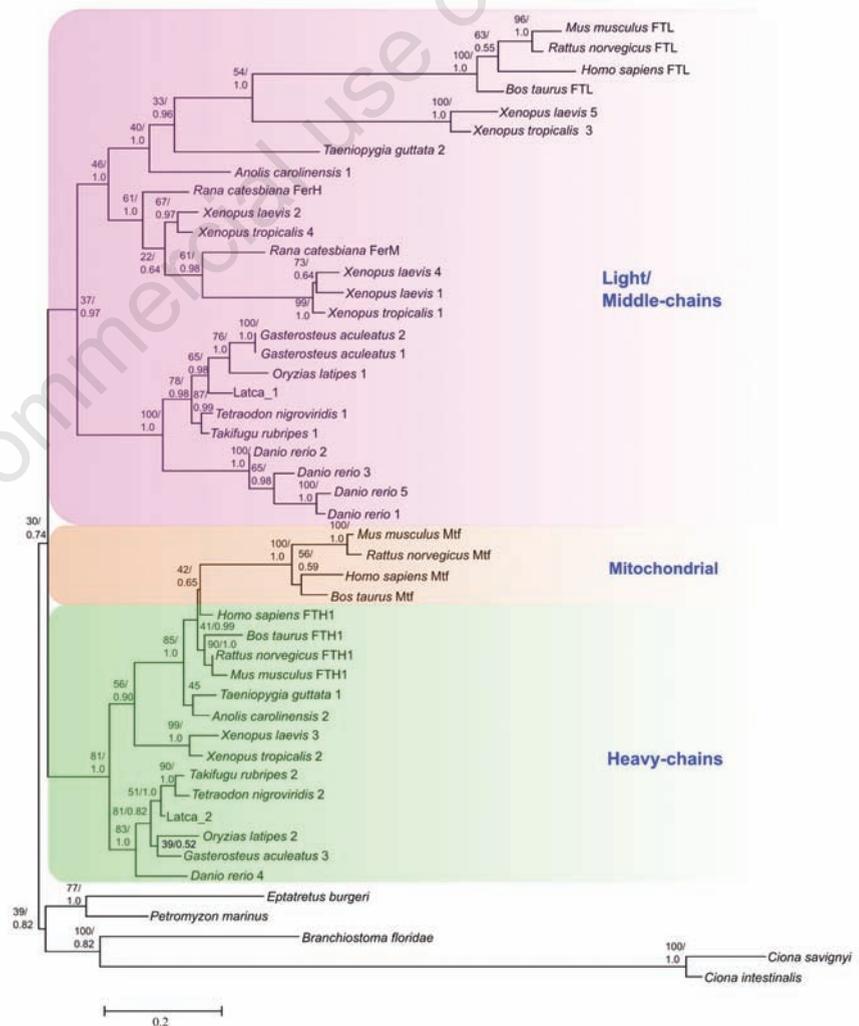
lowed by neo-functionalisation (re-targeting to the mitochondria) occurred independently in two separate lineages.

Mitochondria are major producers of reactive oxygen species due to their roles in cellular iron utilisation and conversion of iron into bioactive forms through the haem and iron-sulphur pathways.<sup>76</sup> Within the mitochondria, the frataxin protein has been suggested to function as an iron chaperone and storage protein, as well as to have roles in the synthesis of iron-sulphur clusters and haem.<sup>77-79</sup> Frataxin is found in the mitochondria of all eukaryotes (including yeast) and has been shown to bind more than 2000 iron atoms into a ferrihydrite structure, similar to ferritin.<sup>80</sup> In contrast, the extremely limited expression of Mtf in *Drosophila* and mammals within high-metabolic cell types suggests that the presence of Mtf is not critical for mitochondrial function in all cell types.<sup>6,15</sup> Expression studies of Mtf in frataxin-deleted yeast cells also showed that it

is able to restore most of the functions of frataxin and prevented iron overload and mitochondrial DNA damage.<sup>81</sup> A suggested hypothesis therefore, would be that the relocation of a ferritin copy to the mitochondrion may have been advantageous in two distant lineages under a similar pressure to complement frataxin function in certain periods or cell types.

## Phylogeny of vertebrate ferritin

Our initial attempt to clarify the evolutionary relationships of the *L. calcarifer* putative M- and H-chains in relation to other vertebrate H- and L-chains were complicated by the presence of pseudogenes, incomplete sequence information (particularly the 5'-UTR) and incorrect sequence annotations (Supplementary Figure 3). To minimise the effect of long-branch attraction we used several representative sequences from various taxa as outgroup.<sup>82</sup> Overall, all the phylogenetic reconstructions from both ML and



**Figure 4.** Maximum Likelihood tree generated from amino acid sequences of vertebrate ferritins. The various ferritin sequences are clustered according to chain types, with heavy (H)-chains forming an orthology group to the middle(M)/light(L)-chains. Values indicate the Maximum Likelihood bootstrap percentages (1000 replications)/Bayesian posterior probabilities. The scale bar represents the estimated number of substitutions per site.

Bayesian approaches show a very similar topology. While a few branches display relatively large discrepancies between the bootstrap values (BP) and Bayesian posterior probabilities (PP), this could be due to the inherent differences on how these two values are calculated.<sup>83</sup> Some studies have claimed that BP values are too conservative,<sup>84</sup> where 70% BP corresponds to a 95% confidence value and that PP values are excessively high.<sup>85,86</sup> However, there are also studies that indicated PP values are still conservative estimates of phylogenetic accuracy.<sup>87,88</sup> Additionally, several studies have suggested that certain Bayesian approaches are able to provide models that fit the data better compared to Maximum Likelihood.<sup>89,90</sup> We next discuss the most likely scenario according to these topologies, considering the existing congruence between two independent methodologies, as well as parsimony criteria in terms of inferred duplications and losses as additional lines of evidence. Undoubtedly, a larger taxonomic sampling, especially around the less supported nodes, will likely help in future studies to completely resolve the phylogeny.

Monophyly of H-chains is clearly recovered by all approaches within an internal topology that largely reflects the accepted vertebrate phylogeny.<sup>91</sup> The partition of the tree containing L- and M-chains displays most of the variability across the different approaches and contains some inconsistencies within the expected species phylogeny; most remarkably the relative positions of a subset of amphibian sequences (*X. tropicalis* 3, *X. laevis* 5) and reptile/bird sequences. Together with the fact that branches in this part of the tree are significantly long, this suggests the presence of phylogenetic noise. Despite these difficulties, some important conclusions can be drawn from this topology.

First, based on the clustering at the base of the phylogenetic tree, we hypothesise that the ancestral vertebrate ferritin originated from a single chain type (Figure 4). Similar to ferritins in extant basal chordates (tunicates, hagfish, lancelet), such ancestral vertebrate ferritin would possess an IRE in the 5'-UTR. Note that this hypothesis is compatible with the presence of basal chordates with additional ferritin chains resulting from lineage-specific duplications. Thus, earlier interpretations that the secondary *Drosophila* ferritin (Fer2LCH) are more closely related to vertebrate L-chains than to H-chains are unsustainable, since both L- and H-chains shared a common ancestor at the base of the vertebrate phylogeny, thus being co-orthologous to the ferritins in *Drosophila*.<sup>92</sup> Our phylogenetic analysis also supports a duplication in the early vertebrates preceding the teleost-tetrapod divergence, which resulted in a paralogous copy that maintained the exon structure and IRE regions. This gene duplication could have been

part of a whole duplication event thought to have occurred after the lamprey-gnathostomes split.<sup>93</sup> Although such event is believed to have involved two rounds of whole genome duplications (2R hypothesis), which would result in at least four paralogous copies in all vertebrates, it has been shown that a majority of gene families do not exactly follow this pattern due to rampant gene loss immediately following the duplication events.<sup>94-96</sup> The phylogeny of the ferritin family is thus consistent with a rapid loss of two of the four 2R paralogs. Similarly, the proposed whole genome duplications that occurred in teleost (3R hypothesis) is not apparent in our trees, which is again in concordance with the observed rampant gene loss following the whole genome duplications in this lineage.<sup>97</sup>

Secondly, regarding the relationship of L- and M-ferritins, all reconstructed phylogenies support an orthologous relationship between the fish M-chains and mammalian L-chains. This is, moreover, the more plausible scenario in terms of inferred duplications and losses. Indeed, the alternative phylogeny supporting a paralogy between L- and M-chains will require two complementary losses in the fish and amniota lineages, respectively.<sup>30</sup>

Considering these results, we disagree with the hypothesis put forth previously that an ancestral H-chain type protein evolved into an M-chain type with enhanced capability to store iron efficiently, and subsequently, mammals concomitantly lost the M-chain while developing the L-chain that is specialised for iron accumulation.<sup>30</sup> In contrast, we propose that two types of ferritin protein existed concurrently in the ancestral vertebrates (*i.e.* teleosts and tetrapods); one of them being an H-chain type with ferroxidase activity and the other, a paralogous copy that had accumulated additional capabilities. In teleosts, the latter is termed M-chain and has been found to possess both ferroxidase and iron nucleation capabilities. On the other hand, in tetrapods this second copy is termed L-chain and has evolved to almost exclusively mineralise iron into a compact iron core.

Although scarce, functional analyses on teleost M-chains have revealed many parallels with mammalian L-chains. The iron nucleation ability of the M-chain enables it to store up to 3500 iron atoms, which is lower than the L-chain (~4000) but significantly higher than the H-chain.<sup>6,30</sup> In terms of biochemical properties, the M-chain is highly stable and is able to maintain its quaternary structure at pH 2.5 similar to mammalian L-chains, which is an intrinsic feature to incorporate higher amounts of iron.<sup>23,29</sup> This is in contrast to mammalian H-chains, which dissociate at pH 3.0. Finally, the M-chain is found to be the only subunit expressed in the spleens of Antarctic codfish *T. bernacchii* and *T. newne-*

*si*.<sup>29,30</sup> This suggests that the M-chain is the dominant subunit in spleen, and this pattern of expression is similar to mammalian L-chains, which is in line with the role of the organ as the major site of iron storage.<sup>6</sup> In terms of functional, biophysicochemical and localisational properties, the M-chains clearly shares more similarities with the mammalian L-chains than with the H-chains. Therefore, coupled with our phylogenetic analysis, we believe that the current naming of the M-chains in teleosts and L-chains in mammals do not accurately reflect their molecular phylogeny and can be misleading.

The origin of this naming confusion becomes apparent when the position of the first identified M-chain from the bullfrog, *R. catesbeiana* is scrutinised. The H-, M- and L-chains in *R. catesbeiana* were named after their electrophoretic mobility.<sup>28</sup> However, our phylogeny clearly shows that the M- and H-chains in *R. catesbeiana* are closely related paralogs, likely resulting from a recent lineage-specific duplication. While the *R. catesbeiana* H-chain (*R. catesbeiana* FerH), being the dominant subunit in red blood cells, might appear to be compatible with the canonical role of mammalian H-chains, it is clearly not orthologous to the mammalian and teleost H-chains.<sup>28</sup> In addition, the amphibian H-chain is represented in our phylogeny by the *X. tropicalis* 2 and *X. laevis* 3 proteins while the corresponding ortholog in *Rana* remains to be identified in the context of the absence of a complete genomic sequence. Our analysis also show that the annotated *Rana* L-chain (*R. catesbeiana* FerL) is highly divergent from rest of the vertebrate ferritin sequences and resulted in long branch attractions and unresolved topology and was, therefore, subsequently removed (Supplementary Figures 6, 7). Our analysis on the 5'-UTR of *R. catesbeiana* FerL using similar sequences from the GenBank EST database uncovered an IRE-like region, but it did not fulfil the stability criteria of canonical and experimental IREs. We believe that further characterisation of the various amphibian ferritin chains will be necessary to clarify their functions in iron metabolism.

Finally, we hypothesise that the differences in biochemical properties of the teleost M-chains and tetrapod L-chains could be attributed to the adaptations of both taxa to their environments during their evolutionary history; in particular, the transition of tetrapods into a terrestrial environment.<sup>98</sup> Small concentrations of aquatic iron are present in varying amounts depending on several environmental factors such as salinity, pH and the presence of inorganic sulphides.<sup>99</sup> Several studies have shown that there is a significant intake of aquatic iron across the gills of freshwater fish, and to a lesser degree, through intestinal absorption in marine fish that imbibe large quantities of seawater to offset integumental

water loss.<sup>99</sup> This is in contrast to most tetrapods that dwell primarily on land and whose only means of acquiring iron is through dietary intake.<sup>5</sup> Therefore, it would appear logical to assume that the tetrapod L-chains have evolved to mineralise iron more efficiently compared to its orthologous counterpart (M-chains) in teleosts. Consistent with the idea that ferritin iron-content would be modulated to compensate for the lack of access to biologically-available iron, ferritins of marine and brackish fish have been shown to have higher iron contents compared to freshwater fish.<sup>100</sup> Interestingly, the presence of various highly-divergent types of L/M-type ferritins in amphibians may reflect the need to adapt to different aquatic and terrestrial environments during their life cycles.

## Conclusions

Various ferritin subunits have been characterised in vertebrates, but there is a lack of proper annotation based on homology and protein function, especially from the non-mammalian vertebrates. In this study, we cloned and characterised two ferritin sequences from *L. calcarifer* as the H- and M-chains, respectively. Based on our extensive phylogenetic analysis, we show that the various ferritin chains evolved independently in the vertebrate lineage and all of them are equally related to the non-vertebrate ferritins. Subsequently, our favoured scenario supports a duplication event at the chordate/vertebrate split that resulted in a paralogous copy. Although caution must be taken due to the existence of some unresolved cases, the most plausible and supported scenario is that the fish M-chains and tetrapod L-chains are orthologous to each other, with the presence of various M/L-chains in amphibians coupled with large functional and sequential divergence causing previously erroneous interpretations. Our study highlights the importance of applying a detailed and extensive approach (such as detecting pseudogenes and the use of several outgroup sequences) in phylogenetic studies to clarify homology relationships within a gene family. Clearly, the sequencing of additional vertebrate genomes, in particular non-teleost fishes, will help to finally clarify these issues.

## References

- Meneghini R. Iron homeostasis, oxidative stress, and DNA damage. *Free Radic Biol Med* 1997;23:783-92.
- Le NT, Richardson DR. The role of iron in cell cycle progression and the proliferation of neoplastic cells. *Biochim Biophys Acta* 2002;1603:31-46.
- Lukianova OA, David SS. A role for iron-sulfur clusters in DNA repair. *Curr Opin Chem Biol* 2005;9:145-51.
- Valko M, Rhodes CJ, Moncol J, et al. Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem Biol Interact.* 2006;160:1-40.
- Hentze MW, Muckenthaler M, Andrews NC. Balancing acts: molecular control of mammalian iron metabolism. *Cell* 2004;117:285-97.
- Arosio P, Levi S. Cytosolic and mitochondrial ferritins in the regulation of cellular iron homeostasis and oxidative damage. *Biochim Biophys Acta* 2010;1800:783-92.
- Andrews SC. The ferritin-like superfamily: evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochim Biophys Acta* 2010;1800:691-705.
- Carrondo MA. Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *EMBO J* 2003;22:1959-68.
- Chiancone E. Dps proteins, an efficient detoxification and DNA protection machinery in the bacterial response to oxidative stress. *Rend Lincei Sci Fis Nat* 2008;19:261-70.
- Proudhon D, Wei J, Briat J, Theil EC. Ferritin gene organization: differences between plants and animals suggest possible kingdom-specific selective constraints. *J Mol Evol* 1996;42:325-36.
- Nichol H, Law JH, Winzerling JJ. Iron metabolism in insects. *Annu Rev Entomol* 2002;47:535-59.
- Hamburger AE, West Jr AP, Hamburger ZA, et al. Crystal structure of a secreted insect ferritin reveals a symmetrical arrangement of heavy and light chains. *J Mol Biol* 2005;349:558-69.
- Kim Y, Cho JH, Yoo OJ, Ahnn J. Transcriptional regulation and life-span modulation of cytosolic aconitase and ferritin genes in *C. elegans*. *J Mol Biol* 2004;342:421-33.
- Levi S, Arosio P. Mitochondrial ferritin. *Int J Biochem Cell Biol* 2004;36:1887-9.
- Missirlis F, Holmberg S, Georgieva T, et al. Characterization of mitochondrial ferritin in *Drosophila*. *Proc Natl Acad Sci USA* 2006;103:5893-8.
- Torti M, Torti SV. Regulation of ferritin genes and proteins. *Blood* 2002;99:3505-16.
- Thomson AM, Rogers JT, Leedman PJ. Iron-regulatory proteins, iron-responsive elements and ferritin mRNA translation. *Int J Biochem Cell Biol* 1999;31:1139-52.
- Rogers J. Ferritin translation by interleukin-6: the role of sequences upstream of the start codons of they heavy and light subunit genes. *Blood* 1996;87:2525-37.
- Theil EC. Coordinating responses to iron and oxygen stress with DNA and mRNA promoters: The ferritin story. *Biometals* 2007;20:513-21.
- Itoh K, Chiba T, Takahashi S, et al. An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. *Biochem Biophys Res Commun.* 1997;236:313-22.
- Lawson DM, Treffry A, Artymiuk PJ, et al. Identification of the ferroxidase centre in ferritin. *FEBS Lett* 1989;254:207-10.
- Levi S, Santambrogio P, Cozzi A, et al. The role of the L-chain in ferritin iron incorporation: studies of homo and heteropolymers. *J Mol Biol* 1994;238:649-54.
- Santambrogio P, Levi S, Arosio P, et al. Evidence that a salt bridge in the light chain contributes to the physical stability difference between heavy and the light human ferritins. *J Biol Chem* 1992;267:14077-83.
- Santambrogio P, Sonia L, Cozzi A, et al. Evidence that the specificity of iron incorporation into homopolymers of human ferritin L- and H-chains is conferred by the nucleation and ferroxidase centres. *Biochem J* 1996;314:139-44.
- Bou-Abdallah F, Papaefthymiou GC, Scheswohl DM, et al.  $\mu$ -1,2-peroxo-bridged di-iron(III) dimer formation in human H-chain ferritin. *Biochem J* 2002;364:57-63.
- Chasteen ND, Harrison PM. Mineralization in ferritin: an efficient means of iron storage. *J Struct Biol* 1999;126:182-94.
- Harrison PM, Arosio P. The ferritins: molecular properties, iron storage function and cellular regulation. *Biochim Biophys Acta* 1996;1275:161-203.
- Dickey LF, Sreedharan S, Theil EC, et al. Differences in the regulation of messenger RNA for housekeeping and specialized-cell ferritin. *J Biol Chem* 1987;262:7901-7.
- Mignogna G, Chiaraluze R, Consalvi V, et al. Ferritin from the spleen of the Antarctic teleost *Trematomus bernacchii* is an M-type homopolymer. *Eur J Biochem* 2002;269:1600-6.
- Giorgi A, Mignogna G, Bellapadrona G, et al. The unusual co-assembly of H- and M-chains in the ferritin molecule from the Antarctic teleosts *Trematomus bernacchii* and *Trematomus neunesi*. *Arch Biochem Biophys* 2008;478:69-74.
- Strozycki PM, Szymanski M, Szczurek A, et al. A new family of ferritin genes from *Lupinus luteus* - comparative analysis of plant ferritins, their gene structure, and evolution. *Mol Biol Evol* 2010;27:91-101.
- Kazuo M, Sumio S. Oligo-capping: a simple method to replace the eukaryotic mRNAs with oligoribonucleotides. *Gene* 1994;138:

- 171-4.
33. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer tracer using Phred. I. Accuracy assessment. *Genome Res* 1998;8:175-85.
  34. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 1999;41:95-8.
  35. Campillos M, Cases I, Hentze MW, Sanchez M. SIREs: searching for iron-responsive elements. *Nucleic Acids Res* 2010;38:W360-7.
  36. Flicek P, Aken BL, Ballester B, et al. Ensembl's 10th year. *Nucleic Acids Res* 2010;38:D557-62.
  37. Altschul SF, Gish W, Miller W, et al. Basic Local Alignment Search Tool. *J Mol Biol* 1990;215:403-10.
  38. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2007;2:953-71.
  39. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;31:3406-15.
  40. Sierzputowska-Gracz H, McKenzie RA, Theil EC. The importance of a single G in the hairpin loop of the iron-responsive element (IRE) in ferritin mRNA for structure: an NMR spectroscopy study. *Nucleic Acids Res* 1995;23:146-53.
  41. Laing LG, Hall KB. A model of the iron responsive element RNA hairpin loop structure determined from NMR and thermodynamic data. *Biochemistry* 1996;35:13586-96.
  42. Bailey TL, Boden M, Buske FA, et al. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res* 2009;37:W202-8.
  43. Huerta-Cepas, Dopazo H, Dopazo J, Gabaldón T. The human phylome. *Genome Biol* 2007;8:R109.
  44. Landan G, Graur D. Heads or tails: a simple reliability check for multiple sequence alignments. *Mol Biol Evol* 2007;24:1380-3.
  45. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792-7.
  46. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008;9:286-98.
  47. Subramaniam AR, Kaufmann M, Morgenstern B. DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. *Algorithms Mol Biol* 2008;3:6.
  48. Wallace IM, O'Sullivan O, Higgins DG, Notredame C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res* 2006;34:1692-9.
  49. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972-3.
  50. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007;24:1586-91.
  51. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 2000;17:32-43.
  52. Guindon S, Dufayard J, Lefort V, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307-21.
  53. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003;19:1572-4.
  54. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol* 2008;25:1307-20.
  55. Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* 2008;24:581-3.
  56. Loriot A, Boon T, De Smet C. Five new human cancer-germline genes identified among 12 genes expressed in spermatogonia. *Int J Cancer* 2003;105:371-6.
  57. Wang PJ, McCarrey JR, Yang F, Page DC. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 2001;27:422-6.
  58. Piccinelli P, Samuelsson T. Evolution of the iron-responsive element. *RNA* 2007;13:952-66.
  59. Gribaldo S, Philippe H. Ancient phylogenetic relationships. *Theor Popul Biol* 2002;61:391-408.
  60. Beck G, Ellis TW, Habicht GS, et al. Evolution of the acute phase response: iron release by echinoderm (*Asterias forbesi*) coelomocytes, and cloning of an echinoderm ferritin molecule. *Dev Comp Immunol* 2002;26:11-26.
  61. Durand J, Goudard F, Pieri J, et al. *Crassostrea gigas* ferritin: cDNA sequence analysis for two heavy chain type subunits and protein purification. *Gene* 2004;338:187-95.
  62. De Zoyza M, Lee J. Two ferritin subunits from disk abalone (*Haliotis discus discus*): cloning, characterization and expression analysis. *Fish Shellfish Immunol* 2007;23:624-35.
  63. Ke Y, Wu J, Leibold EA, et al. Loops and bulge/loops in iron-responsive element isoforms influence iron regulatory protein binding. *J Biol Chem* 1998;273:23637-40.
  64. Goforth JB, Anderson SA, Nizzi CP, Eisenstein RS. Multiple determinants within iron-responsive elements dictate iron regulatory protein binding and regulatory hierarchy. *RNA* 2010;16:154-69.
  65. Wasserman WW, Fahl WE. Functional antioxidant responsive elements. *Proc Natl Acad Sci USA* 1997;94:5361-6.
  66. Pietsch C, Chan JY, Torti FM, Torti SV. Nrf2 mediates the induction of ferritin H in response to xenobiotics and cancer chemopreventive dithiolethiones. *J Biol Chem* 2003;278:2361-9.
  67. Hintze KJ, Theil EC. DNA and mRNA elements with complementary responses to hemin, antioxidant inducers, and iron control ferritin-L expression. *Proc Natl Acad Sci USA* 2005;102:15048-52.
  68. Tsuji Y. JunD activates transcription of the human ferritin H gene through an antioxidant response element during oxidative stress. *Oncogene* 2005;24:7567-78.
  69. Rushmore TH, Morton MR, Pickett CB. The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity. *J Biol Chem* 1991;266:11632-9.
  70. Fujiwara KT, Kataoka K, Nishizawa M. Two new members of the maf oncogene family, mafK and mafF, encode nuclear b-Zip proteins lacking putative trans-activator domain. *Oncogene* 1993;8:2371-80.
  71. Kataoka K, Igarashi K, Itoh K, et al. Small Maf proteins heterodimerize with Fos and may act as competitive repressors of the NF-E2 transcription factor. *Mol Cell Biol* 1995;15:2180-90.
  72. Takagi Y, Kobayashi M, Li L, et al. MafT, a new member of the small Maf protein family in zebrafish. *Biochem Biophys Res Commun* 2004;320:62-9.
  73. Lawson MJ, Zhang L. Sexy gene conversions: locating gene conversions on the X-chromosome. *Nucleic Acids Res* 2009;37:4570-9.
  74. Campanella A, Rovelli E, Santambrogio P, et al. Mitochondrial ferritin limits oxidative damage regulating mitochondrial iron availability: hypothesis for a protective role in Friedreich ataxia. *Hum Mol Genet* 2009;18:1-11.
  75. Metzendorf C, Lind MI. *Drosophila* mitoferrin is essential for male fertility: evidence for role of mitochondrial iron metabolism during spermatogenesis. *BMC Dev Biol* 2010;10:68.
  76. Levi S, Rovida E. The role of iron in mitochondrial function. *Biochim Biophys Acta* 2009;1790:629-36.
  77. Park S, Gakh O, O'Neill HA, et al. Yeast frataxin sequentially chaperones and stores iron by coupling protein assembly with iron oxidation. *J Biol Chem* 2003;278:31340-51.
  78. Stehling O, Elsässer H, Brückel B, et al. Iron-sulfur protein maturation in human

- cells: evidence for a function of frataxin. *Hum Mol Genet* 2004;13:3007-15.
79. Yoon T, Cowan JA. Frataxin-mediated iron delivery to ferrochelatase in the final step of heme biosynthesis. *J Biol Chem* 2004;279:25943-6.
80. Nichol H, Gakh O, O'Neill HA, et al. Structure of frataxin iron cores: an X-ray absorption spectroscopic study. *Biochemistry* 2003;42:5971-6.
81. Campanella A, Isaya G, O'Neill HA, et al. The expression of human mitochondrial ferritin rescues respiratory function in frataxin-deficient yeast. *Hum Mol Genet* 2004;13:2279-88.
82. Smith AB. Rooting molecular trees: problems and strategies. *Biol J Linn Soc Lond* 1994;51:279-92.
83. Huelsenbeck JP, Rannala B. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst Biol* 2004;53:904-13.
84. Hillis DM, Bull JJ. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* 1993;42:182-1992.
85. Doaudy CJ, Delsuc F, Boucher Y, et al. Comparison of Bayesian and Maximum Likelihood bootstrap measures of phylogenetic reliability. *Mol Biol Evol* 2003;20:248-54.
86. Erixon P, Svennblad B, Britton T, Oxelman B. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst Biol* 2003;52:665-73.
87. Alfaro ME, Zoller S, Lutzoni F. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol Biol Evol* 2003;20:255-66.
88. Wilcox TP, Zwickl DJ, Heath TA, Hillis DM. Phylogenetic relationship of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol Phylogenet Evol* 2002;25:361-71.
89. Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H. Additional molecular support for the new chordate phylogeny. *Genesis* 2008;46:592-604.
90. Nagy L, Urban A, Örstandius L, et al. The evolution of autodigestion in the mushroom family Psathyrellaceae (Agaricales) inferred from Maximum Likelihood and Bayesian methods. *Mol Phylogenet Evol* 2010;57:1037-48.
91. Benton MJ. Phylogeny of the major tetrapod groups: morphological data and divergence dates. *J Mol Evol* 1990;30:409-24.
92. Georgieva T, Dunkov BC, Dimov S, et al. *Drosophila melanogaster* ferritin: cDNA encoding a light chain homologue, temporal and tissue specific expression of both subunit types. *Insect Biochem Mol Biol* 2002;32:295-302.
93. Escriva H, Manzon L, Youson J, Laudet V. Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. *Mol Biol Evol* 2002;19:1440-50.
94. Blomme T, Vandepoele K, De Bodt S, et al. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol* 2006;7:R43.
95. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 2005;3:e314.
96. Huerta-Cepas, Dopazo H, Dopazo J, Gabaldón T. The human phylome. *Genome Biol* 2007;8:R109.
97. Brunet FG, Crollius HR, Paris M, et al. Gene loss and evolutionary rates following whole-genome duplications in teleost fishes. *Mol Biol Evol* 2006;23:1808-16.
98. Long JA, Gordon MS. The greatest step in vertebrate history: a paleobiological review of the fish-tetrapod transition. *Physiol Biochem Zool* 2004;77:700-19.
99. Bury N, Grosell M. Iron acquisition by teleost fish. *Comp Biochem Physiol C Toxicol Pharmacol* 2003;135:97-105.
100. Geetha C, Deshpande V. Purification and characterization of fish liver ferritins. *Comp Biochem Physiol B Biochem Mol Biol* 1999;123:285-94.

Non-commercial use only