RESEARCH ARTICLE

Evolutionary Active Transposable Elements in the Genome of the Coelacanth

DOMITILLE CHALOPIN[1], SHAOHUA FAN[2,3], OLEG SIMAKOV[2,4], AXEL MEYER[2,3], MANFRED SCHARTL[5], AND JEAN-NICOLAS VOLFF[1]*

[1]Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, CNRS UMR 5242, Université Lyon 1, Lyon, France
[2]Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Konstanz, Germany
[3]Konstanz Research School Chemical Biology, University of Konstanz, Konstanz, Germany
[4]European Molecular Biology Laboratory, Heidelberg, Germany
[5]Department Physiological Chemistry, Biocenter, University of Wuerzburg, Wuerzburg, Germany

ABSTRACT

The apparent morphological stasis in the lineage of the coelacanth, which has been called a "living fossil" by many, has been suggested to be causally related to a slow evolution of its genome, with strongly reduced activity of transposable elements (TEs). Analysis of the African coelacanth showed that at least 25% of its genome is constituted of transposable elements including retrotransposons, endogenous retroviruses and DNA transposons, with a strong predominance of non-Long Terminal Repeat (non-LTR) retrotransposons. The coelacanth genome has been shaped by four major general bursts of transposition during evolution, with major contributions of LINE1, LINE2, CR1, and Deu non-LTR retrotransposons. Many transposable elements are expressed in different tissues and might be active. The number of TE families in coelacanth, but also in lungfish, is lower than in teleost fish, but is higher than in chicken and human. This observation is in agreement with the hypothesis of a sequential elimination of many TE families in the sarcopterygian lineage during evolution. Taken together, our analysis indicates that the coelacanth contains more TE families than birds and mammals, and that these elements have been active during the evolution of the coelacanth lineage. Hence, at the level of transposable element activity, the coelacanth genome does not appear to evolve particularly slowly. *J. Exp. Zool. (Mol. Dev. Evol.) 322B:322–333, 2014.* © 2013 Wiley Periodicals, Inc.

How to cite this article: Chalopin D, Fan S, Simakov O, Meyer A, Schartl M, Volff J-N. 2014. Evolutionary active transposable elements in the genome of the coelacanth. J. Exp. Zool. (Mol. Dev. Evol.) 322B:322–333.

The African coelacanth *Latimeria chalumnae* belongs to a lineage of lobe-finned (sarcopterygian) fish with fossils dating back to the Devonian period. Coelacanths have been considered to be extinct since 70 million years (Smith, '39). It was therefore a zoological sensation when in 1938 a living specimen was discovered in South Africa. Thereafter, fewer than 300 individuals have been caught accidentally mostly on the Comoran Island, but also along the Eastern coastline of Africa in Kenya, Tanzania, Mozambique, Madagascar, and South Africa. A second species, *Latimeria menadoensis*, was discovered in 1997 in Indonesia. The African and Indonesian coelacanths are the only known extant species of this once rather diverse lineage of lobe-finned fish. Some fossils show a remarkably high degree of conservation in morphology to the living specimen, suggesting an extended period of stasis in coelacanth phenotypic evolution (Smith, '56). The coelacanth genome has therefore been proposed to be a slowly evolving genome (Thomson, '91; Friedman and Coates, 2006; Amemiya et al., 2010; but also see Meyer and Wilson, '90; Meyer and Dolven, '92; Casane and Laurenti, 2013).

Both the lungfish and the coelacanth lineages hold key phylogenetic positions in the vertebrate phylogeny, as both have been considered to be the closest living relatives to tetrapods (reviewed in Meyer, '95; Zardoya and Meyer, '97a; Meyer and Zardoya, 2003). Divergence between coelacanth, lungfish, and tetrapods has been estimated to have occurred around 390 million years ago (Johanson et al., 2006). Different types of molecular datasets favored the lungfish over the coelacanth as the most closely related living lineage to tetrapods (Meyer and Wilson, '90; Meyer and Dolven, '92; Zardoya and Meyer, '96a,b, '97a,b; Zardoya et al., '98; Brinkmann et al., 2004; Takezaki et al., 2004; Amemiya et al., 2013). Although all previous analyses could rule out the coelacanth as the closest living relative to tetrapods, some studies could not distinguish between the lungfish and the lungfish + coelacanth hypotheses (reviewed in Meyer and Zardoya, 2003). Recently, the complete sequence of the *L. chalumnae* genome has yielded important new insights, for example, into the water-to-land transition, but also into the evolution of vertebrate genomes in general (Amemiya et al., 2013).

Here we focus on the analysis of transposable elements (TEs), one of the major components of eukaryotic genomes, in the coelacanth genome. TEs are repeated mobile genetic elements able to move within and between genomes. They have been considered for a long time as purely junk and selfish DNA, with only negative effects on host genes such as mutational inactivation and silencing (Deininger and Batzer, '99; Slotkin and Martienssen, 2007; Hollister and Gaut, 2009; Hancks and Kazazian, 2010). More recently, it has been proposed that TEs might also act as more positive drivers of genome evolution (Feschotte and Pritham, 2007; Böhne et al., 2008; Oliver and Greene, 2009; Pritham, 2009). Indeed, it has been shown that TEs can promote recombination and genomic rearrangements such as insertions, deletions, duplications, inversions, and translocations,

and therefore can be major contributors to genome dynamics and plasticity (Kazazian, 2004). They can also mobilize and duplicate host coding sequences, or can be "domesticated" by the host genome for the formation of new regulatory sequences, new exons and even new genes (Bejerano et al., 2006; Volff, 2006; Rebollo et al., 2012).

The classification of TEs is based on their transposition mechanism, depending on the use of an RNA transposition intermediate (class I: retrotransposons) or not (class II: DNA transposons) (Finnegan, '89; Wicker et al., 2007). Within classes, TEs are classified into orders and (super)families. Retrotransposons (class I) transpose via the reverse transcription of an RNA intermediate into cDNA. Based on their structure and the molecular phylogeny of their reverse transcriptase, retrotransposons can be divided into five major orders: LTR (long terminal repeat) retrotransposons and retroviruses, *Dictyostelium* intermediate repeat sequence (DIRS) elements, Penelope elements (PLE), LINEs, and SINEs (long and short interspersed nuclear elements, also called non-LTR retrotransposons) (Malik et al., '99; Eickbush and Jamburuthugoda, 2008). In contrast to autonomous protein-coding retrotransposons, SINEs are non-coding and parasitize the enzymatic machinery of other retrotransposons for their own transposition. DNA transposons (class II) transpose through a mechanism without RNA intermediate. This class includes the classical "cut-and-paste" terminal inverted repeat (TIR) transposon order, with double-stranded DNA cleavage and transposition catalyzed by a transposase. Other class II transposon orders such as Helitrons (rolling-circle transposons) and Mavericks/Polintons (self-synthesizing transposons) cut only one strand of DNA to transpose. A last order called Crypton uses a tyrosine recombinase for transposition like DIRS retrotransposons. Miniature inverted transposable elements (MITEs) are non-autonomous elements mobilized *in trans* by DNA transposons that encode an active transposase.

All known types of transposable elements have been found in vertebrate genomes (Volff et al., 2003; Feschotte and Pritham, 2007; Mandal and Kazazian, 2008; Kojima and Jurka, 2011; Huang et al., 2012). However, the distribution and abundance of many elements is patchy in vertebrates, with some types of TEs being present only in some lineages but absent from others. For example, many retrotransposon families detected in teleost genomes are absent from mammals (Volff et al., 2003). Among the 13 distinct families of non-LTR retrotransposons (Malik et al., '99; Ichiyanagi and Okada, 2008), only four (LINE1, LINE2, RTE, and Chicken Repeat-1 [CR1]) are common to mammals, birds, reptiles, amphibians, and fish (Ichiyanagi and Okada, 2008).

The CR1-like retrotransposon superfamily, which is subdivided into three families (CR1, LINE2, and Rex1/Babar) (Volff et al., 2000; Lovšin et al., 2001; Ichiyanagi and Okada, 2008), has been initially discovered and studied in chicken (Stumph et al., '81; Haché and Deeley, '88; Burch et al., '93; Vandergon and

Reitman, '94; Haas et al., '97; Wicker et al., 2005). CR1-like elements have been subsequently identified in other birds (Kaiser et al., 2007; Suh et al., 2012) as well as in reptiles (Nobuhisa et al., '98; Shedlock, 2006; Novick et al., 2009; Castoe et al., 2011), turtles (Kajikawa et al., '97), eutherian mammals and marsupials (Gentles et al., 2007; Mandal and Kazazian, 2008), monotremes (Warren et al., 2008), amphibians (Hellsten et al., 2010), lungfish (Sirijovski et al., 2005; Metcalfe et al., 2012), and ray-finned fish (Poulter et al., '99; Volff et al., 2000; Sugano et al., 2006). CR1-like elements are also present in urochordates, cephalochordates, and insects (Cañestro and Albalat, 2012). Hence, these retrotransposons probably played a significant role in shaping vertebrate genomes during evolution.

Does the coelacanth genome, which has been proposed to be a slowly evolving genome, contain active transposable elements? It has been suggested that phenotypic evolution has stalled in the

Table 1. Coverage of transposable element families in the coelacanth genome compared to other vertebrate genomes.

Transposable element	Genome coverage (%)						
Classes Orders Families	Fugu	Coelacanth	Lungfish	Salamander	Frog	Chicken	Human
Class I retrotransposons	Un	7.20	38.10	41.02	8.65	Un	28.71
Non–LTR retrotransposons	1.32	6.43	30.30	10.39	5.40	3.10	20.42
LINE1/Tx1	0.06	0.58	5.60	1.88	1.20	—	16.89
CR1–like	0.78	4.20	22.50	7.19	3.80	3.10	3.53
CR1	—	2.90	15.20	Un	3.80	Un	0.31
LINE2	0.53	1.30	7.20	Un		Un	3.22
Rex1/Babar	0.25	—	0.10	—		Un	—
RTE	0.39	0.40	1.70	0.76	—	—	—
R4	0.09	<0.01	Un	—	—	—	—
Jockey	—	<0.01	Un	0.03	—	—	—
Deu	Un	1.80	Un	Un	Un	—	—
MIR	Un	<0.01	Un	Un	0.01	—	2.20
Penelope	Un	0.16	0.50	0.51	0.90	—	—
DIRS	0.01	0.62	6.00	5.88	0.60	—	—
LTR retrotransposons	0.30	0.71	1.10	20.94	1.62	0.60	—
Gypsy	0.17	0.09	1.10	20.43	1.30	0.47	—
Ty1/Copia	0.01	—	—	0.04	0.02	—	—
BEL	0.02	—	—	—	0.30	—	—
Endogenous retroviruses	Un	0.20	0.20	3.30	0.13	Un	8.29
ERV1 (class I)	Un	0.19	Un	3.04	0.10	—	2.89
ERVK (class II)	Un	0.01	—	0.03	—	—	0.31
ERVL/Foamy (class III)	Un	<0.01	—	0.23	0.03	Un	1.44
Class II DNA transposons	Un	0.20	1.30	6.37	25	0.61	2.84
Tc-Mariner	Un	0.04	Un	0.48	4.70	0.51	1.15
hAT	Un	0.11	0.10	0.63	6.10	0.10	1.55
Harbinger	Un	0.02	0.20	2.96	4.70	—	—
PiggyBac	Un	—	—	—	1.30	—	0.02
Polinton/Maverick	Un	<0.01	Un	0.91	0.01	Un	—
Helitron	Un	<0.01	—	1.3	0.60	—	—
Total TEs	2.7%	25%	39.4%	47.52%	34.5%	4.3%	45%

Coelacanth data are a summary of Table S1. Values for other species are from the literature (Fugu, Chicken, and Human from Mandal and Kazazian (2008); Salamander (*Aneides flavipunctatus*) from Sun et al. (2012); Frog (*Xenopus tropcicalis*) from Hellsten et al. (2010); Lungfish from Metcalfe et al. (2012), and this analysis). Absence of TE families is shown by "—;" "Un" for unknown indicates that the family is present but no value is available from the literature. Families representing <0.001% of the genome in all five species are not shown. Total TE values include unclassified elements.

coelacanth due to a lack of intense intermittent activity by TE families (Oliver and Greene, 2009). Information currently available on TEs in coelacanth is scarce. Han and Worobey (2012) have identified endogenous copies of a foamy retrovirus in the coelacanth genome. In addition, potential cases of TE-derived regulatory and coding sequences have been reported (Bejerano et al., 2006; Smith et al., 2012). The recent sequencing of the genome of *L. chalumnae* (Amemiya et al., 2013) provides us with the opportunity to analyze TE content and activity at the genome-wide scale, and to better characterize TE evolution in one of the closest living "fish" relatives of tetrapods. We show here that the coelacanth has a relatively low repeat content but possesses more different types of TE families than mammals, and that, apparently, TEs have shaped the coelacanth genome by successive bursts of transposition during evolution.

RESULTS

TE Landscape in the Coelacanth Genome

Analysis of the coelacanth genome draft showed that approximately 30% of the genome is constituted by repeated sequences (Table S1), mainly transposable elements (25% of the genome). Non-LTR retrotransposons are the most abundant (Table 1). Almost all known families reported from the animal kingdom are present except Rex1/Babar and R2 (R2 elements are absent from all vertebrate species analyzed in this study). Major families of autonomous non-LTR retrotransposons in the coelacanth are CR1, LINE2, LINE1/TX1, and RTE, and the non-autonomous Deu SINE. Jockey and R4 elements were also detected, but at a lower copy number. LTR retrotransposons are also present but less diverse and only represented by Gypsy elements. This is in contrast to the situation in teleost fish, which show the highest LTR retrotransposon diversity among vertebrates (Volff et al., 2003). As in teleosts and human, the three classes of endogenous retroviruses were detected, with ERV1 being the most abundant. Teleost fish, coelacanth, salamanders, and mammals are the only lineages where foamy endogenous retroviruses have been described so far (Han and Worobey, 2012; Sun et al., 2012; Schartl et al., 2013). Finally, all orders of DNA transposons except Crypton were identified in the coelacanth genome. Particularly, three main TIR families (Harbinger, Tc-Mariner, and hAT) with many subfamilies constitute the larger fraction of DNA transposons in the genome. Helitrons and Polintons are also present but with a much lower copy number.

Compared to other vertebrate genomes, several types of TEs were observed in coelacanth, salamanders (Sun et al., 2012), and teleost fish, but were absent from chicken and human, including RTE, Jockey, Penelope, and DIRS retrotransposons (Table 1; Penelope is absent from Fugu, but present in other teleost species). Hence, these elements might have been lost in the amniote stem lineage. BEL retrotransposons were found in teleost fish and frog (Hellsten et al., 2010), but neither in coelacanth nor in salamanders, birds, and

mammals. This suggested multiple losses of these sequences in the salamander, coelacanth, and amniote lineages. Taken together, these results indicate that TE diversity is higher in coelacanth than in birds and mammals, similar to that observed in salamanders and frogs, and lower compared to teleost fish.

The relative age of the different TE families was estimated through Copy Divergence Analysis (CDA) using Jukes–Cantor distances between individual copies and their consensus sequence (Jukes and Cantor, '69; Lander et al., 2001). For each TE family, the consensus sequence provides an approximation of the sequence of the ancestral TE. Four main peaks were observed on the TE age plot (Fig. 1), suggesting that four major bursts of transposition occurred during the evolution of the coelacanth genome. In order to better understand the relative contribution of each major coelacanth TE family, the CR1, Deu SINE, LINE2, and LINE1 retrotransposons were also analyzed separately (Fig. 2; Table S1). LINE1 seemed to be the most ancient active family of transposable elements detected in coelacanth, which strongly contributed to the oldest burst of transposition (burst A on Figs. 1 and 2). The second most ancient burst (burst B) was mostly generated by CR1 and LINE2 elements. Contributors to the two more recent bursts were CR1, Deu SINE, and LINE2 for burst C, and CR1 and Deu SINE for the most recent burst D. The absence of significant contribution of LINE2 to burst D suggests that these elements became recently less active, probably explaining their low representation in the genome of extant coelacanths. Concomitant activity of CR1
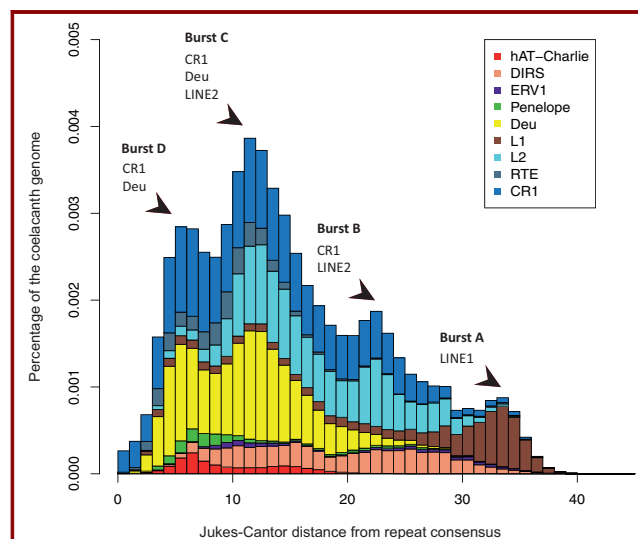

Figure 1. History of transposable element activity in the coelacanth genome through Copy Divergence Analysis. Jukes–Cantor distances to consensus sequences were calculated for the main TE families (frequencies >0.01% of the genome). Major families contributing to the four general bursts are indicated near black arrows.
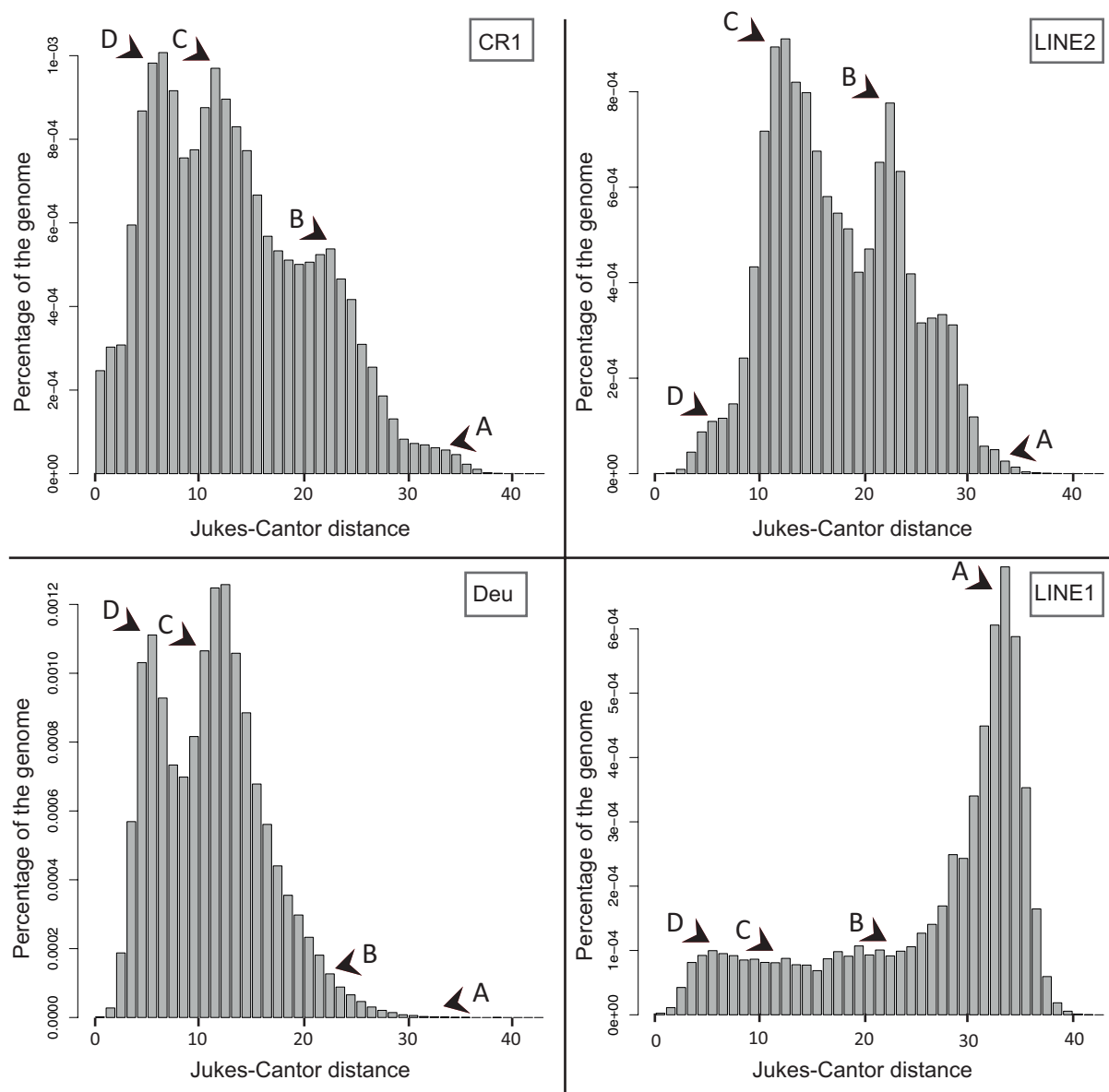
Figure 2. Sequence divergence for major TE families in the coelacanth genome through Copy Divergence Analysis. Jukes–Cantor distances were calculated to the consensus sequence of each TE family. The four major bursts during the coelacanth genome evolution are indicated by the black arrows.

(LINE) and Deu (SINE) families during bursts C and D might reflect the fact that Deu non-autonomous elements use CR1-encoded proteins for their transposition (Kajikawa et al., '97; Nishihara et al., 2006). Analysis of very similar TE sequences suggested that CR1 is currently the most active autonomous TE in the genome of extant coelacanths. TE analysis of the transcriptome of three different adult coelacanth tissues (testis, liver, and muscle) showed that the CR1, Deu SINE, DIRS, and RTE families are particularly highly expressed (see Forconi et al., personal communication).

Comparison of Transposable Elements in Coelacanth and Lungfish Genomes

Analysis of TE content in the genome of lungfish, another sarcopterygian fish lineage more related to tetrapods than to teleost fish, may provide in combination with the coelacanth interesting insights into TE evolution at the water-to-land transition in vertebrates. Due to the large size of lungfish species genomes (from 50 Gb for *Neoceratodus forsteri* to 130 Gb for *Protopterus aethiopicus*), it is currently almost impossible to

sequence and assemble them in order to perform whole genome analyses. However, some genomic (Metcalfe et al., 2012) and transcriptomic (Amemiya et al., 2013) data are available, allowing a first comparison between coelacanth and lungfish mobilomes.

Several TEs have been previously reported in lungfish, including the NfCR1 LINE (from CR1 family) and Lun1 SINE (from the Deu family) (Ogiwara et al., 2002). A recent global analysis based on partial genomic databases estimated that TEs might constitute about 40% of the lungfish genome (Metcalfe et al., 2012). We compared the data obtained by Metcalfe et al. (2012) for lungfish with our results on the coelacanth genome (Table 1). In addition we completed the analysis of the TE repertoire in lungfish through the analysis of a transcriptomic database from liver, kidney, and testis (Amemiya et al., 2013).

Probably linked to its larger size, the lungfish genome contains a higher proportion of TEs than the coelacanth (Metcalfe et al., 2012). However, from a qualitative point of view, the TE content in lungfish and coelacanths is rather similar: both show a strong contribution of non-LTR retrotransposons and a much lower abundance of LTR retrotransposons and DNA transposons. In both species, the most abundant families of non-LTR retrotransposons are the CR1 and L2 families, followed by LINE1/TX1 and RTE. We also detected R4 and Jockey retrotransposons in the transcriptome. Penelope, Gypsy, and DIRS elements are found in both lungfish and coelacanth. However, it should be noted that DIRS elements showed a 10× higher genome-wide abundance in lungfish compared to coelacanth. Beside hAT and Harbinger, which are the major DNA transposon families, Tc1 (widespread among vertebrates) and Polinton sequences were identified though transcriptome analyses. All identified TE families were found in the transcriptome, suggesting that most of them are expressed and probably active in lungfish. As observed in the coelacanth (Forconi et al., personal communication), CR1 and DIRS were particularly highly expressed in lungfish (Amemiya et al., 2013; data not shown).

TE Coverage and Genome Size in Vertebrates

The comparison between genome size and TE coverage in different organisms has previously revealed a general positive trend: larger genomes tend to be composed of a higher percentage of TEs (Hancock, 2002; Vieira et al., 2002; Biémont, 2008). In vertebrates, this correlation is also observed and well-illustrated by the smallest genome of the Fugu (2.7% of TEs) and the largest sequenced genome of the opossum (52% of TEs) (Fig. 3). Compared to other vertebrate genomes of similar size, the coelacanth TE content is relatively low, for example, approximately only half of the estimated fraction in the human genome. The gigantic genomes of lungfish and salamanders (from 14 to 120 Gb; Sun et al., 2012) are not shown on the graph, since no genome draft sequences are available so far. TE coverage has been estimated to about 40% for lungfish and 25–47% for salamanders, values that are large but in range of those observed for large mammalian
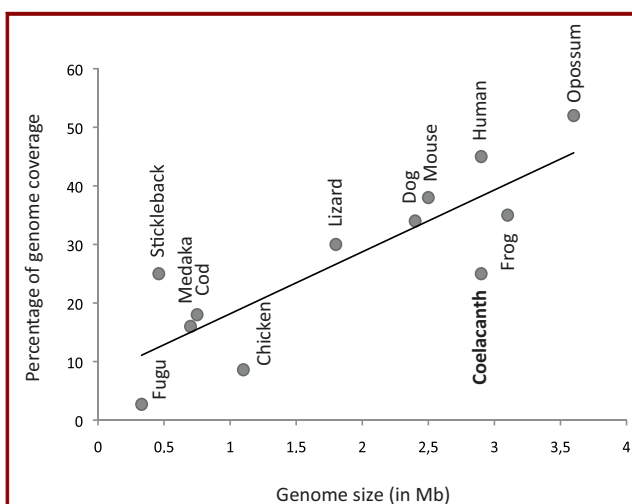

Figure 3. Genome size and transposable element coverage in different vertebrate species. Genome sizes were plotted against percentages of TE coverage for 12 vertebrate species. The black line represents the linear regression of the plot. Data were obtained from the literature for Fugu (Aparicio et al., 2002), stickleback (Jones et al., 2012), medaka (Kasahara et al., 2007), cod (Star et al., 2011), chicken (International Chicken Genome Sequencing Consortium, 2004), lizard (Alföldi et al., 2011), dog (Lindblad-Toh et al., 2005), mouse (Waterston et al., 2002), human (Lander et al., 2001), frog (Hellsten et al., 2010), and opossum (Gentles et al., 2007).

genomes. Additional unknown mechanisms beside TE activity might be involved in genome expansion in these species.

Evolutionary History of the CR1-like Superfamily in Vertebrates and Other Animals

The CR1-like superfamily (CR1, LINE2, and Rex1/Babar) is represented at a high copy number and expressed in both coelacanth and lungfish (Sirijovski et al., 2005; Metcalfe et al., 2012). The evolutionary origin of this superfamily, as well as the phylogenetic relationships between its different families have not been completely resolved (Ichiyanagi and Okada, 2008; Novikova and Blinov, 2009). CR1-like elements have been considered to be vertically transmitted in vertebrates (Malik et al., '99), but there is so far no systematic analysis of their evolutionary history in this group of animals.

All available CR1-like sequences from Repbase (Jurka, 2000; Jurka et al., 2005) and other public databases were added to coelacanth and lungfish sequences to reconstruct the molecular phylogeny of the CR1, LINE2, and Rex1/Babar families. Jockey, which is considered as the closest family (Malik et al., '99), as well as the LINE1 and RTE families were used as outgroups (Fig. 4). Available genome draft sequences were screened for CR1-like elements.
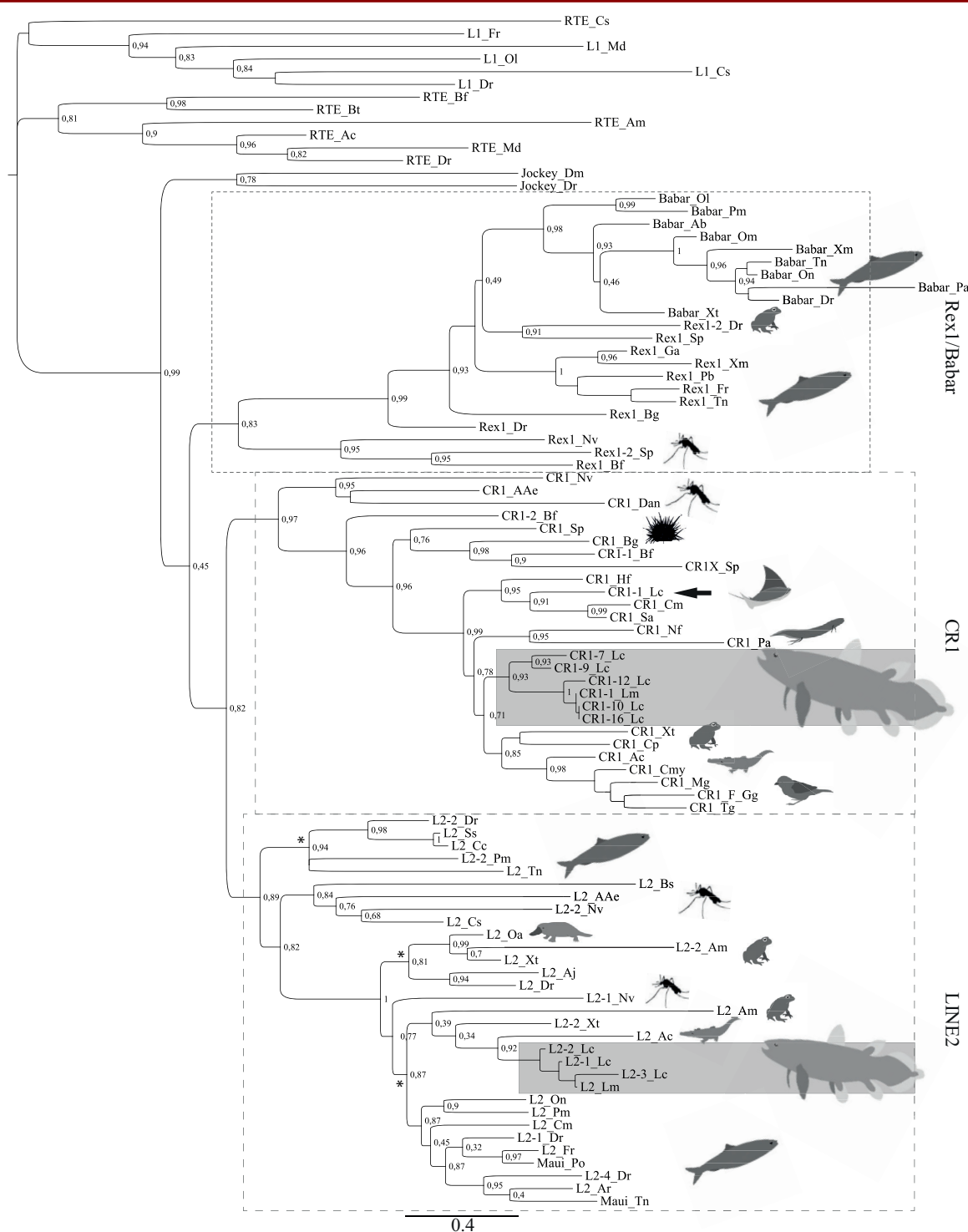
Figure 4. Molecular phylogeny of the CR1-like retrotransposon superfamily in animals. Phylogenetic tree was built from a reverse transcriptase alignment using the maximum likelihood method (calculation on 206 sites). Scale bar represents the number of substitutions per site. Abbreviations for species names are summarized in Table S2. Group of coelacanth sequences are shown by gray boxes. The black arrow shows a single coelacanth sequence in the CR1 family. Asterisks highlight vertebrate branches in the LINE2 family.

The CR1 family branch reflects species phylogeny, with insect sequences at the base of the group, followed by gastropods, echinoderms, and chordates. The CR1 family is present in tetrapods, lobe-finned fish (including coelacanth and lungfish), and cartilaginous fish but might be absent from ray-finned fish and lampreys (Fig. 5). The coelacanth probably contains at least two subfamilies of CR1 since their sequences are not monophyletic. Indeed, most of the sequences are grouped (gray box in Fig. 4) but one is placed in a more basal position among cartilaginous fish sequences (arrow in Fig. 4).

In vertebrates, the LINE2 family appears to be split in three different branches (asterisks in Fig. 4). The first branch supports only ray-finned fish sequences, and might be therefore fish-specific. The second branch contains diverse vertebrate species but no coelacanth sequence. Finally, the last one contains the Maui retrotransposons described in teleosts (Poulter et al., '99). All coelacanth sequences are located within the "Maui" subfamily, suggesting a poor diversity of LINE2 subfamilies in the coelacanth in contrast to what is observed in teleosts. LINE2 elements were not found in birds, turtles, and gastropods (Fig. 5). Both CR1 and
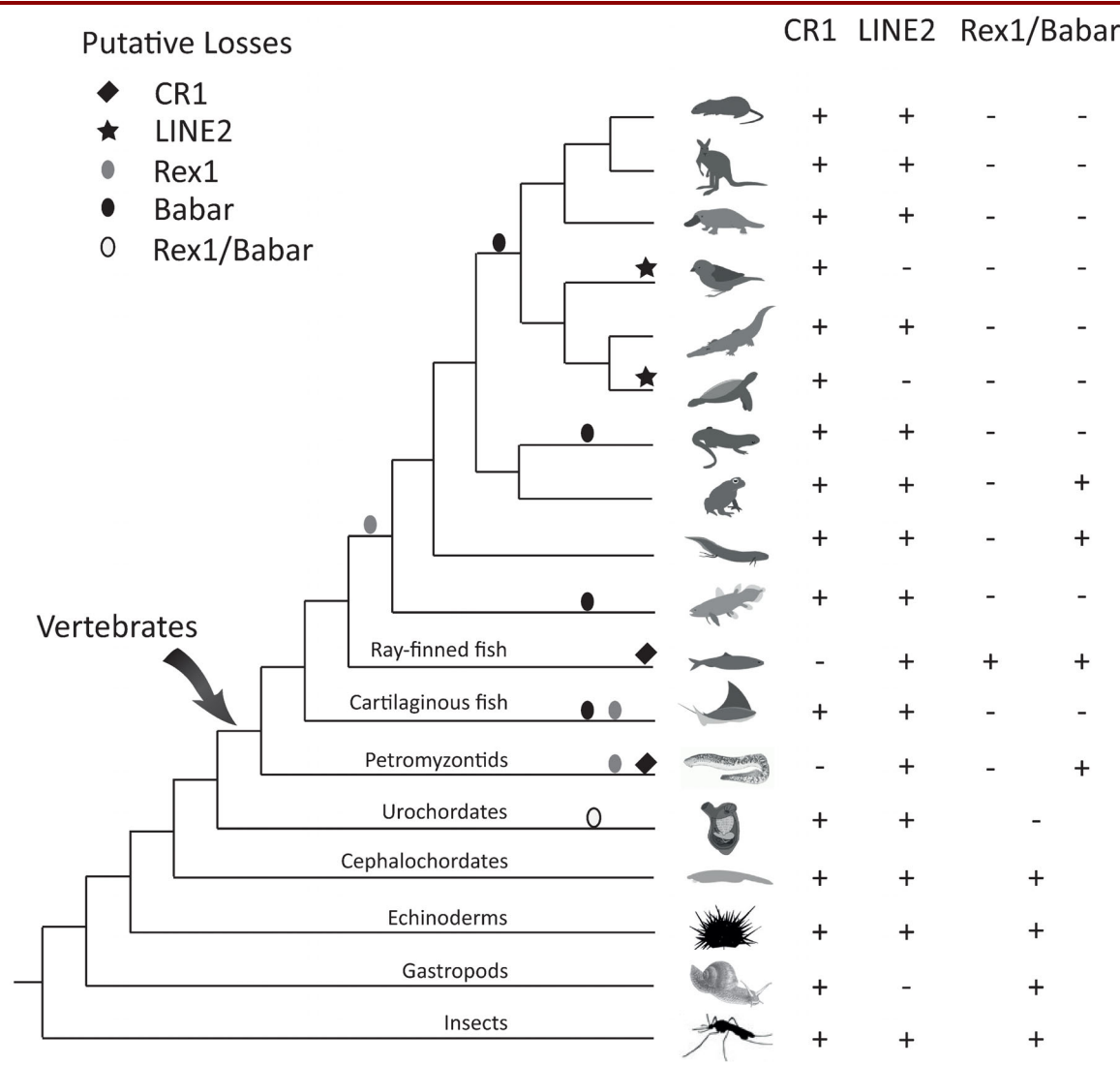

Figure 5. Evolutionary history of the three CR1-like retrotransposon families CR1, LINE2 and Rex1/Babar in animals. This scenario is based on the vertical transmission of elements, even if horizontal transfer cannot be excluded. Presence (+) and absence (−) of the different families are represented on the right side of the figure. Small icons summarized on the left side show loss of the different families.

LINE2 families are present in eutherians and marsupials, but mainly as inactive copies. Therefore, their sequences were not included in the phylogeny.

Rex1/Babar, the third CR1-like retrotransposon family, is subdivided into two subfamilies called Rex1 and Babar in vertebrates (Fig. 5). One ancestral Rex1/Babar group is detected in insects, gastropods, echinoderms, and cephalochordates. No Rex1/Babar element was found in urochordates. In vertebrates, Babar is present in lamprey, ray-finned fish, lungfish, and amphibians, while Rex1 is only found in ray-finned fish (Fig. 5). Neither Rex1 nor Babar could be identified in the coelacanth genome assembly.

Hence, CR1-like is an ancient animal retrotransposon superfamily subdivided into three main families that interestingly show a patchy distribution and multiple lineage-specific losses during vertebrate evolution. The CR1 family, which is widely distributed in animals, has been lost in the ray-finned fish lineage, but was maintained in sarcopterygians. LINE2 is present in ray-finned fish and most sarcopterygians but seems to have been eliminated in turtles and birds. Rex1 is only found in ray-finned fish. Babar is present in ray-finned fish, lungfish, and amphibians, but has been lost in the remaining tetrapods. Babar distribution suggests its elimination in the coelacanth lineage.

CONCLUSION

We have analyzed transposable element content and evolution in the coelacanth genome. At least 25% of this genome is constituted by transposable elements, a relatively low value compared to some other sarcopterygian genomes. TE abundance is approximately twice as high in the human as in the coelacanth genome—both genomes having a similar genome size. At least one type of TEs, the Babar family, has been lost specifically in the coelacanth lineage. In contrast, genome-wide analysis of element diversity showed that the coelacanth contains more different families of transposable elements than the genome of chicken and human. Our study might underestimate both TE content and diversity in the coelacanth genome, since analyses were performed only on the genome draft. Additional TEs and even additional TE families might be present in unassembled parts of the genome.

The analysis of the genome of the coelacanth supports an active history in terms of TE evolution. Copy Divergence Analysis (see Methods Section) suggested that the coelacanth genome has been shaped by at least four general bursts of transposition during its evolutionary history. LINE1 was probably the most ancient active TE family, which might have been supplanted by CR1, Deu SINE, and LINE2 retrotransposons later on. CR1 and other elements are actively transcribed and possibly still active in the genome. Hence, the apparent morphological stasis during coelacanth evolution might not be due to reduced transposable element activity, as proposed by some authors (Oliver and Greene, 2009). Such an observation challenges the possible

relationship between TE diversity and species richness proposed for instance in the teleost fish lineage (Volff, 2005). On the other hand, the designation of coelacanth as a "living fossil" has been called into question, with comparative analysis suggesting that morphological stability of coelacanths is not supported by paleontological evidence (for review Casane and Laurenti, 2013 and references therein).

Previous comparative analyses of TE content in fish and mammals have suggested that many families of transposable elements have been lost in the tetrapod lineage (Volff et al., 2003). Availability of the coelacanth genome, as well as data from lungfish, allowed us to refine this analysis using two species delimiting the water-to-land transition in the sarcopterygian lineage. Coelacanth and lungfish were found to present similar TE landscape and activity, even if some minor differences were detected. Interestingly, the number of TE families in coelacanth and lungfish, which was similar to that found in salamander and frog, was lower than in teleost fish, but higher than in chicken and human. Some major groups of TE might have been eliminated in a common ancestor of birds and mammals (RTE, Penelope, Jockey, and DIRS retrotransposons) and others from tetrapods after the split from coelacanth and lungfish (R4 retrotransposons). Predominant TE families are different in different lineages: CR1 and LINE2 are one of the most abundant families from basal chordates to lungfish and also in monotremes, while Gypsy retrotransposons are the main elements in salamander, hAT transposons in frog and LINE1 retrotransposons in marsupials and eutherians.

The analysis of the coelacanth genome led to a better understanding of the distribution and evolution of the CR1-like superfamily, which is widely represented among vertebrates and other animals. Of the three families constituting the CR1-like superfamily, two (CR1 and LINE2) are present in the coelacanth genome, while the third one (Rex1/Babar) is absent. Our analysis showed multiple lineage-specific losses during evolution. Particularly, we could show that Rex1/Babar is absent from most tetrapods, and that the CR1 family, which is present in tetrapods and more divergent animals, is absent from ray-finned fish. Hence, CR1 might represent the first example of a major family of TEs that was eliminated in ray-finned fish, but maintained in mammals, other tetrapods, and other lobe-finned fish.

METHODS

Repeat Content Analyses
The TE library (Amemiya et al., 2013) was constructed from the assembly both manually and automatically using RepeatScout with an lmer size of 16 (Price et al., 2005) and was manually curated. We annotated TEs based on three methods: first, RepeatMasker version 3.3.0 using Repbase version 14.11, second, TBlastX against Repbase version 14.11, and third, BlastX against a custom non-redundant collection of transposable elements

protein. Then, the coelacanth genome was masked by Repeat-Masker in a "sensitive mode". Coverage and number of copies were determined after genome masking from output files using a home-made script. Missing families were manually searched by Blast analysis on the assembly.

Copy Divergence Analysis based on Jukes–Cantor Distances

The percentage differences between identified TE copies in the genome and the consensus sequences in the TE library were extracted from the RepeatMasker file (.out file), and converted to Jukes–Cantor distance (Jukes and Cantor, '69) by using $d = -(3/4)\log_e(1-(4/3)p)$, where p is the proportion of the different nucleotide sites between the copies of TE and the consensus sequence.

CR1-like Superfamily Phylogenetic Analyses

Vertebrate nucleotide sequences belonging to the CR1-like retrotransposon superfamily were retrieved using Repbase Update (Jurka et al., 2005) as well as through systematic BLAST analysis of public databases accessible from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Ensembl (http://www.ensembl.org/Multi/blast-view) servers, resulting in a set of 844 vertebrate nucleotide sequences. For the coelacanth genome, 219 consensus sequences affiliated to the CR1-like superfamily (from 67 to 6,263 nucleotides in length) were identified in the coelacanth repeat library (Amemiya et al., 2013). These sequences are derived from elements showing at least 80% of nucleotide identity. Sequences were translated into proteins using Softberry from Molquest 2.3.3 (Softberry, Inc.). Only representative good quality translated sequence matching the reverse transcriptase domain were kept for phylogenetic analysis. Invertebrate CR1-like sequences were also included in the phylogeny as well as Jockey, RTE, and LINE1 sequences as outgroups. Reverse transcriptase sequences were aligned using Muscle (Edgar, 2004), resulting in a 206 sites alignment. Phylogenetic tree was reconstructed by Maximum Likelihood with default aLRT (non-parametric branch support) using PhyML (Gouy et al., 2010). The molecular phylogeny presented in Figure 4 includes only representative CR1-like family sequences for different animal lineages. Presence/absence of each CR1-like family was systematically assessed in all genomes analyzed through BLAST analysis using representative sequences as queries.

ACKNOWLEDGMENTS

LITERATURE CITED

Alföldi J, Di Palma F, Grabherr M, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477:587–591.

Amemiya CT, Powers TP, Prohaska SJ, et al. 2010. Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. Proc Natl Acad Sci USA 107:3622–3627.

Amemiya CT, Alföldi J, Lee AP, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature 496:311–316.

Aparicio S, Chapman J, Stupka E, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297:1301–1310.

Bejerano G, Lowe CB, Ahituv N, et al. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441:87–90.

Biémont C. 2008. Genome size evolution: within-species variation in genome size. Heredity (Edinb) 101:297–298.

Böhne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff JN. 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. Chromosome Res 16:203–215.

Brinkmann H, Venkatesh B, Brenner S, Meyer A. 2004. Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. Proc Natl Acad Sci USA 101:4900–4905.

Burch JB, Davis DL, Haas NB. 1993. Chicken repeat 1 elements contain a pol-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. Proc Natl Acad Sci USA 90:8199–8203.

Castoe TA, Hall KT, Guibotsy Mboulas ML, et al. 2011. Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. Genome Biol Evol 3:641–653.

Cañestro C, Albalat R. 2012. Transposon diversity is higher in amphioxus than in vertebrates: functional and evolutionary inferences. Brief Funct Genomics 11:131–141.

Casane D, Laurenti P. 2013. Why coelacanths are not 'living fossils': a review of molecular and morphological data. Bioessays 35:332–338.

Deininger PL, Batzer MA. 1999. Alu repeats and human disease. Mol Genet Metab 67:183–193.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.

Eickbush TH, Jamburuthugoda VK. 2008. The diversity of retro-transposons and the properties of their reverse transcriptases. Virus Res 134:221–234.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–368.

Finnegan DJ. 1989. Eukaryotic transposable elements and genome evolution. Trends Genet 5:103–107.

Friedman M, Coates MI. 2006. A newly recognized fossil coelacanth highlights the early morphological diversification of the clade. Proc Biol Sci 273:245–250.

Gentles AJ, Wakefield MJ, Kohany O, et al. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. Genome Res 17:992–1004.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224.

Haas NB, Grabowski JM, Sivitz AB, Burch JB. 1997. Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames. Gene 197:305–309.

Haché RJ, Deeley RG. 1988. Organization, sequence and nuclease hypersensitivity of repetitive elements flanking the chicken apoVLDLII gene: extended sequence similarity to elements flanking the chicken vitellogenin gene. Nucleic Acids Res 16:97–113.

Han GZ, Worobey M. 2012. An endogenous foamy-like viral element in the coelacanth genome. PLoS Pathog 8:e1002790.

Hancks DC, Kazazian HH. 2010. SVA retrotransposons: evolution and genetic instability. Semin Cancer Biol 20:234–245.

Hancock JM. 2002. Genome size and the accumulation of simple sequence repeats: implications of new data from genome sequencing projects. Genetica 115:93–103.

Hellsten U, Harland RM, Gilchrist MJ, et al. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. Science 328:633–636.

Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19:1419–1428.

Huang CR, Burns KH, Boeke JD. 2012. Active transposition in genomes. Annu Rev Genet 46:651–675.

Ichiyanagi K, Okada N. 2008. Mobility pathways for vertebrate L1, L2, CR1, and RTE clade retrotransposons. Mol Biol Evol 25:1148–1157.

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

Johanson Z, Long JA, Talent JA, Janvier P, Warren JW. 2006. Oldest coelacanth, from the early Devonian of Australia. Biol Lett 2:443–446.

Jones FC, Grabherr MG, Chan YF, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. p 21–132.

Jurka J. 2000. Repbase Update: a database and an electronic journal of repetitive elements. Trends Genet 16:418–420.

Jurka J, Kapitonov VV, Pavlicek A, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467.

Kaiser VB, van Tuinen M, Ellegren H. 2007. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. Mol Biol Evol 24:338–347.

Kajikawa M, Ohshima K, Okada N. 1997. Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. Mol Biol Evol 14:1206–1217.

Kasahara M, Naruse K, Sasaki S, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719.

Kazazian HH Jr. 2004. Mobile elements: drivers of genome evolution. Science 303:1626–1632.

Kojima KK, Jurka J. 2011. Crypton transposons: identification of new diverse families and ancient domestication events. Mob DNA 2:12.

Lander ES, Linton LM, Birren B, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438:803–819.

Lovšin N, Gubensek F, Kordis D. 2001. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. Mol Biol Evol 18:2213–2224.

Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16:793–805.

Mandal PK, Kazazian HH Jr. 2008. SnapShot: vertebrate transposons. Cell 135:192–192. e1.

Meyer A. 1995. Molecular evidence on the origin of tetrapods and the relationships of the coelacanth. Trends Ecol Evol 10:111–116.

Meyer A, Dolven SI. 1992. Molecules, fossils, and the origin of tetrapods. J Mol Evol 35:102–113.

Meyer A, Wilson AC. 1990. Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J Mol Evol 31:359–364.

Meyer A, Zardoya R. 2003. Recent advances in the (molecular) phylogeny of vertebrates. Annu Rev Ecol Syst 34:311–338.

Metcalfe CJ, Filée J, Germon I, Joss J, Casane D. 2012. Evolution of the Australian lungfish (*Neoceratodus forsteri*) genome: a major role for CR1 and L2 LINE elements. Mol Biol Evol 29:3529–3539.

Nishihara H, Smit AF, Okada N. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res 16:864–874.

Nobuhisa I, Ogawa T, Deshimaru M, et al. 1998. Retrotransposable CR1-like elements in crotalinae snake genomes. Toxicon 36:915–920.

Novick PA, Basta H, Floumanhaft M, McClure MA, Boissinot S. 2009. The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. Mol Biol Evol 26:1811–1822.

Novikova OS, Blinov AG. 2009. Origin, evolution, and distribution of different groups of non-LTR retrotransposons among eukaryotes. Genetika 45:149–159.

Ogiwara I, Miya M, Ohshima K, Okada N. 2002. V-SINEs: A new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. Genome Res 12:316–324.

Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. Bioessays 31:703–714.

Poulter R, Butler M, Ormandy J. 1999. A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons. Gene 227:169–179.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. Bioinformatics 1:i351–i358.

Pritham EJ. 2009. Transposable elements and factors influencing their success in eukaryotes. J Hered 100:648–655.

Rebollo R, Romanish MT, Mager DL. 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet 46:21–42.

Schartl M, Walter RB, Shen Y, et al. 2013. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. Nature Genet 45:567–572.

Shedlock A. 2006. Phylogenomic investigation of CR1 LINE diversity in reptiles. Syst Biol 55:902–911.

Sirijovski N, Woolnough C, Rock J, Joss JM. 2005. NfCR1, the first non-LTR retrotransposon characterized in the Australian lungfish genome, *Neoceratodus forsteri*, shows similarities to CR1-like elements. J Exp Zool B Mol Dev Evol 304:40–49.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8:272–285.

Smith JLB. 1939. A living fish of Mesozoic type. Nature 143:455–456.

Smith JLB. 1956. Old fourlegs: the story of the coelacanth. publisher. London: Longmans, Green & Co.

Smith JJ, Sumiyama K, Amemiya CT. 2012. A living fossil in the genome of a living fossil: Harbinger transposons in the coelacanth genome. Mol Biol Evol 29:985–993.

Star B, Nederbragt AJ, Jentoft S, et al. 2011. The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210.

Stumph WE, Kristo P, Tsai MJ, O'Malley BW. 1981. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. Nucleic Acids Res 9:5383–5397.

Sugano T, Kajikawa M, Okada N. 2006. Isolation and characterization of retrotransposition-competent LINEs from zebrafish. Gene 365:74–82.

Suh A, Kriegs JO, Donnellan S, Brosius J, Schmitz J. 2012. A universal method for the study of CR1 retroposons in nonmodel bird genomes. Mol Biol Evol 29:2899–2903.

Sun C, Shepard DB, Chong RA, et al. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol Evol 4:168–183.

Takezaki N, Figueroa F, Zaleska-Rutczynska Z, Takahata N, Klein J. 2004. The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of forty-four nuclear genes. Mol Biol Evol 21:1512–1524.

Thomson KS. 1991. Living fossil: the story of the coelacanth. New York: WW Norton.

Vandergon TL, Reitman M. 1994. Evolution of chicken repeat 1 (CR1) elements: evidence for ancient subfamilies and multiple progenitors. Mol Biol Evol 11:886–898.

Vieira C, Nardon C, Arpin C, Lepetit D, Biémont C. 2002. Evolution of genome size in Drosophila. Is the invader's genome being invaded by transposable elements? Mol Biol Evol 19:1154–1161.

Volff JN, Körting C, Schartl S. 2000. Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. Mol Biol Evol 17:1673–1684.

Volff JN, Bouneau L, Ozouf-costaz C, Fischer C. 2003. Diversity of retrotransposable elements in compact pufferfish genomes. Trends Genet 19:674–678.

Volff JN. 2005. Genome evolution and biodiversity in teleost fish. Heredity (Edinb) 94:280–294.

Volff JN. 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28:913–922.

Warren WC, Hillier LW, Marshall Graves JA, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. Nature 453:175–183.

Waterston RH, Lindblad-Toh K, Birney E, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.

Wicker T, Robertson JS, Schulze SR, et al. 2005. The repetitive landscape of the chicken genome. Genome Res 15:126–136.

Wicker T, Sabot F, Hua-Van A, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982.

Zardoya R, Meyer A. 1996a. The complete nucleotide sequence of the mitochondrial genome of the African lungfish, *Protopterus dolloi*, supports its phylogenetic position as a close relative of land vertebrates. Genetics 142:1249–1263.

Zardoya R, Meyer A. 1996b. Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene. Proc Natl Acad Sci USA 93:5449–5454.

Zardoya R, Meyer A. 1997a. Molecular phylogenetic information on the identity of the closest living relatives(s) of land vertebrates. Naturwissenschaften 84:389–397.

Zardoya R, Meyer A. 1997b. The complete DNA sequence of the mitochondrial genome of the "living fossil" the coelacanth *Latimeria chalumnae*. Genetics 146:995–1010.

Zardoya R, Cao Y, Hasegawa M, Meyer A. 1998. Searching for the closest living relative(s) of tetrapods through analyses of mitochondrial and nuclear DNA data sets. Mol Biol Evol 15:506–517.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.