Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2011, Article ID 680673, 8 pages
doi:10.1155/2011/680673

Research Article

The Roles and Evolutionary Patterns of Intronless Genes in Deuterostomes

Ming Zou,[1,2] Baocheng Guo,[3,4] and Shunping He[1]

[1] *The key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China*
[2] *Institute of Hydrobiology, Graduate University of the Chinese Academy of Sciences, Beijing 100039, China*
[3] *Institute of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland*
[4] *The Swiss Institute of Bioinformatics, Quartier Sorge-Batiment Genopode, 1015 Lausanne, Switzerland*

Correspondence should be addressed to Shunping He, heshunping@gmail.com

Received 23 July 2010; Revised 13 April 2011; Accepted 22 June 2011

Academic Editor: J. Peter W. Young

Genes without introns are a characteristic feature of prokaryotes, but there are still a number of intronless genes in eukaryotes. To study these eukaryotic genes that have prokaryotic architecture could help to understand the evolutionary patterns of related genes and genomes. Our analyses revealed a number of intronless genes that reside in 6 deuterostomes (sea urchin, sea squirt, zebrafish, chicken, platypus, and human). We also determined the conservation for each intronless gene in archaea, bacteria, fungi, plants, metazoans, and other eukaryotes. Proportions of intronless genes that are inherited from the common ancestor of archaea, bacteria, and eukaryotes in these species were consistent with their phylogenetic positions, with more proportions of ancient intronless genes residing in more primitive species. In these species, intronless genes belong to different cellular roles and gene ontology (GO) categories, and some of these functions are very basic. Part of intronless genes is derived from other intronless genes or multiexon genes in each species. In conclusion, we showed that a varying number and proportion of intronless genes reside in these 6 deuterostomes, and some of them function importantly. These genes are good candidates for subsequent functional and evolutionary analyses specifically.

1. Introduction

Most eukaryotic genes are interrupted by one or more noncoding sequences called introns, and intronless genes are a characteristic feature of prokaryotes. However, researches on intronless genes in eukaryotes have been reported over the past few decades [1–4]. Many human genes, like G protein-coupled receptor genes, are intronless [5] and the human genome report identified 901 predicted intronless genes [6]. Recently, Tay et al. found that many single-copy primate-specific human transcriptional units are single exon [7]. Moreover, Yang et al. found that species-specific genes in *Arabidopsis*, *Oryza,* and *Populus* are enriched with intronless genes [8]. A retrogene, which is formed by homologous recombination between the genomic copy of a gene and an cDNA [9], is also considered to be intronless, and it has been reported that many retrogenes exist in eukaryotic genomes [10–12]. Intronless genes in eukaryotes, because of their prokaryotic architecture, provide interesting datasets for comparative genomics and evolutionary studies. Studying these genes can help to understand the evolutionary patterns of related genes and genomes. As a result, systematical researches on intronless genes in many species from mammals to plants have been reported [13–18]. Several databases of these single exon genes, such as SEGE [19] and Genome SEGE [20], have been set up and are of important use for evolutionary and functional studies. However, former evolutionary researches on intronless genes have usually been limited to 1 to 2 species and studies within a phylogenetic framework are rare. With the development of sequencing technology, more and more complete genomes have been sequenced and annotated, which makes comprehensive comparative analysis on intronless genes possible. The present study was designed to identify and analyse intronless genes in

6 deuterostomes, sea urchin (*Strongylocentrotus purpuratus*), sea squirt (*Ciona intestinalis*), zebrafish (*Danio rerio*), chicken (*Gallus gallus*), platypus (*Ornithorhynchus anatinus*), and human (*Homo sapiens*), which were selected because of their pivotal phylogenetic positions. We compared the functions and conservation of these genes between and within species in an attempt to gain some evolutionary meaningful insights.

2. Materials and Methods

2.1. Data Source of Intronless Genes. The annotated genomes (GenBank Flat File Format) of sea urchin, sea squirt, zebrafish, chicken, platypus, and human were downloaded from the NCBI ftp server (ftp://ftp.ncbi.nih.gov/genomes/, 10 Jun 2009). Using a customized Perl script, we extracted protein sequences for all the intron and intronless genes from each annotated genome. During our processing, a gene was classified as intron-containing if the "CDS" line in the FEATURES contains a "join"; otherwise, it was classified as an intronless gene. Proteins that encoded by mitochondrial genomes were removed. To avoid any ambiguity, proteins encoded by genes which have the symbol "<" or ">" in their annotation ("<" indicates partial on the 5′ end and ">" indicates partial on the 3′ end) were also discarded.

2.2. Functional Assignment and Category. ProtFun is an online procedure designed to produce *ab initio* predictions of protein functions from sequences and combines 14 different sequence-based functional prediction methods. ProtFun queries a large number of other feature prediction servers to obtain information on various posttranslation and localisation aspects of the protein to predict protein function, rather than relying on sequence similarity compared with other protein function prediction procedures [21, 22]. Therefore, functional assignments of intronless genes in our study were done with the webserver ProtFun (http://www.cbs.dtu.dk/services/ProtFun/) and sequences were clustered according to their cellular roles and gene ontology (GO) categories.

2.3. Distribution, Conservation, and Paralogue Identification of Intronless Genes. Genes (both intronless and intron-containing genes) in archaea, bacteria, fungi, plants, metazoans, and other eukaryotes homologous with our intronless genes (BLAST score more than 100), were determined on the basis of sequence similarity using BLink (BLAST Link), which is a tool that displays the precomputed results of BLAST searches that have been completed for every protein sequence in the Entrez proteins data domain [24] and is available at NCBI.

CD-HIT is a program for clustering the entries in a large protein database according to sequence identity (with a high threshold of identity). CD-HIT can remove redundant sequences and generate a database of only the representatives [25]. To determine the conservative intronless genes among the 6 deuterostome species in this study, we clustered all of our intronless genes using CD-HIT. In order to determine the relationships among these intronless genes, we clustered them and identified nonredundant intronless genes in each genome. We also clustered intron-containing genes to produce nonredundant multiexon genes in each genome. We clustered these nonredundant intronless genes with nonredundant multiexon genes in the same genome and produced a list of corresponding intronless and intron-containing genes to determine the relationships between intronless and intron-containing genes. All these data handling were done with CD-HIT.

3. Results and Discussion

3.1. Intronless Genes in Deuterostomes. Sea urchin, sea squirt, zebrafish, chicken, platypus, and human were selected to represent the major groups of deuterostomes and the intronless genes in their genomes were identified. Gi number and protein sequence for each intronless gene in each species were obtained from processing their annotated genomes. As a result, there are abundant intronless genes in each of the 6 deuterostome genomes. The numbers of intronless genes in each species is given in Table 1 and details are given in supplementary material online at doi:10.1155/2011/680673. Among the selected species, human has the maximum number of intronless genes (6229) and platypus has the least (930). We can see the maximum one is nearly seven times the number of the least one. However, the difference among numbers of intronless genes in sea urchin (2482), zebrafish (2169), chicken (1659), and sea squirt (1448) is not significant and these numbers should increase and be more accurate when their well-annotated genomes are available. Since a few previous studies reported a bit lower numbers of the number of intronless gene [16, 20], we compared protein numbers (encoded by intron and intronless genes) from Ensemble (Table 1) with ours, and found the former was always larger. Compared to their numbers, proportions that intronless genes are accounting for total genes do not differ significantly, and the maximum one is about twice the number of the least one (Table 1). In fact, former researches reported that 11109, 5846, and 5085 intronless genes reside in rice, *Arabidopsis* and mouse genomes, accounting for 19.9%, 21.7%, and 18.9% genes correspondingly [13]. Given that the total gene numbers and annotation qualities between species are different, these data may indicate that although the number of intronless genes varied significantly between species, the proportions that they account for total genes are nearly constant. However, the number and percentages of intronless genes do not correlate with their genome sizes ($P > 0.6$, $|r| < 0.3$, Spearman's test). The human genome has the largest number of intronless genes, which might be due to the following reasons. Firstly, human has the most complete expression data, which could result in more annotated genes compared with other species during the genome annotation process. Secondly, the human genome has many more retrogenes compared with other species [26, 27]. Thirdly, duplications of intronless genes are common in the human genome (see later). Plenty of intronless genes exist in the 6 deuterostomes indicating they may play important roles during deuterostome evolution. Earlier, Jain et al. found that

Table 1: The *C*-values and statistics for genes (intron and intronless) in each species.

Species	Sea-Urchin	Sea-Squirt	Zebrafish	Chicken	Platypus	Human
C-value (pg)	0.89	0.20	1.75	1.25	3.06	3.5
N	—	19858	40585	22194	26836	88237
I	2482 (8.6)	1448 (11.2)	2169 (8.6)	1659 (10.2)	930 (7.9)	6229 (16.7)
NR*	676	1263	856	1029	516	2290
NRI*	8792	8502	10620	9502	7840	12823
*R***	621 (92)	110 (9)	274 (32)	156 (15)	86 (17)	1321 (58)
*C***	212 (31)	191 (15)	269 (31)	186 (18)	133 (26)	665 (29)

C: values are obtained from http://www.genomesize.com/.
N: number of proteins encoded by intron and intronless genes, obtained from Ensemble (http://www.ensembl.org/index.html).
I: number of intronless genes in each genome, numbers in parentheses are percentages of intronless genes account for total genes.
NR: number of nonredundant intronless gene clusters.
NRI: number of nonredundant intron gene clusters.
R: number of nonredundant clusters that represent more than one intronless gene.
C: number of clusters that represent both a nonredundant intronless and a nonredundant intron-containing gene.
*Clustered using CD-Hit (identity = 0.3).
**Numbers in parentheses are percentages they account for all nonredundant intronless gene clusters.

intronless genes have a strong bias towards encoding shorter proteins [13]. Here we testified that the average length of intronless genes is significantly shorter than multiexon genes in all the selected species ($P < 0.001$, Mann-Whitney Test). The average length for intronless genes in sea urchin, sea squirt, zebrafish, chicken, platypus, and human is 341.75 bp, 389.34 bp, 378.78 bp, 259.53 bp, 294.43 bp, 241.26 bp, and for intron genes is 530.59 bp, 540.58 bp, 553.04 bp, 541.31 bp, 528.11 bp, and 503.31 bp, respectively.

Among the selected species, chromosomes were well assembled in human, chicken, and zebrafish. To study the distribution of intronless genes in each selected genome, we counted the numbers of intronless genes on each of their chromosomes (Figure 1). Spearman's test showed that the number of intronless genes is significantly correlated with the length of their chromosomes in human ($P < 0.001$, $r = 0.721$) and chicken ($P < 0.001$, $r = 0.712$). The correlation may be also significant in zebrafish ($P = 0.119$, $r = 0.320$) given that nonparametric tests have less "power" to detect a significant difference. Therefore, we proved that the distributions of intronless genes in human and chicken (and maybe in zebrafish) are stochastic, just like previous studies in mouse, rice, and *Arabidopsis* [13, 18]. However, several clusters of intronless genes exist in certain chromosomes and some of these clusters have been reported. For example, the olfactory receptor gene clusters on human chromosome 17 and odorant receptor genes in the zebrafish genome [28, 29].

3.2. Functional Assignment of Intronless Genes. It has been shown that the distribution of intronless human genes across molecular function categories is nonrandom [17]. In order to study the molecular function categories of intronless genes in the 6 selected species, their cellular roles and GO categories were predicted using ProtFun (available via webserver http://www.cbs.dtu.dk/services/ProtFun/). Figure 2 shows the distribution of intronless genes among each cellular role in 6 species. As in plants, intronless genes that functionally belong to translation and energy metabolism are the commonest in most species, followed by the cell envelope and amino acid biosynthesis [13]. Furthermore, in these 6 deuterostomes, transport and binding, followed by regulatory functions and central intermediary metabolism, are also well represented compared with other function categories. The percentage of intronless genes with the same cellular role among the total intronless genes varies significantly between species (Figure 2). For example, 7% of intronless genes are transported and binding in sea squirt is significantly fewer than in other species. The number of cellular roles, such as amino acid biosynthesis and central intermediary metabolism, are quite similar in sea urchin, sea squirt, zebrafish, and human. However, this is not the case for chicken or platypus. GO categories can be assigned to more than 70% of intronless genes except in sea squirt (which is more than 60%), and the distribution of genes according to each GO category is shown in Figure 3. As in plants, proteins associated with the GO category growth factor, transcription regulation, transport, immune response and structural proteins are overrepresented in these species [13]. Furthermore, proteins associated with the GO category transcription, which might be different between plants and animals, are well represented in deuterostomes. The percentage of total intronless genes that proteins with a certain GO category, such as growth factor and transporter proteins, varied significantly among these species. According to their cellular roles and GO categories, the functional category distribution of intronless genes in each selected genome is very similar to those reported for rice and *Arabidopsis* [13]. This result might indicate that biological mechanisms related to intronless genes are common in the biological kingdom. On the basis of earlier work and this analysis, we concluded that most plant and deuterostome intronless genes have the same characteristics, but deuterostomes still have some lineage-specific and species-specific functional intronless genes.

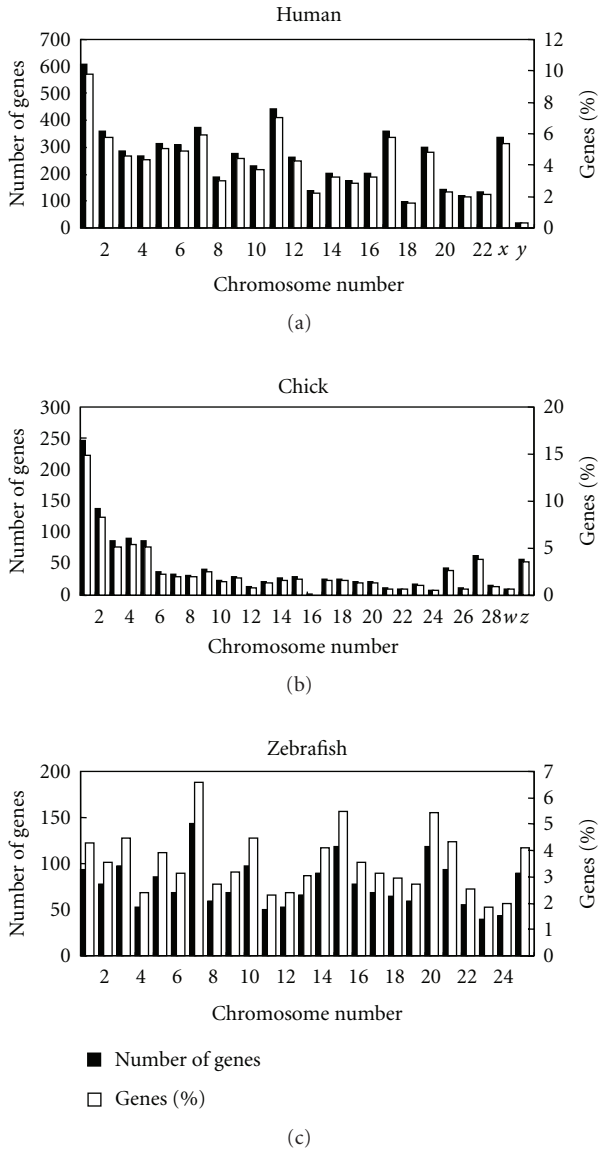

Figure 1: The numbers of intronless genes on each chromosome in human, chicken, and zebrafish. Both numbers and percentages are shown.

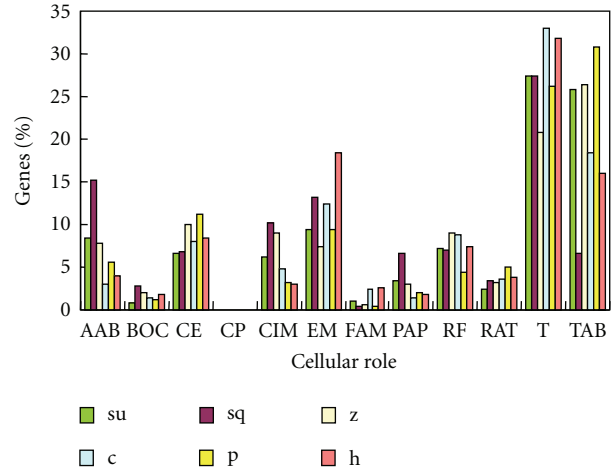

Figure 2: Distribution of intronless genes among different cellular roles in each species. Su: sea urchin; sq, sea squirt; z: zebrafish; c: chicken; p: platypus; h: human. AAB: amino acid biosynthesis; BOC: biosynthesis of cofactors; CE: cell envelope; CP: Cellular processes; CIM: Central intermediary metabolism; EM: energy metabolism; FAM: fatty acid metabolism; PAP: purines and pyrimidines; RF: regulatory functions; RAT: replication and transcription; T: translation; TAB: transport and binding.

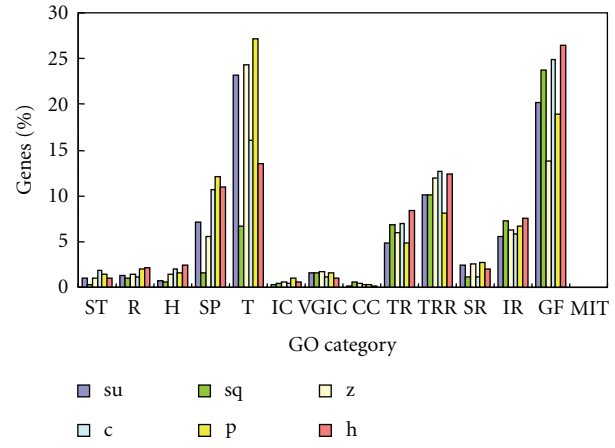

Figure 3: Distribution of intronless genes among different kinds of gene ontology (GO) categories in each species. su: sea urchin; sq: sea squirt; z: zebrafish; c: chicken; p: platypus; h: human. ST: signal transducer; R: receptor; H: hormone; SP: structural protein; T: transporter; IC: ion channel; VGIC: voltage-gated ion channel; CC: cation channel; TR: transcription; TRR: transcription regulation; SR: stress response; IR: immune response; GF: growth factor; MIT: metal ions transport.

3.3. Taxonomic Distribution. To study the evolutionary patterns of intronless genes in major taxonomic groups, we used BLink, a tool that displays the precomputed results of BLAST searches for every protein sequence from the entrez proteins data domain [24], to determine the evolutionarily conserved proteins among different taxonomic groups (archaea, bacteria, fungi, metazoans, plants, and other eukaryotes). The results of intronless gene clustering on the basis of homology with each taxonomic group are given in Table 2, and this will change as more genome sequences become available. We divided these genes into 7 types of combination according to their conservation among archaea (A), bacteria (B), and eukaryote (E) and the distributions are shown in Figure 4. Majority of intronless genes in each species that have homologues only in eukaryotes (E) suggested that most intronless genes emerged after the eukaryotes diverged from prokaryotes. Another important category of intronless genes is ABE, in which intronless genes are conserved in all major biological kingdoms, and these genes are considered to be functionally important and evolved slowly [30]. Intronless genes belonging to ABE account for 39% of the total intronless genes in sea squirt and 30% in sea urchin, which together

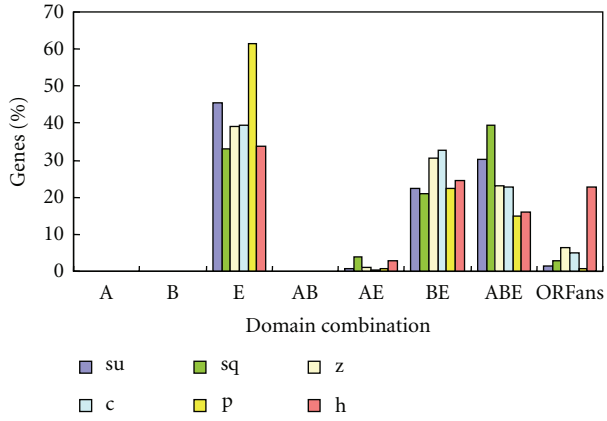

Figure 4: Distribution of intronless genes specific to different taxonomic group combinations for each species. su: sea urchin; sq: sea squirt; z: zebrafish; c: chicken; p: platypus; h: human. A: archaea; B: bacteria; E: eukaryote; AB: archaea and bacteria; AE: archaea and eukaryote; BE: bacteria and eukaryote; ABE: archaea, bacteria and eukaryote; ORFans: homologs not found in other organisms.

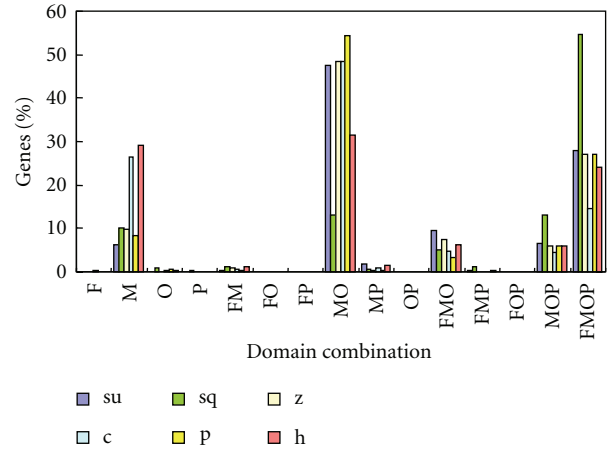

Figure 5: Distribution of eukaryote-specific intronless genes specific to different eukaryotic taxonomic group combinations for each species. su: sea urchin; sq: sea squirt; z: zebrafish; c: chicken; p: platypus; h: human. F: fungi; M: metazoans; O: other eukaryotes; P: plants; FM: fungi and metazoans; FO: fungi and other eukaryotes; FP: fungi and plants; MO: metazoans and other eukaryotes; MP: metazoans and plants; OP: other eukaryotes and plants; FMO: fungi, metazoans and other eukaryotes; FMP: fungi, metazoans and plants; FOP: fungi, other eukaryotes and plants; MOP: metazoan, other eukaryotes and plants FMOP: fungi, metazoans, other eukaryotes and plants.

form the first class. The second class contains zebrafish (23%), chicken (22%), and the third class includes human (16%) and platypus (14%). Given their phylogenetic positions, the first class is more primitive than the second class, which is more primitive than the third class. These data show that higher percentage of intronless genes in primitive species are inherited from the common ancestor of archaea, bacteria, and eukaryotes than in higher species. More than 20% of intronless genes are conserved in bacteria and eukaryotes (BE) in each species, but less than 5% are conserved in archaea and eukaryotes (AE). This could be because archaea have lost more homologues with eukaryotes than bacteria, or because bacteria have obtained more. Moreover, the percentage of genes conserved in bacteria and eukaryotes that account for total intronless genes is significantly higher in zebrafish and chicken than that in other species, suggesting that these 2 species have a greater percentage of intronless genes inherited from the common ancestor of bacteria and eukaryotes. No gene is conserved in archaea or/and bacteria except one human gene in bacteria and this might be an example of lateral gene transfer (LGT) from bacteria to human.

More than 30% of intronless genes are eukaryote specific in all these species, especially in platypus (61.6%). To investigate their distributions in eukaryotic groups, we divided these proteins according to their homogeneity in fungi (F), metazoans (M), other eukaryotes (O), and plants (P) and formed 15 types of combination (Figure 5). Generally, majority of genes have homologues in the combination MO (metazoans and other eukaryotes) in each species except in sea squirt, in which only 13% of eukaryote-specific intronless genes are of this kind. Less than 10% of genes are metazoan-specific (M) in many species, but in chicken and human there are 26% and 29%, respectively. Genes conserved in fungi, metazoans, other eukaryotes, and plants (FMOP), including histones and ribosomal proteins, were thought to be very conservative because they are essential for the survival of all eukaryotes [30]. The number of FMOP genes is very similar in sea urchin (314), sea squirt (263) and zebrafish (228) but the percentage of total eukaryote-specific intronless genes is much greater in sea squirt (54.7%) than that in sea urchin (27.8%) and zebrafish (26.9%). Except those cases mentioned above, very few genes are conserved in other taxonomic combinations. Moreover, some genes in some species have homologues in fungi, plants or other eukaryotes but not in metazoans. These genes might be examples of lateral gene transfer (LGT) between eukaryotes, which has been demonstrated recently [31, 32]. Since fungi and plants diverged from metazoan ahead of other eukaryotes, the distribution pattern of eukaryote-specific intronless genes in these species can be explained by that much more homologs have been lost in fungi and plants plus lots of others have been obtained after their divergence. However, lots of essential genes (FMOP) were still preserved. Therefore, the distribution pattern of eukaryote-specific intronless genes and the gain and loss patterns in this work are in accord with earlier reports [13, 15, 33].

The predicted cellular role of each kind of combination is shown in the supplementary material. Amino acid biosynthesis, cell envelope, energy metabolism, translation, transport and binding are usually well represented. Furthermore, the distribution of basic functional categories, such as amino acid biosynthesis, energy metabolism, and translation, are overrepresented in intronless genes conserved in all major biological kingdoms (ABE) or all eukaryotic groups (FMOP) compared to others.

Table 2: Number of intronless genes with homologous genes in other taxonomic groups.

Taxonomic group	Sea urchin	Sea squirt	Zebrafish	Chicken	Platypus	Human
Archaea	764	625	524	385	143	1184
Bacteria	1303	873	1161	920	345	2524
Fungi	1492	1149	1067	817	413	2513
Plants	1480	1209	1070	809	441	2617
Metazoans	2447	1403	2027	1575	918	4807
Other eukaryotes	2337	1343	1897	1388	867	4055
ORFans	35	39	136	80	7	1416

ORFans [23]: homologues not found in other species.

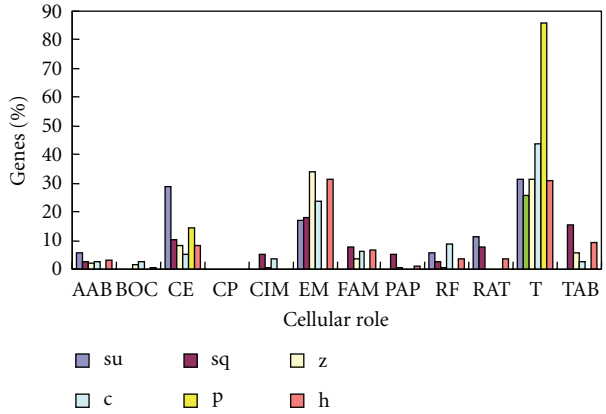

Figure 6: Distribution of ORFans according to their functional categories in each species. The description of functional categories and species is the same as that given for Figure 2.

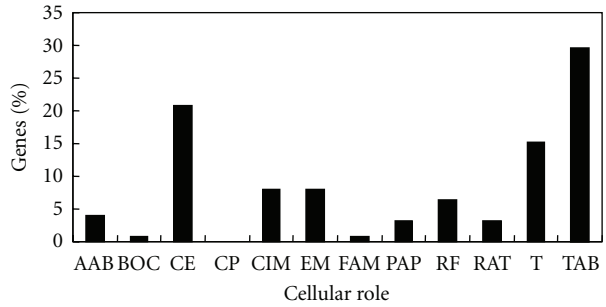

Figure 7: Functional distribution of conserved intronless genes in vertebrates. The description of functional categories is same as that given in Figure 2.

3.4. ORFans. The protein sequences that have no homologue in other species are termed ORFans [23]. These proteins could be responsible for some species-specific characteristics, and most of these proteins might have evolved faster than others [34]; in fact, they are part of the most interesting genome content. Thus, it is important to experimentally characterize these proteins or use more sensitive bioinformatic approaches to understand their roles and functions [15]. We found very few ORFans in these species except in human (Table 2), about 22.7% of whose intronless genes are ORFans, and this might be due to their complexities because they are viviparous and mammalian. Most of these proteins are annotated as hypothetical; however, majority of ORFans in all species, except platypus, have mRNA or EST supports when we checked their annotations. Thus, most of these ORFans might not be misannotated. Figure 6 shows the predicted cellular role distribution of ORFans in each species. It is interesting to note that translation and energy metabolism are the most frequently represented cellular roles in these species. The pattern is similar to earlier reports of plants and human [13, 15], suggesting that most species-specific intronless genes in plants and animals have the same functions and even the components of basic cellular machinery might evolve to perform species-specific functions in all these species [13, 15]. Moreover, we found the cellular role of cell envelope is well represented in sea urchin ORFans and more than half of the intronless genes that have the cellular role of fatty acid metabolism are ORFans.

3.5. Conserved Intronless Genes in Deuterostomes. To examine the conservation of intronless genes in deuterostomes, we clustered them together for these 6 species using CD-Hit (identity = 0.3). Only 6 nonredundant sequences (NR) were shared by these species, and they might perform pivotal functions in deuterostomes. The predicted cellular roles were translation for 4 NRs, regulatory function for one NR, and energy metabolism for one NR. The GO categories of these NRs were associated with transcription regulation, growth factor, and transport. When we compared the shared NRs between any two species, we found that sea urchin and sea squirt have less than 100 shared NRs with other species, but the number was more than 200 between any two vertebrate species (data not shown), suggesting that significantly more intronless genes are shared by vertebrates than those shared by deuterostomes. As expected, we found 125 NRs shared by vertebrates, and Figure 7 shows the distribution of the predicted functions of these NRs. Most of these proteins are involved in basic cellular processes, such as transport and binding, cell envelope, and translation, and these genes could be one of the important reasons for the emergence of vertebrates.

3.6. Paralogues of Intronless Genes. Intronless genes in eukaryotic genomes have many origins other than inheritance

from ancient prokaryotes, such as duplication (whole genome duplication or tandem duplication) of existing intronless genes and retroposition of intron-containing genes (retroduplicated genes). Also, there is evidence that ancient intronless genes were the origin of multiexon genes [35, 36]. To investigate these latter patterns, we clustered intronless genes and nonredundant intronless genes with multiexon genes using CD-Hit (identity = 0.3), and the results were shown in Table 1. It shows that about 92% of sea urchin nonredundant clusters have more than 1 intronless gene and the value is 58% in human, 32% in zebrafish, and only 9% in sea squirt. These data suggest that most intronless genes may originate from other intronless genes in sea urchin, human, and zebrafish, but much fewer intronless genes have the same origin in other species. Table 1 also shows the frequency of correspondence between nonredundant intronless genes and nonredundant intron-containing genes. About 30% of nonredundant intronless gene clusters have corresponding nonredundant intron-containing genes in each species, but in sea squirt and chicken, the proportion is only 15% and 18%, respectively. This might be due to the activity of LINE retrotransposable elements in their genomes. Active LINE retrotransposons that can reversibly transcribe polyadenylated mRNAs are thought to be the main reason for the emergence of retrogenes [37, 38]. More than 20% of the human genome is composed of LINE retrotransposable elements [39], and many studies have suggested a high rate of retroposition in human [26, 40, 41], which might result in the emergence of intronless genes. In chicken, only about 8% of the genome is comprised of the CR1 (chicken repeat 1) [42] and this kind of LINE-1 is not thought to reversibly transcribe polyadenylated mRNAs [42]. Majority of intronless genes that have intronless or intron-containing homologs are associated with cellular roles transport and binding, cell envelope, and translation. It has long been believed that duplicated genes (including retroduplicated genes) provide material for the evolution of genes with new functions [43], but there is evidence that retrogenes function as their parent genes during the spermatogenesis X chromosome inactivation of meiosis in mammals [12] and in the fruit fly [10]. Thus, the selective advantage of retention of these duplicated intronless genes might be that these genes can evolve new functions or help to buffer crucial functions similar to earlier reports on duplicated genes in angiosperms [44].

4. Conclusion

Both this and earlier studies indicate that the evolutionary patterns of intronless genes among deuterostomes, as well as between deuterostomes and plants, have many common characteristics and might be appropriate for all major eukaryote kingdoms. However, there are still some lineage-specific and species-specific characteristics on the evolution of intronless genes, and this might be one of the reasons for the existence of biodiversity in this world. As more genome sequences are sequenced and more exhaustive and accurate genes are annotated, the evolutionary patterns of intronless genes will become clearer, providing insights into understanding the evolutionary mechanisms underlying gene or genome evolution in eukaryotes.

Acknowledgments

The authors are thankful to four anonymous reviewers and M. Yu for their critical reading of this manuscript and helpful comments and suggestions that greatly improved the paper. This work was supported by a Grant from the Major State Basic Research Development Program of China (973 Program, no. 2007CB411601).

References

[1] K. B. Gatermann, A. Hoffmann, G. H. Rosenberg, and N. F. Kaufer, "Introduction of functional artificial introns into the naturally intronless ura4 gene of Schizosaccharomyces pombe," *Molecular and Cellular Biology*, vol. 9, no. 4, pp. 1526–1535, 1989 (English).

[2] B. Bhandari, W. J. Roesler, K. D. DeLisio, D. J. Klemm, N. S. Ross, and R. E. Miller, "A functional promoter flanks an intronless glutamine synthetase gene," *Journal of Biological Chemistry*, vol. 266, no. 12, pp. 7784–7792, 1991 (English).

[3] A. V. Makeyev, A. N. Chkheidze, and S. A. Liebhaber, "A set of highly conserved RNA-binding proteins, alpha CP-1 and alpha CP-2, implicated in mRNA stabilization, are coexpressed from an intronless gene and its intron-containing paralog," *Journal of Biological Chemistry*, vol. 274, no. 35, pp. 24849–24857, 1999 (English).

[4] A. Sugiyama, K. Noguchi, C. Kitanaka et al., "Molecular cloning and chromosomal mapping of mouse intronless myc gene acting as a potent apoptosis inducer," *Gene*, vol. 226, no. 2, pp. 273–283, 1999 (English).

[5] A. J. Gentles and S. Karlin, "Why are human G-protein-coupled receptors predominantly intronless?" *Trends in Genetics*, vol. 15, no. 2, pp. 47–49, 1999 (English).

[6] J. C. Venter, "the sequence of the human genome," *Science*, vol. 292, no. 5507, pp. 1304–1351, 2001 (English).

[7] S. K. Tay, J. Blythe, and L. Lipovich, "Global discovery of primate-specific genes in the human genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 29, pp. 12019–12024, 2009 (English).

[8] X. Yang, S. Jawdy, T. J. Tschaplinski, and G. A. Tuskan, "Genome-wide identification of lineage-specific genes in Arabidopsis, Oryza and Populus," *Genomics*, vol. 93, no. 5, pp. 473–480, 2009 (English).

[9] G. R. Fink, "Pseudogenes in yeast?" *Cell*, vol. 49, no. 1, pp. 5–6, 1987 (English).

[10] E. Betrán, K. Thornton, and M. Long, "Retroposed new genes out of the X in Drosophila," *Genome Research*, vol. 12, no. 12, pp. 1854–1859, 2002.

[11] Y. Zhang, Y. Wu, Y. Liu, and B. Han, "Computational identification of 69 retroposons in Arabidopsis," *Plant Physiology*, vol. 138, no. 2, pp. 935–948, 2005.

[12] J. J. Emerson, H. Kaessmann, E. Betrán, and M. Long, "Extensive gene traffic on the mammalian X chromosome," *Science*, vol. 303, no. 5657, pp. 537–540, 2004.

[13] M. Jain, P. Khurana, A. K. Tyagi, and J. P. Khurana, "Genome-wide analysis of intronless genes in rice and Arabidopsis," *Functional and Integrative Genomics*, vol. 8, no. 1, pp. 69–78, 2008 (English).

[14] S. M. Agarwal, "Evolutionary rate variation in eukaryotic lineage specific human intronless proteins," *Biochemical and Biophysical Research Communications*, vol. 337, no. 4, pp. 1192–1197, 2005 (English).

[15] S. M. Agarwal and J. Gupta, "Comparative analysis of human intronless proteins," *Biochemical and Biophysical Research Communications*, vol. 331, no. 2, pp. 512–519, 2005 (English).

[16] M. K. Sakharkar et al., "Computational prediction of SEG (single exon gene) function in humans," *Frontiers in Bioscience*, vol. 10, pp. 1382–1395, 2005 (English).

[17] A. E. Hill and E. J. Sorscher, "The non-random distribution of intronless human genes across molecular function categories," *FEBS Letters*, vol. 580, no. 18, pp. 4303–4305, 2006 (English).

[18] K. R. Sakharkar, M. K. Sakharkar, C. T. Culiat, V. T. K. Chow, and S. Pervaiz, "Functional and evolutionary analyses on expressed intronless genes in the mouse genome," *FEBS Letters*, vol. 580, no. 5, pp. 1472–1478, 2006 (English).

[19] M. K. Sakharkar, P. Kangueane, D. A. Petrov, A. S. Kolaskar, and S. Subbiah, "SEGE: a database on 'intron less/single exonic' genes from eukaryotes," *Bioinformatics*, vol. 18, no. 9, pp. 1266–1267, 2002 (English).

[20] M. K. Sakharkar and P. Kangueane, "Genome SEGE: a database for 'intronless' genes in eukaryotic genomes," *BMC Bioinformatics*, vol. 5, article 67, 2004 (English).

[21] M. Punta and Y. Ofran, "The rough guide to in silico function prediction, or how to use sequence and structure information to predict protein function," *PLoS Computational Biology*, vol. 4, no. 10, Article ID e1000160, 2008 (English).

[22] L. J. Jensen, R. Gupta, N. Blom et al., "Prediction of human protein function from post-translational modifications and localization features," *Journal of Molecular Biology*, vol. 319, no. 5, pp. 1257–1265, 2002 (English).

[23] D. Fischer and D. Eisenberg, "Finding families for genomic ORFans," *Bioinformatics*, vol. 15, no. 9, pp. 759–762, 1999 (English).

[24] D. L. Wheeler, T. Barrett, D. A. Benson et al., "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research*, vol. 33, pp. D39–D45, 2005 (English).

[25] W. Li, L. Jaroszewski, and A. Godzik, "Clustering of highly homologous sequences to reduce the size of large protein databases," *Bioinformatics*, vol. 17, no. 3, pp. 282–283, 2001 (English).

[26] A. C. Marques, I. Dupanloup, N. Vinckenbosch, A. Reymond, and H. Kaessmann, "Emergence of young human genes after a burst of retroposition in primates," *PLoS Biology*, vol. 3, no. 11, article e357, pp. 1970–1979, 2005.

[27] Z. Yu, D. Morais, M. Ivanga, and P. M. Harrison, "Analysis of the role of retrotransposition in gene evolution in vertebrates," *BMC Bioinformatics*, vol. 8, article 308, 2007.

[28] N. Ben-Arie, D. Lancet, C. Taylor et al., "Olfactory receptor gene cluster on human chromosome 17: possible duplication of an ancestral receptor repertoire," *Human Molecular Genetics*, vol. 3, no. 2, pp. 229–235, 1994 (English).

[29] J. C. Dugas and J. Ngai, "Analysis and characterization of an odorant receptor gene cluster in the zebrafish genome," *Genomics*, vol. 71, no. 1, pp. 53–65, 2001 (English).

[30] I. King Jordan, I. B. Rogozin, Y. I. Wolf, and E. V. Koonin, "Essential genes are more evolutionarily conserved than are nonessential genes in bacteria," *Genome Research*, vol. 12, no. 6, pp. 962–968, 2002 (English).

[31] R. Kamikawa, Y. Inagaki, and Y. Sako, "Direct phylogenetic evidence for lateral transfer of elongation factor-like gene," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 19, pp. 6965–6969, 2008 (English).

[32] M. E. Rumpho, J. M. Worful, J. Lee et al., "Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug Elysia chlorotica," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 46, pp. 17867–17871, 2008 (English).

[33] E. V. Koonin, N. D. Fedorova, J. D. Jackson et al., "A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes," *Genome biology*, vol. 5, no. 2, p. R7, 2004 (English).

[34] T. Domazet-Loso and D. Tautz, "An evolutionary analysis of orphan genes in Drosophila," *Genome Research*, vol. 13, no. 10, pp. 2213–2219, 2003 (English).

[35] A. Lecharny, N. Boudet, I. Gy, S. Aubourg, and M. Kreis, "Introns in, introns out in plant gene families: a genomic approach of the dynamics of gene structure," *Journal of Structural and Functional Genomics*, vol. 3, no. 1–4, pp. 111–116, 2003 (English).

[36] N. Boudet, S. Aubourg, C. Toffano-Nioche, M. Kreis, and A. Lecharny, "Evolution of intron/exon structure of DEAD helicase family genes in Arabidopsis, Caenorhabditis and Drosophila," *Genome Research*, vol. 11, no. 12, pp. 2101–2114, 2001 (English).

[37] E. V. Gogvadze and A. A. Buzdin, "New mechanism of retrogene formation in mammalian genomes: in vivo recombination during RNA reverse transcription," *Molekulyarnaya Biologiya*, vol. 39, no. 3, pp. 364–373, 2005 (Russian).

[38] C. Esnault, J. Maestre, and T. Heidmann, "Human LINE retrotransposons generate processed pseudogenes," *Nature Genetics*, vol. 24, no. 4, pp. 363–367, 2000.

[39] E. S. Lander, L. M. Linton, B. Birren et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001 (English).

[40] Z. Zhang, P. M. Harrison, Y. Liu, and M. Gerstein, "Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome," *Genome Research*, vol. 13, no. 12, pp. 2541–2558, 2003 (English).

[41] K. Ohshima, M. Hattori, T. Yada, T. Gojobori, Y. Sakaki, and N. Okada, "Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates," *Genome Biology*, vol. 4, no. 11, article R74, 2003.

[42] N. B. Haas, J. M. Grabowski, A. B. Sivitz, and J. B. E. Burch, "Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames," *Gene*, vol. 197, no. 1-2, pp. 305–309, 1997.

[43] A. Wagner, "The fate of duplicated genes: loss or new function?" *BioEssays*, vol. 20, no. 10, pp. 785–788, 1998 (English).

[44] B. A. Chapman, J. E. Bowers, F. A. Feltus, and A. H. Paterson, "Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 8, pp. 2730–2735, 2006 (English).