Biochemical and Biophysical Research Communications 449 (2014) 146–150

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Different evolutionary patterns among intronless genes in maize genome

Hanwei Yan [1], Wei Zhang [1], Yongxiang Lin, Qing Dong, Xiaojian Peng, Haiyang Jiang, Suwen Zhu, Beijiu Cheng *

Key Laboratory of Crop Biology of Anhui Province, Anhui Agricultural University, Hefei 230036, China

ARTICLE INFO

Article history:
Received 16 April 2014
Available online 10 May 2014

Keywords:
Intronless
Maize
Function
Evolution

ABSTRACT

Intronless genes, as a characteristic feature of prokaryotes, are an important resource for the study of the evolution of gene architecture in eukaryotes. In the study, 14,623 (36.87%) intronless genes in maize were identified and the percentage is greater than that of other monocots and algae. The number of maize intronless genes on each chromosome has a significant linear correlation with the number of total genes on the chromosome and the length of the chromosomes. Intronless genes in maize play important roles in translation and energy metabolism. Evolutionary analysis revealed that 2601 intronless genes conserved among the three domains of life and 2323 intronless genes that had no homology with genes of other species. These two sets of intronless genes were distinct in genetic features, physical locations and function. These results provided a useful source to understand the evolutionary patterns of related genes and genomes and some intronless genes are good candidates for subsequent functional analyses specifically.

1. Introduction

Introns are the parts of eukaryotic genes that are not translated. Introns must spliced out of precursor RNA to form mature mRNAs [1]. However, prokaryotic genes lack introns. Although intronless genes are a characteristic feature of prokaryotes, eukaryotes have both intron-containing and intronless genes. The percentage of intronless genes varies from 2.7% to 97.7% of the genes in eukaryotic genomes [2]. Studying intronless genes enables the analysis of the evolution of gene architecture by identifying the expression patterns of related genes and genomes [3]. Furthermore, the presence of intronless genes provides an opportunity to understand why it is advantageous for a gene to be intronless. Systematic analyses of intronless genes in many species, from mammals to plants, have been reported over the past few decades. Some large gene families, such as G-protein-coupled receptors and olfactory receptors in human and mouse, are enriched in intronless genes [4,5]. Several databases for intronless genes in eukaryotes have been constructed [2,6,7]. The resources make comprehensive comparative analysis of intronless genes possible. However, research concerning intronless genes in maize is still in its infancy. Maize is a globally important agricultural crop for humans and animals. The maize genome has been sequenced, which provides a valuable resource to further understand the evolutionary history of intronless genes.

In this study, maize intronless genes were identified and distribution of intronless genes on chromosomes was then analyzed statistically. The function of maize intronless genes was also predicted. In addition, we analyzed the patterns of evolution across maize and other groups of organisms. Then gene duplication events in existing intronless genes and the retroposition of intron-containing genes were investigated to understand the evolutionary and genetic mechanisms that function in the maize genome. Expression patterns based on microarray data were analyzed. Our results may enhance bio-computational studies of intronless genes, and may also provide a biological reference for cloning intronless genes, which may contribute to genetic breeding in maize.

2. Materials and methods

2.1. Identification of maize intronless genes

A stringent protocol was used to reliably identify maize intronless genes [8]. First, the coding sequences (CDS) were compared to the corresponding gene sequences using Java program. When the

* Corresponding author. Fax: +86 551 5786021.
E-mail address: beijiucheng@ahau.edu.cn (B. Cheng).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bbrc.2014.05.008
0006-291X/

coding sequence of a gene matched its genomic sequence with 100% identity, the gene was chosen for further analysis. Second, all selected genes were examined to exclude redundant genes (i.e., different gene models representing the same genetic loci). Finally, a gene was selected if it was annotated as a protein-coding gene. In other words, transposable elements and pseudogenes were deleted. The final list of non-redundant intronless genes in maize was subjected to further analysis.

2.2. Predictions of protein functions

The prediction of the protein functions or GO categories of intronless genes in maize was performed by submitting all of the complete amino acid sequences to ProtFun [9–11].

2.3. Microarray analysis

To analyze the spatial and temporal expression patterns of intronless genes during development, transcriptome data of the genome-wide gene expression atlas of maize inbred line B73 made by the NimbleGen microarray technology was downloaded from Plexdb (ZM37). The microarray data of intronless genes were imported into R and Bioconductor for expression analysis. Then, the pheatmap package was used to make the heatmaps.

2.4. Taxonomic group(s)

Blink (http://www.ncbi.nlm.nih.gov/sutils/blink.cgi?mode=query) offers a very convenient way to run a precomputed BLAST search against the query protein in the Entrez protein database by taxonomic group at the protein level (including the groups archea, bacteria, fungi, plants, metazoans, and other eukaryotes).

2.5. Paralogs of maize intronless genes

To investigate paralogs of intronless genes in maize, the complete protein sequences of maize intronless genes were merged as an input, and multiple fasta format sequence alignments were then generated using CD-HIT (Cluster Database at High Identity with Tolerance) [12]. This program enabled the rapid clustering of similar sequences in the large protein database according to sequence identity. Each cluster file, which was obtained as output, was submitted to get the gene cluster groups. This method also can be applied to identify intronless genes in the maize genome using intron-containing paralogs.

3. Results and discussion

3.1. Identification and characterization of maize intronless genes

14,623 (36.87%) intronless genes were identified in maize (Supplementary Table 1). To determine whether the number of intronless genes varies considerably among plants, we characterized and compared the intronless genes in monocots and algae (Supplementary Fig. 1). Compared with the intronless genes identified in chlamydomonas (6.5%) and volvox (16.2%), the number of intronless genes in monocots was significantly higher, especially a greater proportion in maize intronless genes. One reason for such a difference may be that intronless genes can arise via reverse-transcript-mediated recombination [13], which has been confirmed to create a large amount of new genes. Additionally, this difference may be a result of the "intron early" hypothesis, posited that protein-coding genes were interrupted by numerous introns even at the earliest stages of life's evolution [14].

The number of intronless genes in each of the 10 chromosomes ranged from 1006 to 2237 (Supplementary Table 1). Of the intronless genes in maize, 15.3% were present on chromosome 1, which is the longest chromosome. Strikingly and interestingly, the ratio of intronless genes to genes was constant along chromosome (roughly 37%). In order to test whether the number of intronless genes on each chromosome correlated with the number of genes (including intronless and intron-containing genes) on the chromosome and the length of the chromosomes in maize, linear equation was employed. The result showed that the number of intronless genes on each chromosome are linear with the number of genes on the one (R^2 = 0.9935) as well as the length of their respective chromosome (R^2 = 0.9274; Fig. 1). We speculated whether the linear relationship is a universal law in other plants. To test this hypothesis, we further explored the relationship in *Sorghum bicolor*, *Brachypodium distachyon*. As we expected, the results were also in accordance with the linear correlation. In addition, neither the intronless genes nor the intron-containing genes were distributed evenly on the maize chromosomes. Our results showed a trend that intronless genes enrich both ends of all 10 chromosomes (Supplementary Table 2). On the contrary, the position of centromere on each chromosome exhibited lower number of intronless genes, perhaps because highly repetitive sequences of centromere hindered the construction of biological genetic and physical map.

3.2. Functional groups of maize intronless genes

The functions of the 14,623 intronless genes in maize are only partially understood. A significantly large percentage of intronless genes were annotated as uncharacterized proteins (81%) or even had no description information (1%) in NCBI (Supplementary Table 1). This is mainly due to the large amount of manual work required for re-annotating individual sequences and integrating the data. To investigate the functions of intronless genes in maize, the cellular roles and GO categories of these genes were predicted using ProtFun. This analysis revealed that the largest percentage of intronless genes (32.37%) is included in the translation (T) category, followed by energy metabolism (EM; 22.55%; Fig. 2A). Therefore, intronless genes may play a crucial role in translation and in various metabolic pathways in maize. In order to further explore the conservation of intronless genes among different functional categories, the cellular role of intronless genes in maize were compared with those in sorghum (Supplementary Table 3). Although the number of intronless genes in maize (14,623) is significantly larger than that in sorghum (6197), the number of intronless genes belonging to central intermediary metabolism (CIM) category is similar in maize (675) and sorghum (656). The percentages of intronless genes with the same cellular roles were generally similar between maize and sorghum except one; specifically, about 32.37% of maize intronless genes were associated with translation, which was higher than that of the sorghum intronless genes (23.04%). A comparative analysis of intronless genes in different species revealed that intronless genes in the CIM category are conserved, suggesting that maize intronless genes in the functional category share no great expansion after the split of maize and sorghum. In addition, higher percentage of intronless genes involved in translation was demonstrated that the proteins of intronless genes play significant role in basic cellular processes. To learn more about the functions of the intronless genes in maize, the GO category profile of all 14,623 intronless genes were also assessed. A total of 12,147 of these genes fit into 13 different of GO categories (Fig. 2B). Among these, growth factor (GF) category contained the most genes (27.73%), while the fewest genes (0.34%) were present in cation channel (CC) category. Strikingly, compared the distribution of GO category of intronless genes in maize and that in chlamydomonas (Supplementary Table 4), the biggest proportion

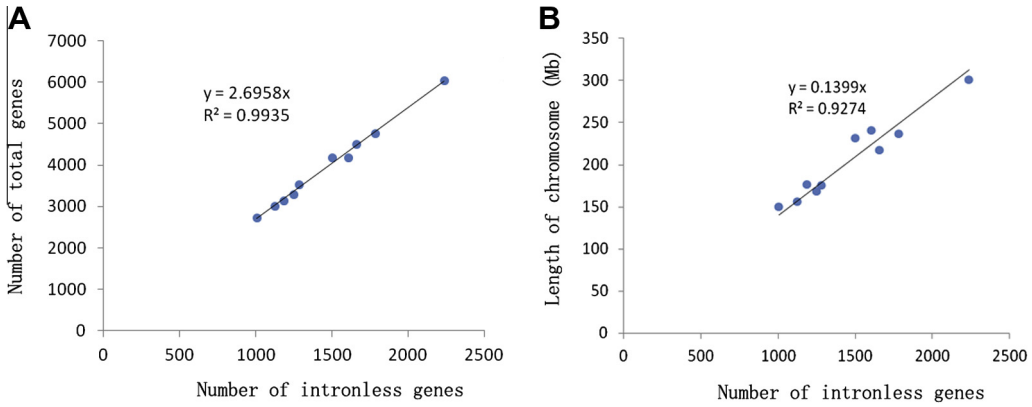

Fig. 1. Linear correlation among intronless genes, total genes, and the length of chromosome. (A) The linear regression equation between the number of intronless genes on each chromosome and the number of genes (including intronless and intron-containing genes) on the chromosome. (B) The linear regression equation between the number of intronless genes on each chromosome and the length of their respective chromosome.

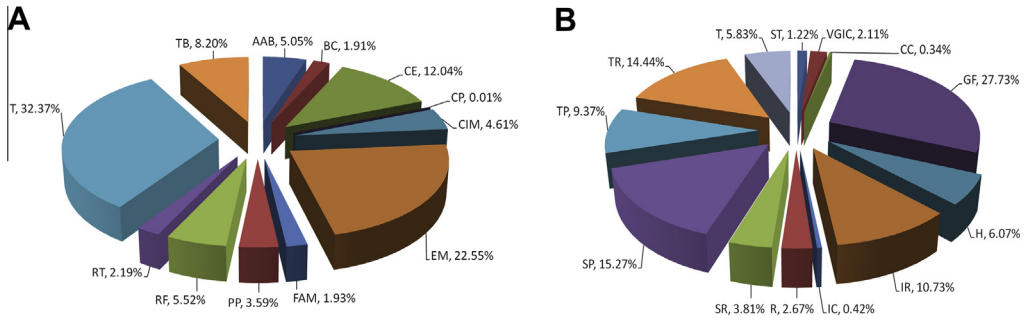

Fig. 2. Functional categorization of maize intronless genes. (A) Distribution of maize intronless genes among 12 cellular role categories: amino acid biosynthesis (AAB), biosynthesis of cofactors (BC), cell envelope (CE), cellular processes (CP), central intermediary metabolism (CIM), energy metabolism (EM), fatty acid metabolism (FAM), purines and pyrimidines (PP), regulatory functions (RF), replication and transcription (RT), translation (T), transport and binding (TB). (B) Distribution of maize intronless genes among 13 GO categories: signal transducer (ST), voltage-gated ion channel (VGIC), cation channel (CC), growth factor (GF), hormone (H), immune response (IR), ion channel (IC), receptor (R), stress response (SR), structural protein (SP), transcription (TP), transcription regulation (TR) and transporter (T).

increase was seen on "immune response", implying a significantly larger expansion of the intronless genes involved in immune response from lower plant to monocot plant. Genes involved in the immune response are rapidly regulated. Genes that are rapidly regulated during stress responses tend to lack introns; intronless genes are transcribed without undergoing the process of splicing, enabling the organism to rapidly respond to stress conditions [15]. Therefore, immune response genes are expected to have few or no introns. Furthermore, the significant increase of immunity proteins in maize intronless genes indicated that maize intronless genes involved in immunity should help maize better to respond to different biotic and abiotic stresses. Maize spreads all over the world because of its ability to grow in diverse climates. Although it is grown mainly in wet, hot climates, it has been reported to thrive in cold, hot, dry or wet conditions, meaning that it is an extremely versatile crop. Maize can be better adapt to various biotic and abiotic stresses, such as drought, high salt content, high or freezing temperature, and insect or pathogen attack [16,17]. Therefore, the enrichment of intronless genes associated with immune response may be an important reason for the successful diversification.

3.3. Expression profile of maize intronless genes

To understand the temporal and spatial transcription patterns of intronless genes, a comprehensive expression analysis across 60 distinct tissues representing 11 major organ systems of inbred line B73 was performed by utilizing the publicly available microarray data for maize [18]. Expression profiles were identified for a total of 10,063 intronless genes (Fig. 3, Supplementary Table 5), whereas no corresponding probe sets were available for the remaining intronless genes in the dataset. We postulated that the absent genes in the datasets might because these genes were expressed under specific environmental conditions or were very specific to organs and/or developmental stages that were not covered in the dataset. The expression profiles demonstrated that majority intronless genes have rather broad expression patterns with presence in diverse libraries. These genes may be essential for maintaining cell survival. The genes are being divided into three clusters based on expression level. 12.19% of the detected intronless genes belonged to Cluster I, which had relatively high expression levels. Cluster II contained 6978 intronless genes with relatively low expression levels. Genes in Cluster III were expressed in a mid-level. In order to investigate the differential expression patterns in 60 developmental stages and diverse tissues, the coefficient of variation (CV value; standard deviation/mean) of each gene were calculated, and the results showed huge variation among all the intronless genes with the CV values ranging from 0.93% to 54.78%. 7856 intronless genes with a CV value of <15% had the least expression variability [19]. And these genes were expressed stably across all tissue, which may work synergistically with other genes during plant growth and development. Conversely, the remaining 2207 intronless genes with the CV values of $\geqslant$15%, including 1095 genes with a CV value greater than 20%, were considered to be expressed at stage-specific development. It is meaningful to study the intronless genes with a CV values of $\geqslant$15% for further discovering new organ- or tissue-specific genes in maize. Microarray data showed that the mean normalized

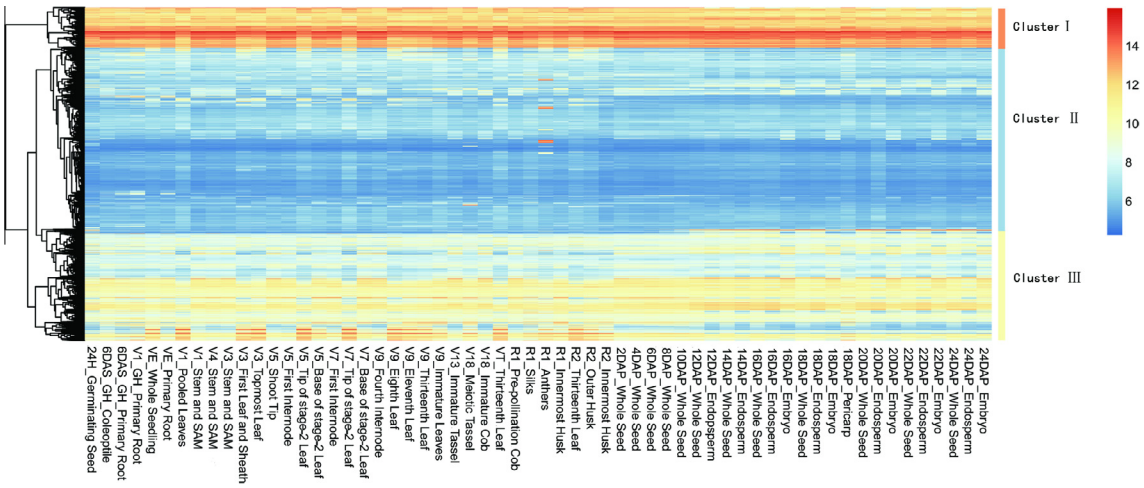

Fig. 3. Expression profiles of intronless genes across the 60 different development stages. The heatmap showed clustering of intronless genes according to the signal intensities of 10,063 detected transcripts in maize. Red, yellow and blue indicated high, medium and low levels of gene expression, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

level of gene expression was generally higher before pollination than after pollination (*t*-test, $p < 0.01$; Supplementary Fig. 2). Expression analysis combined with function prediction confirmed the potential roles of intronless genes as growth factor. Interestingly, upon sorting the normalized expression levels of the intronless genes for the 60 development stage from highest to lowest, we found that the largest proportion of genes having the highest expression level was in the R1 (reproductive 1, with silks emerged from the husk) anthers stage (2.39%). Thus, we speculated that these genes may play a role in meiosis network and pollen wall synthesis network.

3.4. Evolutionary analysis of maize intronless genes

According to Woese's theory [20], life is divided into three domains, i.e., archaea, bacteria and eukaryotes. To analyze the evolutionary conservation of maize intronless genes among these taxonomic groups, we conducted searches using Blink (Fig. 4). Among the 14,473 intronless genes examined, 25 genes that were present in bacteria but absent in archaea and eukaryotes, and 6067 genes

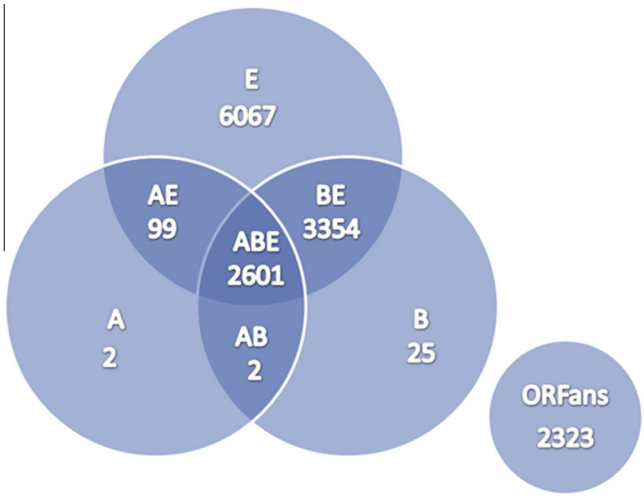

Fig. 4. Classification of maize intronless genes among different taxonomic groups. Venn diagram showed the classification of 14,473 maize intronless genes into three domains of life based on homology, including archaea (A), bacteria (B), eukaryotes (E), archaea and bacteria (AB), archaea and eukaryotes (AE), bacteria and eukaryotes (BE), archaea, bacteria and eukaryotes (ABE).

were present in eukaryotes but absent in archaea and bacteria. However, only 2 genes were identified with a homology cluster only in archaea. According to the origin and evolution of species, there are two possible reasons for these distributional profiles. First, the genes may have been present in a common ancestor and later lost in one of the domains. Second, the genes may have been horizontally transferred after the domains split. In contrast, 2601 intronless genes, accounting for 17.97% of the detected intronless genes, were conserved across all three biological kingdoms (i.e., the ABE group).

Interestingly, 2323 (16.05%) of the intronless genes are maize-specific genes that share no homology with genes of other species. These sequences are known as ORFans [21]. Apart from being abundant, ORFans have been thought to be important for maize-specific traits and adaptations [22]. However, about 96.13% of ORFans have no description information in NCBI. Therefore, we attempted to classify the maize intronless ORFans into diverse functional categories using ProtFun. We found that 924 of these genes take part in translation, followed by approximately 781 intronless genes that function in energy metabolism. Clearly, 73.40% of ORFans corresponded to essential and basic cellular machinery. Although abundant in quantity and important in functions, the evolutionary origin of ORFans is still enigmatic. There are some hypotheses about the origin of ORFans: they may have been involved in gene duplication followed by rapid sequence divergence, they may have arisen from lateral gene transfer (LGT), have an accelerated rate of evolution, or be artifacts of genome annotation [23]. To further explore the difference between ABE and ORFans, we characterized the two set of these genes by the following features, including gene size, GC content, protein size, isoelectric point (PI), molecular weight (MW), physical locations and functional prediction (Supplementary Table 6). The results showed that the average gene size, average GC content, average protein size, average MW of ORFans were all lower than that of the ABE group. However, the PI of ORFans was higher than that of the ABE group. It has been reported that younger genes tend to have shorter gene size, shorter proteins than order genes [24,25]. Thus, ORFans were regarded as younger genes that have arisen in relatively recent years. Interestingly, the number of ORFans in the region of centromere on each chromosome was significantly higher than that of ABE group (one-way ANOVA, $p < 0.05$). This dramatic difference may be because centromeric DNA sequences have no homolog in other species [26–28]. Thus, some maize-specific genes in the region of centromere may play important role in segregating

chromosomes into daughter cells. In addition, the distribution of the functional categories of the two sets is different. 465 (17.88%) of ABE group were associated with the cell envelope, which was significantly higher than that of ORFans (6.63%). With respect to the GO categories, the structural protein category (SP; about 20.20%) was a relatively abundant category for ORFans, while the percentage of SP for ABE group was only 8.73%. Therefore, ORFans were distinct from the ABE group in terms of genic features, physical locations and function.

Gene duplications were shown to play a crucial role in the succession of genomic rearrangements and expansions [29]. CD-HIT was employed to further investigate gene duplication events in existing intronless genes and the retroposition of intron-containing genes, which provided some valuable information. First, we clustered maize intronless genes to determine the relationship between the intronless genes. We found that 3833 (26.21%) of the maize intronless genes have at least one intronless gene paralog. It was suggested that more than a quarter of intronless genes in maize may have been derived from other intronless maize genes. The most remarkable outcome of a gene duplication event is the evolution of a novel function, as sister genes may produce different functional proteins by alternative splicing. Conversely, the function of paralogous intronless genes was similar. To prove there notions, we investigated the functional relationships of paralogs using Pfam website [30]. As expected, almost paralogous genes shared same family (Supplementary Table 1). Gene duplications include tandem and segmental duplication events. Through analyzing paralogous genes on chromosomal location, we can better evaluate tandem gene duplications in maize intronless genes. In our study, 1056 paralogous genes were clustered on the maize chromosomes, that is, almost a third paralogs were generated via tandem duplications. This finding suggested that segmental duplications of maize intronless genes may be more prevalent than tandem duplications. In addition, we clustered intronless genes with intron-containing genes. The results showed that 270 intronless genes have paralogs of intron-containing genes. The possible explanation can be given for this phenomenon. Due to the reverse transcription of mature mRNAs from parental source genes, these genes are therefore devoid of introns, which is the main reason for the performance of retrogenes [31]. Gene duplication via retrotransposition has been shown to be an important mechanism in evolution, affecting gene dosage and allowing for the acquisition of new gene functions [32].

In summary, a comprehensive genome-wide analysis of identification, characterization on maize intronless genes was performed. The number of intronless genes on each of chromosome significantly correlated with the number of total genes and the length of their respective chromosome. In addition, intronless genes were classified distinct groups based on different functional and evolutionary patterns. Investigation into the groups of ABE and ORFans will be important for understanding lineage-specific adaptation in maize. To study intronless genes could help to understand the evolutionary patterns of related genes and genomes.

Acknowledgments

This work was supported by Grants from the Science & Technology Program of Anhui Province (11010301026) and the Genetically Modified Organisms Breeding Major Projects (2011ZX08003-002). We also thank members of the Key Laboratory of Crop Biology of Anhui Province for their assistance in this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2014.05.008.

References

[1] R.A. Padgett, P.J. Grabowski, M.M. Konarska, S. Seiler, P.A. Sharp, Splicing of messenger RNA precursors, Annu. Rev. Biochem. 55 (1986) 1119–1150.
[2] A. Louhichi, A. Fourati, A. Rebai, IGD: a resource for intronless genes in the human genome, Gene 488 (2011) 35–40.
[3] K.R. Sakharkar, M.K. Sakharkar, C.T. Culiat, V.T. Chow, S. Pervaiz, Functional and evolutionary analyses on expressed intronless genes in the mouse genome, FEBS Lett. 580 (2006) 1472–1478.
[4] A.J. Gentles, S. Karlin, Why are human G-protein-coupled receptors predominantly intronless?, Trends Genet 15 (1999) 47–49.
[5] X. Zhang, S. Firestein, The olfactory receptor gene superfamily of the mouse, Nat. Neurosci. 5 (2002) 124–133.
[6] M.K. Sakharkar, P. Kangueane, D.A. Petrov, A.S. Kolaskar, S. Subbiah, SEGE: a database on 'intron less/single exonic' genes from eukaryotes, Bioinformatics 18 (2002) 1266–1267.
[7] M.K. Sakharkar, P. Kangueane, Genome SEGE: a database for 'intronless' genes in eukaryotic genomes, BMC Bioinform. 5 (2004) 67.
[8] M. Jain, P. Khurana, A.K. Tyagi, J.P. Khurana, Genome-wide analysis of intronless genes in rice and *Arabidopsis*, Funct. Integr. Genom. 8 (2008) 69–78.
[9] L.J. Jensen, R. Gupta, N. Blom, D. Devos, J. Tamames, C. Kesmir, H. Nielsen, H.H. Staerfeldt, K. Rapacki, C. Workman, C.A. Andersen, S. Knudsen, A. Krogh, A. Valencia, S. Brunak, Prediction of human protein function from post-translational modifications and localization features, J. Mol. Biol. 319 (2002) 1257–1265.
[10] L.J. Jensen, R. Gupta, H.H. Staerfeldt, S. Brunak, Prediction of human protein function according to gene ontology categories, Bioinformatics 19 (2003) 635–642.
[11] L.J. Jensen, D.W. Ussery, S. Brunak, Functionality of system components: conservation of protein function in protein feature space, Genom. Res. 13 (2003) 2444–2449.
[12] W. Li, L. Jaroszewski, A. Godzik, Clustering of highly homologous sequences to reduce the size of large protein databases, Bioinformatics 17 (2001) 282–283.
[13] J.M. Logsdon Jr., A. Stoltzfus, W.F. Doolittle, Molecular evolution: recent cases of spliceosomal intron gain?, Curr Biol. 8 (1998) R560–R563.
[14] E.V. Koonin, The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?, Biol Dir. 1 (2006) 22.
[15] D.C. Jeffares, C.J. Penkett, J. Bahler, Rapidly regulated genes are intron poor, Trends Genet. 24 (2008) 375–378.
[16] D. Kizis, V. Lumbreras, M. Pages, Role of AP2/EREBP transcription factors in gene regulation during abiotic stress, FEBS Lett. 498 (2001) 187–189.
[17] T. Yang, L. Zhang, T. Zhang, H. Zhang, S. Xu, L. An, Transcriptional regulation network of cold-responsive genes in higher plants, Plant Sci. 169 (2005) 987–995.
[18] R.S. Sekhon, H. Lin, K.L. Childs, C.N. Hansey, C.R. Buell, N. de Leon, S.M. Kaeppler, Genome-wide atlas of transcription during maize development, Plant J. 66 (2011) 553–563.
[19] T. Ishida, S. Hattori, R. Sano, K. Inoue, Y. Shirano, H. Hayashi, D. Shibata, S. Sato, T. Kato, S. Tabata, K. Okada, T. Wada, Arabidopsis TRANSPARENT TESTA GLABRA2 is directly regulated by R2R3 MYB transcription factors and is involved in regulation of GLABRA2 transcription in epidermal differentiation, Plant Cell 19 (2007) 2531–2543.
[20] C.R. Woese, Bacterial evolution, Microbiol. Rev. 51 (1987) 221–271.
[21] I. Yomtovian, N. Teerakulkittipong, B. Lee, J. Moult, R. Unger, Composition bias and the origin of ORFan genes, Bioinformatics 26 (2010) 996–999.
[22] M. Long, E. Betran, K. Thornton, W. Wang, The origin of new genes: glimpses from the young and old, Nat. Rev. Genet. 4 (2003) 865–875.
[23] L. Yang, M. Zou, B. Fu, S. He, Genome-wide identification, characterization, and expression analysis of lineage-specific genes within zebrafish, BMC Genom. 14 (2013) 65.
[24] M.T. Donoghue, C. Keshavaiah, S.H. Swamidatta, C. Spillane, Evolutionary origins of *Brassicaceae* specific genes in *Arabidopsis thaliana*, BMC Evol. Biol. 11 (2011) 47.
[25] D.J. Lipman, A. Souvorov, E.V. Koonin, A.R. Panchenko, T.A. Tatusova, The relationship of protein conservation and sequence length, BMC Evol. Biol. 2 (2002) 20.
[26] A. Houben, I. Schubert, DNA and proteins of plant centromeres, Curr. Opin. Plant Biol. 6 (2003) 554–560.
[27] H.S. Malik, S. Henikoff, Conflict begets complexity: the evolution of centromeres, Curr. Opin. Genet. Dev. 12 (2002) 711–718.
[28] K. Nagaki, K. Kashihara, M. Murata, A centromeric DNA sequence colocalized with a centromere-specific histone H3 in tobacco, Chromosoma 118 (2009) 249–257.
[29] T.J. Vision, D.G. Brown, S.D. Tanksley, The origins of genomic duplications in Arabidopsis, Science 290 (2000) 2114–2117.
[30] R.D. Finn, J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S.R. Eddy, E.L. Sonnhammer, A. Bateman, Pfam: clans, web tools and services, Nucleic Acids Res. 34 (2006) D247–D251.
[31] M. Fablet, M. Bueno, L. Potrzebowski, H. Kaessmann, Evolutionary origin and functions of retrogene introns, Mol. Biol. Evol. 26 (2009) 2147–2156.
[32] D.R. Schrider, K. Stevens, C.M. Cardeno, C.H. Langley, M.W. Hahn, Genome-wide analysis of retrogene polymorphisms in *Drosophila melanogaster*, Genom. Res. 21 (2011) 2087–2095.