Massive Horizontal Gene Transfer in Amphioxus Illuminates the Early Evolution of Deuterostomes

Authors: Qing Xiong[1,2]†, Kevin Yi Yang[1,2]†, Xi Zeng[3]†, Mingqiang Wang[1,4]†, Patrick Kwok-Shing Ng[1], Jun-Wei Zhou[1], Judy Kin-Wing Ng[1], Cherie Tsz-Yiu Law[1], Qiao Du[3], Kejin Xu[3], Laura Jane Falkenberg[5], Bingyu Mao[6], Jun-Yuan Chen[7], Stephen Kwok-Wing Tsui[1,2,8] *

Affiliations:

1 School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong.
2 Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Hong Kong.
3 Agricultural Bioinformatics Key Laboratory of Hubei Province, 3D Genomics Research Centre, College of Informatics, Huazhong Agricultural University, Wuhan, China.
4 Stanford Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, 94305, USA.
5 Simon FS Li Marine Science Laboratory, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong.
6 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 650223, Kunming, Yunnan, China.
7 State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing, China.
8 Centre for Microbial Genomics and Proteomics, The Chinese University of Hong Kong, Hong Kong.

* Correspondence to Professor Stephen Kwok-Wing Tsui, School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. E-mail: kwtsui@cuhk.edu.hk; Tel: 852-3943 6381; Fax: 852-2603 5123.

† These authors contributed equally to this work.

Abstract

Background

Horizontal gene transfer (HGT) enables novel gene establishment, has frequently reported in the genomes of protostomes but rarely in that of deuterostomes. As one of the closest living invertebrate relatives of vertebrates, amphioxus is considered as an important group for the evolutionary study of vertebrates and deuterostomes.

Results

In this study, a high-quality genome of amphioxus was *de novo* assembled and annotated. Using the genomes of four amphioxi, massive HGT events were identified and introduced novel genes in different functional groups such as follows. Innate immunity of amphioxi that has no adaptative immune system was enriched by HGT genes including that for a group of lipoxygenases and a DEAD-box helicase. Anaerobic adaptation of amphioxi may have been enabled by the HGT of genes for alcohol dehydrogenase and ferredoxin. A proximally arrayed cluster of EF-hand calcium-binding protein genes similar to that of bacteria, of which the copy number was linearly correlated to the sea salinity of the collection region, may foster their survival at different calcium concentrations.

Conclusions

Massive HGT events were involved in a range of amphioxus adaptations such as innate immunity, anaerobic adaptation, and calcium balance. Combining phylogenetic analysis and conservation among species, the comprehensive analysis revealed the neglected roles of HGT in the early genome evolution of vertebrates and deuterostomes.

Keywords

Horizontal gene transfer, Gene novelty, Genome evolution, Amphioxus, Vertebrate, Deuterostome

Background

Horizontal gene transfer (HGT) is the horizontal movement of genetic materials between organisms without a parent–offspring relationship and has been widely recognized among prokaryotes and eukaryotes [1-3]. HGT-involved genes (referred to as HGT genes) have been frequently reported in protostomes, especially in insects, but rarely identified in deuterostomes, which is a sister group composed of Chordata (Vertebrata, Urochordata and Cephalochordata) and Ambulacraria (Hemichordata and Echinodermata) [2-7]. On the deuterostome stem, many unusual phylogenetically distributed genes were identified as possible HGT genes, and they enabled a wide range of gene novelties [8]. Although genes similar to a type II antifreeze protein (AFP) gene are horizontally transferred among fishes [9], HGT genes from other nonmetazoan cellular species, such as bacteria, remains controversial in vertebrate deuterostome groups [10-12]. For nonvertebrate deuterostome groups, a cellulose synthase gene and a GH6 hydrolase-like gene of the tunicate (Urochordata) *Ciona intestinalis* were proposed as HGT genes of bacterial origin [13, 14], although their origin was possibly chimeric [15, 16]. Among the three extant chordates, the cephalochordate is regarded as evolutionary stasis [17], providing valuable evidence for our understating of the origination both of chordates and vertebrates. The living fossils of cephalochordate, amphioxi are an appropriate outgroup for evolutionary studies of deuterostomes, chordates, and vertebrates. Using annotated genomes of four amphioxi, this study identified a wide range of HGT events that introduced many functional genes related to amphioxus adaptations such as innate immunity, anaerobic adaptation, and calcium balance, and a comparative analysis provides many insights into the early evolution of deuterostomes.

Results and discussion

A high-quality genome of the Beihai amphioxus, *Branchiostoma* (*B.*) *belcheri beihai* [18], was *de novo* assembled, annotated, and compared with three other genome assemblies of the Xiamen amphioxus (*B. belcheri xiamen*) [19], Florida amphioxus (*B. floridae*) [20] and European amphioxus (*B. lanceolatum*) [21] (Supplementary Note 1.1, Table S1, Fig. S1). For the genome assembly of *B. belcheri beihai*, the scaffold N50 length (4,185,906 bp) was second only to that of *B. floridae* (25,441,410 bp), the gap content was the lowest among the four amphioxi and the BUSCO completeness was one of the highest (97.2%) (Supplementary Note 1.2, Table S1). In the genome annotation of *B. belcheri beihai*, 44,745 protein-coding genes were annotated with 95.2% BUSCO completeness. The genome of *B. lanceolatum* was annotated with 38,190 protein-coding genes with 91.9% completeness (Supplementary Note 1.3, Table S1).

Along with other genome assemblies (Supplementary Note 2.1, Table S2), a phylogenetic analysis was performed, with the sponge *Amphimedon queenslandica* used as the outgroup. The deuterostome phylogeny of this organism was consistent with current results [22, 23] and supported the monophylies of both chordates and deuterostomes (Fig. 1). Additional deuterostome phylogenetic analyses further supported the monophylies and refuted confusing phylogenies (Supplementary Note 2.2, Fig. S2). In all of these phylogenetic analyses, amphioxi were placed at a core position among deuterostomes and suggested as the outermost or first-evolving chordate group (Fig. 1 and S2A).

Within four annotated amphioxus genomes (Table S1), we identified a wide range of conserved HGT events among them (Table 1), in which the HGT genes possibly originated from bacteria, fungi,

plants or other eukaryotes. In addition to cautious filtering and rigorous validation to avoid false positivity, the high conservation of those HGT genes among four independently collected, sequenced, and assembled genomes of amphioxi (Table 1) precluded the occurrence of contamination. Notably, to avoid repeating studies on the gene novelties of stem deuterostomes [8], all HGT candidates that are also conserved in jawed vertebrates, such as the sialidase gene, which may have arisen in the deuterostome stem (Fig. S3, Supplementary Note 3.1), were excluded in this study. Therefore, this HGT study only explored those HGT genes that are not observed in jawed vertebrates but are possibly conserved in other deuterostome groups (Table 1 and S3). We only explored apparent cross-kingdom HGT events, such as HGT genes from bacteria, but not HGT genes from viruses or deuterostomes, such as the type II AFP of fishes [9]. Moreover, HGT in deuterostomes is difficult to distinguish with gene loss in other groups like protostomes and/or extreme divergence [8]. Therefore, low-confidence HGT candidates were discarded and only those candidates highly similar with that of unusual taxonomic groups like bacteria were kept in the final list (Table 1), unlike the green fluorescent proteins (GFPs) without apparently unusual taxonomic origin [24].

An HGT event could generate many novel genes via gene duplications; for example, in *B. belcheri beihai*, 28 HGT events generated 108 new genes (Table 1). According to their biological functions, these HGT genes were classified into six categories, namely, innate immunity, anaerobic adaptation, calcium balance, amino acid metabolism, other metabolic genes, and others without explicit functional domains (Table 1). The expression level of the HGT genes was analyzed in different tissues of *B. lanceolatum* because of its high-quality transcriptome data, and all HGT genes were

expressed in at least one tissue (Fig. S4). Regardless of the inconsistent gene expression in tissue duplicates of muscle and male gonads, the most marked expression feature was that HGT genes coding for calcium-binding proteins were highly expressed in the guts and hepatic diverticula (Fig. S4) (Supplementary Note 3.2). Subsequently, we explored these HGT genes in three important functional groups, namely innate immunity, anaerobic adaptation, and calcium balance.

Innate immunity

Innate immunity is the first-line nonspecific immune system against infections, and this immune system is observed in jawed and jawless vertebrates as well as invertebrates [25]. Successful immunity evolution confers survival traits to a species. In the absence of an adaptative immune system in amphioxus [25], innate immunity is the most important weapon of amphioxi in response to various pathogens. In this study, we identified several innate immunity-related HGT genes in amphioxi (Table 1).

Group of lipoxygenases

Lipoxygenases (LOXs) are important in arachidonic acid metabolism, participate in inflammatory lipid signaling and further regulate inflammatory and allergic responses [26, 27]. A group of duplicated LOX genes of amphioxi were identified as HGT genes similar to those of cyanobacteria or proteobacteria (Table 1). Notably, these LOX genes have unexpected homologs from only two living groups of jawless vertebrates (Cyclostomata) [28], the sea lamprey *Petromyzon marinus* and the hagfish *Eptatretus burger* (Fig. S5); however, such homologs are not observed in other metazoa.

In the phylogenetic tree, these LOX genes of amphioxi are close to those of sea lamprey, cyanobacteria and proteobacteria (Fig. 2A). Three splice sites (locations of exon–intron boundaries) were conserved among the LOX genes of four amphioxi and the sea lamprey (Fig. S5). In the absence of genome annotation, two fragments of the LOX gene were recovered from the genome of the hagfish, and three other conserved splice sites were identified between those of the two jawless vertebrates (Fig. S5). Therefore, we suggest that the LOX genes of four amphioxi and two jawless vertebrates originated from the same eukaryotic source. To the best of our knowledge, this HGT gene is the first reported HGT gene of vertebrates from bacteria.

The phylogenetic analysis revealed that these LOX genes were duplicated and diversified among the four amphioxi and *LOX_Bf_9* was an outlier gene (Fig. 2B). In the domain structure, all LOXs contain a C2-like domain (Fig. 2B). Except for the five LOXs of *B. floridae* and *B. lanceolatum*, the LOXs possessed one to five domains of the PYR (pyrabactin resistance)/PYL (PYR1-like)/RCAR (regulatory components of ABA receptor) family in the SRPBCC superfamily (SRPBCC domains, Fig. 2B) and these SRPBCC domains could be further divided into three groups (Fig. S6). In Fig. 2C, the gene synteny alignment revealed that many LOXs (12 of 27 genes) were proximally located (distance <20 Kb) and possibly generated by tandem gene duplications. *LOX_Bl_5* had an apparently wide expression in the gills, gut, and hepatic diverticula (Fig. S4).

These unusual LOX genes containing a C2-like domain were reported in amphioxi [26, 27] and first in jawless vertebrates, although the SRPBCC domains were not identified, and an HGT analysis was not performed. Since SRPBCC superfamily proteins contain a deep hydrophobic ligand-binding

pocket and can bind diverse ligands [29], we consider that these SRPBCC domains may also function as receptor domains in these LOX genes in response to various signals. A multifunctional LOX containing a SRPBCC domain was reported in the red alga *Pyropia haitanensis* and presented substrate flexibility, oxygenation position flexibility, and functional versatility [29]. The HGT of LOX genes may enrich the inflammatory lipid signaling pathway of amphioxi.

DEAD/DEAH-box helicase

DEAD-box helicases (DBHs) are known as sensors for nucleic acids and mediators of antiviral innate immunity [30]. A DBH gene was identified as a single-copy HGT gene originating from bacteria, and it was conserved among four amphioxi (Table 1). In the phylogenetic tree (Fig. 3A), this DBH gene of amphioxi is close to those of the charophyte alga *Klebsormidium* (*K.*) *nitens*, soft coral *Dendronephthya gigantea*, hemichordate *Saccoglossus kowalevskii*, and a group of bacteria. Intriguingly, the DBH genes of amphioxi shared three conserved splice sites with those of charophyte alga, soft coral, and hemichordate (Fig. 3B and S7), which suggested that the DBH genes of four amphioxi, soft coral and hemichordate shared a common source with that of the charophyte alga. A homologous DBH gene was also identified in the genome of the other hemichordate, *Ptychodera flava*, in the absence of annotation. This DBH gene was widely expressed at low levels in the tissues of *B. lanceolatum* (Fig. S4). The single-copy DBH gene may serve as a novel sensor in the antiviral innate immunity of amphioxi. A DEAD-like helicase was reported as an HGT gene in coral species [31], although it was not identical to this DBH gene.

In addition, another nine conserved splice sites in orange lines (Fig. 3B and S7) indicated close

phylogenetic relationships among the DBHs of the four amphioxi, soft coral and hemichordate; another four conserved splice sites in blue lines (Fig. 3B and S7) suggested that the DBHs of the soft coral and hemichordate originated from the same ancestral gene (Fig. 1 and S2), while two unusually conserved splice sites marked with black lines (S1 and S2, Fig. 3B and S7) could be explained by intron loss during gene evolution, which indicated S1 loss in ancestral amphioxus but S2 loss in both the ancestral amphioxus and hemichordate.

Ricin B-like lectins and TIR-domain-containing genes

Ricin B-like lectins (RBLs) are a family of carbohydrate-binding proteins that were reported to function in antifungal immunity of the Colorado potato beetle [32]. An RBL gene was identified as an HGT gene of amphioxi (Table 1). Three proximally arrayed RBL genes were identified in *B. floridae* (Fig. S8A). Except for *RBL_Bl_1* (Fig. S8A) of *B. lanceolatum*, other RBL genes encode two tandem lectin domains with low similarity. Their N- and C- terminal domains are two distinct ricin B-like lectins (Fig. S8B) that share only 42% identity (in *RBL_Bbb_1*), although the conserved splice sites supporting two domains were generated via duplication of one ancestral gene (Fig. S8C).

Toll/interleukin-1 receptor (TIR) domains are protein interaction domains responsible for interactions among TIR-containing proteins in innate immunity [33]. Two groups of TIR domains (TIR-a and TIR-b) originating from bacteria were assembled with various immune-related functional domains. TIR-a is a smaller group than TIR-b (Table 1), and all TIR-a-containing genes are proximally arrayed (distance <20 Kb) in all four amphioxi (Fig. S9A). Four TIR-a-containing genes of *B. belcheri beihai*, five genes of *B. belcheri xiamen*, and two genes of *B. floridae* and *B. lanceolatum* were

identified, and some of these genes were combined with two DEATH domains or a DED domain (Fig. S9A). Compared with TIR-a-containing genes, TIR-b-containing genes are much more complicated. A number of TIR-b-containing genes were analyzed to determine the domain structure (Fig. S9B) and a group of Myd88-like genes that contain a CARD domain, a group of RIG-I/MDA5-like genes containing two helicase domains and a group of LRR-containing genes were identified in the domain structure (Fig. S9B). None of the Myd88-like and RIG-I/MDA5-like genes containing the TIR-b domain were identified in *B. lanceolatum* (Fig. S9B). Compared with the other three amphioxi, *B. lanceolatum* has fewer and less diverse gene models containing these two TIR domains. Toll-like receptor (TLR) models in the amphioxus genome were reported to be much more complex than those in the human genome [34]. These TIR domain-containing genes are widely expressed and relatively highly expressed in gills, which are important structures in the amphioxus immune system (Fig. S4) [35].

Anaerobic adaptation

Amphioxi largely live in subtidal shallow water with extremely fluctuating dissolved oxygen level [36] and burrow in sandy substrate where is often rich in excrement and anaerobic [37]. Amphioxi do not have a mature respiratory system and breathe mainly through at the surface over the coelomic cavity [38]. Under anoxic conditions, the soft tissues of amphioxus (*B. lanceolatum*) can survive for as long as 124 days [39]. Additionally, another amphioxus species, *Asymmetron inferum*, survives in anaerobic and sulfide-rich bottom environments (229 m depth) [40]. Therefore, we suggest that amphioxi are tolerant to an anaerobic environment.

Alcohol dehydrogenase and ferredoxin

Alcohol dehydrogenases (ADHs) catalyze the oxidation of alcohols to aldehydes and ketones [41]. In the phylogenetic tree (Fig. 3C), an ADH gene containing zinc-binding sites and belonging to the conserved medium-chain dehydrogenase/reductase (MDR) family was much closer to those of the sponge *Amphimedon queenslandica*, the copepod *Eurytemora affinis*, and a bacterial group than to those of other metazoa. The ADH genes of the sponge and copepod were also confirmed to be HGT genes. The splice sites of the ADH gene of amphioxi are highly conserved with those of the sponge but not the copepod (Fig. 3D and S10), which suggested that the ADH genes of amphioxi and *Amphimedon queenslandica* originated from the same eukaryotic source. Among the 100 best hits of this ADH gene in the NCBI NR database, we noticed over ten hits from anaerobic bacteria, such as the ADH gene of *B. belcheri beihai*, which shared up to 64% identity with the alcohol dehydrogenase catalytic domain-containing protein of *Anaerolineae bacterium* (GenBank accession: MCD6286762.1).

Ferredoxin (FDX) contains an iron-sulfur cluster and is known as a mediator of electron transfer in many metabolic pathways [42]. A highly conserved FDX gene was identified as the HGT gene in amphioxi and was highly conserved with the genes in certain plants, bacteria, or other eukaryotes (identities over 68%) but not with that of other metazoa (excluding Amphioxiformes, Table 1). This FDX gene was tandemly duplicated into two copies on opposite strands in *B. floridae*. All these FDX genes contain a Ubl_FUBI (Ubiquitin-like protein FUBI) domain (Fig. S11A), which may protect the FDX gene from ubiquitination and subsequent degradation [43].

Genes encoding ferredoxins (FDXs), nitroreductases, NADH oxidase, and alcohol dehydrogenase (ADH) 3 have been suggested to be horizontally transferred from anaerobic prokaryotes to two intestinal protozoan parasites [44] and facilitate their survival under anaerobic conditions within the intestinal lumen. Collectively, we suggest that the HGT of the ADH and FDX genes may play similar roles in amphioxi, thereby facilitating their survival in low-oxygen and even anoxic environments. Similarly, under anoxic conditions, crucian carp and goldfish could present ethanol formation but not toxic lactate accumulation with muscle-specific ADH and novel pyruvate decarboxylase (PDC) [45].

In addition, a group of atypical short-chain dehydrogenase/reductase subgroup 4 (SDR_a4) genes were identified mostly proximally arrayed (distance <20 Kb) and divided into two clusters in four amphioxi (Fig. S11B and C). The SDR_a4 gene was tandemly duplicated in all four amphioxi (Fig. S11B) and could be divided into two clusters in the phylogenetic tree (Fig. S11C). This group of SDR_a4 genes that function in redox sensor mechanisms may also contribute to anaerobic adaptation [46, 47].

Therefore, we propose that amphioxus developed adaptation and tolerance to a low-oxygen environment through HGT, considering their niche adaptation in shallow subtidal sand bottom of seas. In addition, as an ancient animal has existed for over 500 million years [48], amphioxus experienced three events with extreme drops in oxygen concentration [49]. Since HGT events related to anaerobic adaptation are conserved among the four amphioxi in different oceans, they should have occurred before the geographical separation caused by the breakup of Pangaea (initiated

approximately 200 Ma) [50]. Therefore, we propose that the anaerobic adaptation-related HGT events were possibly the result of the ancient extreme drops in oxygen concentration that occurred from 200-500 million years ago, especially the two major drops that occurred in this period [49].

Calcium balance

Calcium balance and osmoregulation of fishes are important for a wide range of physiological functions [51]. A voltage-dependent calcium channel, EF-hand calcium-binding protein group and calcium-binding protein (CML26) were identified as HGT genes (Table 1).

Voltage-dependent calcium channel

An HGT gene of amphioxi was annotated as voltage-dependent calcium channel (VDCC, Table 1), which is a complex of several different subunits [52]; however, whether this gene functions independently as a calcium channel deserves further investigation. As shown in Fig. 3E, the VDCC genes of amphioxi, the hemichordate *Saccoglossus kowalevskii*, and three starfishes were located closer to those of other eukaryotes and plants, especially the VDCC genes of the choanoflagellate *Salpingoeca rosetta* (bootstrap value 53), whereas they were located far away from other metazoa, including many protostomes. The VDCC genes of amphioxi, hemichordate and three starfishes shared three conserved splice sites (Fig. 3F and S12), although no conserved splice sites are shared with choanoflagellate. This VDCC gene was proposed to be related to electrophysiological function and was widely expressed among tissues; however, it had no apparent tissue-specific expression (Fig. S4).

EF-hand calcium-binding proteins

EF-hand calcium-binding proteins (EF-CaBPs) may function in maintaining a high calcium concentration in the endoplasmic reticulum [53] and could be involved in bone formation [54]. Of the 73 newly identified EF-CaBP genes in four amphioxi, the bit scores of the best hits to bacteria were significantly higher than those to other metazoa (Fig. 4A and S13), which indicated that these EF-CaBPs were HGT genes originating from bacteria.

This group of EF-CaBP genes was proximally arrayed in four amphioxi and presented high copy number variation. *B. belcheri xiamen* has 25 EF-CaBP genes, *B. belcheri beihai* has 19 genes, *B. floridae* has 17 genes, and *B. lanceolatum* has only 12 genes. Marine salinity is linearly correlated with the calcium ion concentration [55]. Therefore, we further explored the relationship of the gene number variation of these proximally arrayed EF-CaBP genes and the sea surface salinity of the collection locations (Fig. 4B). The Mediterranean Sea has a much higher than average seawater salinity (approximately 35 parts per thousand) [56, 57], and the Atlantic Ocean is globally saltier than the Pacific Ocean [58]. Because of the absence of high-quality salinity data for these nearshore regions where the four amphioxi were obtained, we used the average sea surface salinity data in the analysis. The sea surface salinities of *B. belcheri beihai*, *B. belcheri xiamen*, *B. floridae* and *B. lanceolatum* were estimated as 27.65, 22.9, 34.7 and 37 PSU (practical salinity unit) respectively [59-62], and these values were linearly correlated with the gene number of the proximally arrayed EF-CaBP genes ($R^2 = 0.9041$, Fig. 4B). In addition, the calcium ion concentration could also be negatively related to the concentration of carbon dioxide [63]. This finding was consistent with the

greater gene number of these EF-CaBPs in *B. belcheri* and *B. floridae*, which were collected from nearshore areas and showed a higher population density compared with *B. lanceolatum* (Supplementary Note 1.1).

In the phylogenetic tree (Fig. 4C), except for the outlier gene *Bf_10*, the EF-CaBP genes were classified into two large groups, Clan A and Clan B. Clan A was further classified into three clusters, Cluster 1, Cluster 2 and Cluster 3. In Cluster 1, no genes of *B. floridae* were identified while eight genes of two *B. belcheri* were assigned to four pairs, and they included four highly conserved genes, *Bbb_2*, *Bbb_3*, *Bbx_21* and *Bbx_22*. However, three and two tandemly arrayed genes in Cluster 2 were identified in *B. floridae* and *B. lanceolatum*, respectively, and variable duplications were observed in Cluster 3. The phylogenetic analysis indicated that at least three genes in Clan A appeared in the ancestral amphioxus species. For Clan B, five relatively unique genes of *B. floridae*, *Bf_1-5* in Cluster 4, were proximally arrayed, and many paired genes were identified in two *B. belcheri*, such as three pairs in Cluster 5. All 73 EF-CaBPs were proximally arrayed in four amphioxi (Fig. 4D) and all spaces between genes were below 20 Kb. Notably, in the four amphioxi, four longer genes, namely, *Bbb_6*, *Bbx_18*, *Bf_11* and *Bl_6*, contained more EF-hand motifs (in Cluster 3 of Fig. 4C).

Except for the large group of proximally arrayed EF-CaBPs above, the calmodulin-like protein 26 (CML26) gene was a calcium-binding HGT gene with a shorter length (Table 1). This CML26 gene was tandemly duplicated in *B. floridae* and *B. lanceolatum* (Table 1). Initially, we considered the CML26 gene (Table 1) to belong to another HGT event, but the conserved splice sites (Fig. S14)

suggested that the CML26 gene and the large group of EF-CaBPs possibly originated from a single HGT event when no eukaryotic gene source was identified.

Along with the two calcium-binding protein CML26 genes, these EF-CaBP genes were apparently highly expressed in the gut and hepatic diverticulum, which are both considered digestive organs for calcium absorption (Fig. S4). Thus, the expansion of these EF-CaBP genes was proposed to occur in response to marine salinity and calcium concentration changes. The *EF-CaBP_Bl_11* and the two CML26 genes had the highest expression levels (Fig. S4).

Conclusions

As an important group of invertebrate chordates, amphioxi play a pivotal role in the evolutionary study of deuterostomes and chordates. In this study, the genome of the Beihai amphioxus *B. belcheri beihai* was *de novo* assembled and annotated (Table S1). Combined with the three amphioxus genomes (Table S1), we performed a phylogenetic analysis of the deuterostome genomes (Fig. 1 and S2), which supported the monophylies of deuterostomes and chordates.

We identified massive HGT events conserved among the four amphioxi (Table 1, Supplementary Note 3.1). Among these HGT events, except for the HGT of the TIR domains of amphioxi, which was mentioned in a previous report [64], all HGT genes were newly identified and reported in this study. Certain HGT genes had homologs from other deuterostome groups except for the jawed vertebrate group (Table 1 and S3). That is, the LOX gene had homologs from two jawless vertebrates; 8 HGT genes had homologs from either tunicates or echinoderms or hemichordates; and

only the VDCC gene had homologs from both echinoderms and hemichordates. The conservation of these HGT genes cannot be well explained as a common ancestor in the deuterostome phylogeny (Fig. 1) and provides additional but confusing information for the tree of life, which can be attributed to frequent gene gain and loss in the early and rapid evolution of deuterostomes (Table S3) [23]. For example, the LOX gene was suggested to be an HGT gene that occurred at the deuterostome stem but decayed in tunicates and jawed vertebrates. Although most HGT genes are uniquely conserved among four *Branchiostoma* species, the lack of high-quality genomic resource of other genera impeded our further comparison among cephalochordates. Those HGT events could be well dated according to their conservation among different deuterostome groups, such as three HGT genes conserved with urochordates (SCP24, MCO and UF-TM3) were proposed as results of HGT events at the chordate stem, while their absence in vertebrates is due to the selective gene lose.

The HGT genes were further analyzed in depth, thus providing additional insights into the early evolution of deuterostomes as follows. A wide range of adaptation novelties were enabled by these HGT genes, such as innate immunity-related HGT events that introduced more complexity to the immune system in amphioxi, especially for amphioxus without an adaptative immune system [25].

Proposed mechanisms of HGT

Despite a lack of evidence for the molecular mechanism, the HGT in amphioxi genomes could be explained by external fertilization based on the weak-link model, which suggests that the unicellular fertilized egg cells of amphioxi are the potential entry points for these foreign HGT genes [9, 65], especially when many potential HGT origin species are marine microbes, such as cyanobacteria.

Additionally, the HGT of amphioxus could be explained by the 'you are what you eat hypothesis' [66], which means that the HGT genes originated from the food eaten by amphioxus. All modern deuterostomes were considered to have evolved from a filter-feeding ancestor [8]. Along with other deuterostomes, such as tunicates, echinoderms, and hemichordates, amphioxus is known as a filter feeder [67, 68] and may obtain genes from marine microbes, including bacteria and algae, which are potential foods. For example, the charophyte alga *K. nitens* frequently appeared in the analysis of the HGT genes, including the DBH, SCP24 and UF-TM1 genes (Fig. 3A and B, Fig. S15 and Table 1). The HGT DBH gene, which originated from bacteria, is unusually identified in the charophyte alga *K. nitens* (Fig. 3A). The highly conserved splice sites among the DBH genes of the charophyte alga, soft coral, hemichordate, and amphioxi (Fig. 3B and S5) suggested that their DBH genes originated from a common ancient eukaryote, such as some marine algae. Although the charophyte alga *K. nitens* is considered a freshwater/terrestrial and nonmarine species [69], we proposed that the ancestor of *K. nitens* was possibly distributed in marine environments and served as food for these marine animals, especially because of the sister marine species *K. marinum* (no available genome sequences). Similarly, the SCP24 gene (Fig. S12) of amphioxi and tunicates and UF-TM1 gene (Table 1) possibly originated from the marine ancestral alga of *K. nitens*, which are the food source.

Therefore, those HGT genes conserved with the charophyte alga provided profound examples of genes horizontally transferred from plant to deuterostomes. Coincidently, the first HGT gene from plant to insect was recently reported in the whitefly *Bemisia tabaci* [70].

New gene generation after HGT

The origin of new genes is a fundamental process of genome evolution, and HGT enables the rapid generation of new genes [71]. In this study, the comparative analysis of HGT genes revealed insights into new gene generation. Regardless of whether common gene duplication and diversification occurred, many HGT genes of amphioxi underwent possible gene fusion or fission to generate new genes. Three novel models of new gene generation based on HGT were summarized, including gene fusion (Fig. S24A), gene fission (Fig. S24B) and partial duplication or deletion models (Fig. S24C).

In the gene fusion model, we revealed that two distinct TIR domains fused with a wide range of endogenous genes (or domains) with immune functions (Fig. S9) and generated a series of new immune genes. In addition, the FDX gene fused with an endogenous Ubl_FUBI domain (Fig. S11A). Both gene fission and partial duplication or deletion were observed in the LOX family of amphioxi (Fig. 2B). Since we found that the LOX genes of amphioxi were mostly similar to bacterial LOX genes containing up to three SRPBCC domains and three groups of SRPBCC domains could be clustered with three SRPBCC domains of the cyanobacterium *Coleofasciculus sp. FACHB-125* (GenBank accession: WP_190414990.1) (Fig. S6), we considered that the original LOX gene of amphioxi contains three SRPBCC domains. Then, they were split into LOX genes with only C2 domains and genes with only SRPBCC domain via gene fission (Fig. S24B) and generated a series of LOX genes with variable numbers of SRPBCC domains (Fig. 2B) via partial gene duplication or deletion (Fig. S24C).

Limitations of this study

During the identification of HGT events, we ignored those candidates with homologous genes in jawed vertebrates. Although this process could filter false-positive candidates to a large extent, it may miss some important genes. Therefore, the ancestral gene novelties remaining among all deuterostome groups were not explored or discussed in this study. Meanwhile, the filter standard of HGT identification may be too high to achieve highly comprehensive results, and many HGT genes may be missed during filtration. Moreover, the lack of animal model experiments impeded our experimental validation of the metabolic functions of these HGT genes.

Declarations

Availability of data and materials

All data are accessible under NCBI BioProject accession: PRJNA804338.

Competing interests

We have no competing interest to disclose.

Funding

This work was made possible by grants from the General Research Fund from Research Grants Council (Reference numbers: 464710, 475113, 14119219, 14119420, 14175617) and Health and Medical Research Fund from Food and Health Bureau (Reference numbers: 06171016, 07181266) in Hong Kong.

Authors' contributions

Qing Xiong: Investigation, Methodology, Conceptualization, Writing–Original Draft.

Kevin Yi Yang: Investigation, Conceptualization, Methodology.

Xi Zeng: Investigation, Data curation, Methodology.

Mingqiang Wang: Investigation, Data curation, Methodology.

Patrick Kwok-Shing Ng: Investigation, Methodology.

Jun-Wei Zhou: Investigation, Methodology.

Judy Kin-Wing Ng: Investigation, Data curation.

Cherie Tsz-Yiu Law: Investigation, Methodology.

Qiao Du: Data curation, Methodology.

Kejin Xu: Data curation, Methodology.

Laura Jane Falkenberg: Investigation, Conceptualization.

Bingyu Mao: Investigation, Resources.

Jun-Yuan Chen: Investigation, Resources, Supervision.

Stephen Kwok-Wing Tsui: Funding acquisition, Conceptualization, Project administration, Writing–Original Draft, Writing–Review & Editing.

References

1. Keeling PJ, Palmer JD: **Horizontal gene transfer in eukaryotic evolution.** *Nature Reviews Genetics* 2008, **9:**605-618.
2. Soucy SM, Huang J, Gogarten JP: **Horizontal gene transfer: building the web of life.** *Nature Reviews Genetics* 2015, **16:**472-482.
3. Husnik F, McCutcheon JP: **Functional horizontal gene transfer from bacteria to eukaryotes.** *Nature Reviews Microbiology* 2018, **16:**67-79.
4. Nakabachi A: **Horizontal gene transfers in insects.** *Curr Opin Insect Sci* 2015, **7:**24-29.
5. Drezen JM, Gauthier J, Josse T, Bézier A, Herniou E, Huguet E: **Foreign DNA acquisition by invertebrate genomes.** *Journal of Invertebrate Pathology* 2017, **147:**157-168.
6. Dunning Hotopp JC: **Grafting or pruning in the animal tree: lateral gene transfer and gene loss?** *BMC Genomics* 2018, **19:**470.
7. Xiong Q, Wan AT-Y, Liu X, Fung CS-H, Xiao X, Malainual N, Hou J, Wang L, Wang M, Yang KY, et al: **Comparative Genomics Reveals Insights into the Divergent Evolution of Astigmatic Mites and Household Pest Adaptations.** *Molecular Biology and Evolution* 2022, **39**.
8. Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F, et al: **Hemichordate genomes and deuterostome origins.** *Nature* 2015, **527:**459-465.
9. Graham LA, Li J, Davidson WS, Davies PL: **Smelt was the likely beneficiary of an antifreeze gene laterally transferred between fishes.** *BMC Evolutionary Biology* 2012, **12:**190.

10. Stanhope MJ, Lupas A, Italia MJ, Koretke KK, Volker C, Brown JR: **Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates.** *Nature* 2001, **411:**940-944.

11. Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G: **Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes.** *Genome Biology* 2015, **16:**50.

12. Salzberg SL: **Horizontal gene transfer is not a hallmark of the human genome.** *Genome Biology* 2017, **18:**85.

13. Nakashima K, Yamada L, Satou Y, Azuma J, Satoh N: **The evolutionary origin of animal cellulose synthase.** *Dev Genes Evol* 2004, **214:**81-88.

14. Li K-L, Nakashima K, Inoue J, Satoh N: **Phylogenetic Analyses of Glycosyl Hydrolase Family 6 Genes in Tunicates: Possible Horizontal Transfer.** *Genes* 2020, **11:**937.

15. Syvanen M, Ducore J: **Whole genome comparisons reveals a possible chimeric origin for a major Metazoan assemblage.** *Journal of Biological Systems* 2010, **18:**261-275.

16. Boto L: **Horizontal gene transfer in the acquisition of novel traits by metazoans.** *Proceedings Biological sciences* 2014, **281:**20132450-20132450.

17. Bányai L, Kerekes K, Trexler M, Patthy L: **Morphological Stasis and Proteome Innovation in Cephalochordates.** *Genes* 2018, **9:**353.

18. Yang KY, Chen Y, Zhang Z, Ng PK-S, Zhou WJ, Zhang Y, Liu M, Chen J, Mao B, Tsui SK-W: **Transcriptome analysis of different developmental stages of amphioxus reveals dynamic changes of distinct classes of genes during development.** *Scientific Reports* 2016, **6:**23195.

19. Huang S, Chen Z, Yan X, Yu T, Huang G, Yan Q, Pontarotti PA, Zhao H, Li J, Yang P, et al: **Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes.** *Nat Commun* 2014, **5:**5896.

20. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, et al: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453:**1064-1071.

21. Marlétaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la Calle-Mustienes E, Bertrand S, Burguera D, et al: **Amphioxus functional genomics and the origins of vertebrate gene regulation.** *Nature* 2018, **564:**64-70.

22. Satoh N, Rokhsar D, Nishikawa T: **Chordate evolution and the three-phylum system.** *Proc Biol Sci* 2014, **281:**20141729.

23. Kapli P, Natsidis P, Leite DJ, Fursman M, Jeffrie N, Rahman IA, Philippe H, Copley RR, Telford MJ: **Lack of support for Deuterostomia prompts reinterpretation of the first Bilateria.** *Science Advances* 2021, **7:**eabe2741.

24. Yue J-X, Holland ND, Holland LZ, Deheyn DD: **The evolution of genes encoding for green fluorescent proteins: insights from cephalochordates (amphioxus).** *Scientific Reports* 2016, **6:**28350.

25. Bartl S, Baish M, Weissman IL, Diaz M: **Did the Molecules of Adaptive Immunity Evolve from the Innate Immune System?** *Integrative and Comparative Biology* 2003, **43:**338-346.

26. Yuan D, Zou Q, Yu T, Song C, Huang S, Chen S, Ren Z, Xu A: **Ancestral genetic complexity of arachidonic acid metabolism in Metazoa.** *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 2014, **1841:**1272-1284.

27. Yuan D, Xu A: **Chapter 10 - Evolution of Inflammation-Related Lipid Metabolism Pathway and Inflammatory Lipid Signaling in Amphioxus.** In *Amphioxus Immunity.* Edited by Xu A: Academic Press; 2016: 189-205

28. Janvier P: **Modern look for ancient lamprey.** *Nature* 2006, **443:**921-923.

29. Chen H-m, Zhu Z-j, Chen J-J, Yang R, Luo Q-j, Xu J-l, Shan H, Yan X-J: **A multifunctional lipoxygenase from Pyropia haitanensis— The cloned and functioned complex eukaryotic algae oxylipin pathway enzyme.** *Algal Research* 2015, **12:**316-327.

30. Taschuk F, Cherry S: **DEAD-Box Helicases: Sensors, Regulators, and Effectors for Antiviral Defense.** *Viruses* 2020, **12**.

31. Bhattacharya D, Agrawal S, Aranda M, Baumgarten S, Belcaid M, Drake JL, Erwin D, Foret S, Gates RD, Gruber DF, et al: **Comparative genomics explains the evolutionary success of reef-forming corals.** *eLife* 2016, **5:**e13288.

32. Rotskaya UN, Kryukov VY, Kosman E, Tyurin M, Glupov VV: **Identification of the Ricin-B-Lectin LdRBLk in the Colorado Potato Beetle and an Analysis of Its Expression in Response to Fungal Infections.** *J Fungi (Basel)* 2021, **7**.

33. Pawson T, Nash P: **Assembly of cell regulatory systems through protein interaction domains.** *Science* 2003, **300:**445-452.

34. Dishaw LJ, Haire RN, Litman GW: **The amphioxus genome provides unique insight into the evolution of immunity.** *Briefings in functional genomics* 2012, **11:**167-176.

35. Han Y, Huang G, Zhang Q, Yuan S, Liu J, Zheng T, Fan L, Chen S, Xu A: **The primitive immune system of amphioxus provides insights into the ancestral structure of the vertebrate immune system.** *Developmental & Comparative Immunology* 2010, **34:**791-796.

36. Trowbridge CD, Davenport J, Cottrell DM, Harman L, Plowman CQ, Little C, McAllen R: **Extreme oxygen dynamics in shallow water of a fully marine Irish sea lough.** *Regional Studies in Marine Science* 2017, **11:**9-16.

37. Carvalho JE, Lahaye F, Schubert M: **Keeping amphioxus in the laboratory: an update on available husbandry methods.** *Int J Dev Biol* 2017, **61:**773-783.

38. Schmitz A, Gemmel M, Perry SF: **Morphometric partitioning of respiratory surfaces in amphioxus (Branchiostoma lanceolatum Pallas).** *Journal of Experimental Biology* 2000, **203:**3381-3390.

39. Briggs DE, Kear AJ: **Decay of Branchiostoma: implications for soft-tissue preservation in conodonts and other primitive chordates.** *Lethaia* 1993, **26:**275-287.

40. Kon T, Nohara M, Yamanoue Y, Fujiwara Y, Nishida M, Nishikawa T: **Phylogenetic position of a whale-fall lancelet (Cephalochordata) inferred from whole mitochondrial genome sequences.** *BMC Evolutionary Biology* 2007, **7:**127.

41. Di L, Balesano A, Jordan S, Shi SM: **The role of alcohol dehydrogenase in drug metabolism: beyond ethanol oxidation.** *The AAPS Journal* 2021, **23:**1-21.

42. Meyer J: **The evolution of ferredoxins.** *Trends in Ecology & Evolution* 1988, **3:**222-226.

43. Rossman TG, Visalli MA, Komissarova EV: **fau and its ubiquitin-like domain (FUBI) transforms human osteogenic sarcoma (HOS) cells to anchorage-independence.** *Oncogene* 2003, **22:**1817-1821.

44. Nixon JE, Wang A, Field J, Morrison HG, McArthur AG, Sogin ML, Loftus BJ, Samuelson J: **Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to Giardia lamblia**

and Entamoeba histolytica. *Eukaryot Cell* 2002, **1:**181-190.

45. Fagernes CE, Stensløkken K-O, Røhr ÅK, Berenbrink M, Ellefsen S, Nilsson GE: **Extreme anoxia tolerance in crucian carp and goldfish through neofunctionalization of duplicated genes creating a new ethanol-producing pyruvate decarboxylase pathway.** *Scientific Reports* 2017, **7:**7884.

46. Kavanagh KL, Jörnvall H, Persson B, Oppermann U: **Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes.** *Cell Mol Life Sci* 2008, **65:**3895-3906.

47. Green J, Paget MS: **Bacterial redox sensors.** *Nature Reviews Microbiology* 2004, **2:**954-966.

48. Yuan S, Ruan J, Huang S, Chen S, Xu A: **Amphioxus as a model for investigating evolution of the vertebrate immune system.** *Developmental & Comparative Immunology* 2015, **48:**297-305.

49. Berner RA, VandenBrooks JM, Ward PD: **Oxygen and Evolution.** *Science* 2007, **316:**557-558.

50. Dietz RS, Holden JC: **The breakup of Pangaea.** *Scientific American* 1970, **223:**30-41.

51. Guerreiro PF, Juan: **Control of calcium balance in fish.** In *Fish Osmoregulation.* USA: Science Publisher; 2007

52. Dolphin AC: **A short history of voltage-gated calcium channels.** *British journal of pharmacology* 2006, **147 Suppl 1:**S56-S62.

53. Camacho P, John L, Li Y, Paredes RM, Roderick HL: **Calnexin and calreticulin, ER associated modulators of calcium transport in the ER.** In *Calreticulin.* Springer; 2003:

126-132

54. Allen LH, Hall TE: **Calcium Metabolism, Intestinal Calcium-binding Protein, and Bone Growth of Rats Fed High Protein Diets.** *The Journal of Nutrition* 1978, **108:**967-972.

55. He H, Li Y, Wang S, Ma Q, Pan Y: **A High Precision Method for Calcium Determination in Seawater Using Ion Chromatography.** *Frontiers in Marine Science* 2020, **7**.

56. Millot C, Taupier-Letage I: **Circulation in the Mediterranean Sea.** In *The Mediterranean Sea.* Edited by Saliot A. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005: 29-66

57. Millero FJ, Feistel R, Wright DG, McDougall TJ: **The composition of Standard Seawater and the definition of the Reference-Composition Salinity Scale.** *Deep Sea Research Part I: Oceanographic Research Papers* 2008, **55:**50-72.

58. Jones CS, Cessi P: **Size Matters: Another Reason Why the Atlantic Is Saltier than the Pacific.** *Journal of Physical Oceanography* 2017, **47:**2843-2859.

59. **Tidal Flat Planning of Aquaculture Waters in Dianbai District, Maoming City.** China: The Government of Dianbai District, Maoming City; 2019.

60. **The marine environment of Xiamen City.** China: The Government of Xiamen City; 2016.

61. Meyers SD, Linville AJ, Luther ME: **Alteration of residual circulation due to large-scale infrastructure in a coastal plain estuary.** *Estuaries and Coasts* 2014, **37:**493-507.

62. Dańko A, Schaible R, Dańko MJ: **Salinity Effects on Survival and Reproduction of Hydrozoan Eleutheria dichotoma.** *Estuaries and Coasts* 2020, **43:**360-374.

63. Richards M, Xu W, Mallozzi A, Errera RM, Supan J: **Production of Calcium-Binding Proteins in Crassostrea virginica in Response to Increased Environmental CO2 Concentration.** *Frontiers in Marine Science* 2018, **5**.

64. Huang S, Yuan S, Guo L, Yu Y, Li J, Wu T, Liu T, Yang M, Wu K, Liu H, et al: **Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity.** *Genome research* 2008, **18:**1112-1126.

65. Huang J: **Horizontal gene transfer in eukaryotes: the weak-link model.** *BioEssays : news and reviews in molecular, cellular and developmental biology* 2013, **35:**868-875.

66. Doolittle WF: **You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes.** *Trends in Genetics* 1998, **14:**307-311.

67. Riisgård HU, Svane I: **Filter Feeding in Lancelets (Amphioxus), Branchiostoma lanceolatum.** *Invertebrate Biology* 1999, **118:**423-432.

68. Peterson KJ, Eernisse DJ: **The phylogeny, evolutionary developmental biology, and paleobiology of the Deuterostomia: 25 years of new techniques, new discoveries, and new ideas.** *Organisms Diversity & Evolution* 2016, **16:**401-418.

69. Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, Sato S, Yamada T, Mori H, Tajima N, et al: **Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation.** *Nature Communications* 2014, **5:**3978.

70. Xia J, Guo Z, Yang Z, Han H, Wang S, Xu H, Yang X, Yang F, Wu Q, Xie W, et al: **Whitefly hijacks a plant detoxification gene that neutralizes plant toxins.** *Cell* 2021, **184:**1693-1705.e1617.

71. Long M, Betrán E, Thornton K, Wang W: **The origin of new genes: glimpses from the young and old.** *Nature Reviews Genetics* 2003, **4:**865-875.

Table 1. Horizontal gene transfer events identified in amphioxus

Functional category	Abbrev.	Representative gene ID①	Amino acid length	Best hit, metazoan②	Best hit③			Copy number④			
					E-value	Identity	Taxonomy	*Bbb*	*Bbx*	*Bf*	*Bl*
Innate immunity	LOX*	Bbb_152860.1⑤	696	3E-62	7E-168	39.6%	Bacteria, cyanobacteria/proteobacteria	6	5	9	7
	DBH*	Bbb_034850.1	1253	2E-141	0.0	39.1%	Bacteria, cyanobacteria/proteobacteria	1	1	1	1
	RBL	Bbb_161350.1	282	8E-17	4E-23	29.7%	Bacteria, Terrabacteria group	2	1	3	1
	TIR-a	Bbb_152643.1⑥	126	1E-4	4E-21	42.5%	Bacteria, *Solibacter usitatus*	4	5	2	2
	TIR-b	Bbb_258280.1⑥	95	N. A.	1E-10	43.3%	Bacteria, cyanobacteria	11	14	18	9
Anaerobic adaptation	ADH*	Bbb_340330.1	355	3E-78	1E-129	51.3%	Bacteria, Gammaproteobacteria	1	1	1	1
	FDX	Bbb_233460.1	171	1E-12	6E-46	74.5%	Plants, *Micractinium conductrix*	1	1	2	1
	SDR_a4	Bbb_177610.1	263	N. A.	5E-35	34.6%	Bacteria, *Methylomarinum vadi*	3	3	3	2
Calcium balance	VDCC*	Bbb_084320.1	524	3E-36	9E-83	37.2%	Other eukaryotes, *Salpingoeca rosetta*	1	1	1	1
	EF-CaBP	Bbb_077420.1	248	2E-6	5E-11	27.4%	Bacteria, Gammaproteobacteria	19	25	17	12
	CML26	Bbb_165620.1	153	4E-6	3E-11	36.2%	Bacteria, Gammaproteobacteria	1	1	2	2
Amino acid metabolism	SCP24*	Bbb_236620.1	474	1E-83	8E-136	46.6%	Plants, *Pycnococcus provasolii*	1	1	1	1
	GSA*	Bbb_006430.1	438	2E-32	2E-113	53.6%	Bacteria, Deltaproteobacteria	2	1	1	1
	GNAT	Bbb_266880.1	325	3E-26	7E-60	39.3%	Bacteria, Alphaproteobacteria	3	5	5	5
	DDC*	Bbb_001110.1	586	4E-49	2E-141	43.4%	Bacteria, *Bacteroidetes bacterium*	9	8	16	7
	DOD	Bbb_303310.1	173	3E-8	3E-36	42.5%	Other eukaryotes, *Tetrahymena*	2	1	1	1
	FR	Bbb_207740.1	241	9E-24	3E-72	54.0%	Other eukaryotes, Chlorarachniophyte	2	3	5	4
	DHFR	Bbb_115460.1	186	2E-29	1E-43	44.0%	Bacteria, *Ignavibacteriae bacterium*	1	1	1	1
Other metabolic genes	GH27⑦	Bbb_359430.1	419	8E-32	6e-82	40.1%	Bacteria, Terrabacteria group	1	1	1	1
	MCO*	Bbb_246040.1	640	4E-10	7E-67	30.7%	Plants, *Tetraselmis astigmatica*	5	3	8	4
	anmK	Bbb_029670.1	292	7E-29	2E-51	52.2%	Bacteria, cyanobacteria	1	1	1	1
Others	TPR	Bbb_379800.1	884	1E-86	0.0	37.6%	Other eukaryotes, Amoeba	1	1	1	1
	ApoLp	Bbb_173200.1	1205	N. A.	1E-158	36.8%	Bacteria, Gamma/Deltaproteobacteria	3	2	3	3
	HintN	Bbb_173190.1	789	2E-5	1E-118	36.3%	Bacteria, Gamma/Deltaproteobacteria	3	1	2	3
	CytP460	Bbb_256630.1	435	N. A.	2E-94	41.3%	Fungi, *Punctularia strigosozonata*	5	3	8	2
	UF-TM1	Bbb_158830.1	341	N. A.	6E-39	35.0%	Plants, *Klebsormidium nitens*	6	6	7	12
	UF-TM2*	Bbb_202120.1	1108	8E-7	4E-34	23.9%	Fungi, *Aspergillus thermomutatus*	1	1	1	1
	UF-TM3*	Bbb_320160.1	268	N. A.	1E-8	29.8%	Bacteria, *Streptomyces*	12	10	6	7

① To describe the gene features, a gene of *B. belcheri beihai* was selected as the representative gene. The protein sequences are listed in Table S4.

② The best hits of metazoan were of matches to genes of metazoa (excluding Amphioxiformes) in UniRef50 database. N.A., Not applicable, because no best hit has E-value smaller than the cutoff 0.001. In some cases, the HGT gene has homologs in other metazoa species and marked in star (*); then these metazoa species were excluded.

③ The best hits in UniRef50 database were listed and displayed in detail. In many cases, the best matches were from a group of microbes, but not exactly a species, so the group name like cyanobacteria was listed.

④ Since many HGT genes underwent gene duplications, copy number of four amphioxi were compared. Bbb, *B. belcheri beihai*; Bbx, *B. belcheri xiamen*; Bf, *B. floridae*; Bl, *B. lanceolatum*.

⑤ For the LOX gene Bbb_152860.1 contains many domains, only the partial 696-aa sequence containing its LOX domain was analyzed.

⑥ For the TIR-a and TIR-b, only their TIR domains were analyzed, which covered 35-160 of Bbb_152640.3 and 107-201 of Bbb_258280.1, respectively.

⑦ For the GH27 gene Bbb_359430.1, the best hit in NR database (6E-82 from Terrabacteria) has a much lower E-value than the best hit in UniRef50 database (6E-74 from plants), therefore we chose to present the best hit in NR database.

Abbreviation:

Innate immunity

LOX* lipoxygenase; homologs in the sea lamprey *Petromyzon marinus* and the hagfish *Eptatretus burger*;
DBH* DEAD/DEAH-box helicase; homologs in the soft coral *Dendronephthya gigantea* and hemichordates;
RBL ricin B-like lectin;
TIR-a Toll/interleukin-1 receptor domain a;
TIR-b Toll/interleukin-1 receptor domain b;

Anaerobic adaptation

ADH* alcohol dehydrogenase; homolog in the sponge *Amphimedon queenslandica*;
FDX ferredoxin;
SDR_a4 atypical SDR subgroup 4, containing a glycine-rich NAD(P)-binding motif;

(to be continued)

Calcium balance

VDCC *	voltage-dependent calcium channel (subunit alpha-2/delta-4-like); homologs in hemichordates and echinoderms;
EF-CaBP	EF-hand calcium-binding protein;
CML26	calmodulin-like protein 26;

Amino acid metabolism

SCP24 *	serine carboxypeptidase 24; homologs in tunicates;
GSA *	glutamate-1-semialdehyde aminotransferase; homologs in echinoderms;
GNAT	GNAT family N-acetyltransferase;
DDC *	DOPA decarboxylase; homologs in hemichordates;
DOD	DOPA 4,5-dioxygenase;
FR	folate receptor;
DHFR	dihydrofolate reductase;

Other metabolic genes

GH27	glycoside hydrolase family 27;
MCO *	multicopper oxidase; homologs in tunicates;
anmK	anhydro-N-acetylmuramic acid kinase;

Others

TPR	tetratricopeptide repeat protein;
ApoLp	apolipophorin III-like domain-containing protein;
HintN	Hedgehog/Intein domain-containing protein;
CytP460	cytochrome P460 family like protein;
UF-TM1	transmembrane protein 1 of unknown function;
UF-TM2 *	transmembrane protein 2 of unknown function; homologs in hemichordates;
UF-TM3 *	transmembrane protein 3 of unknown function; homologs in tunicates.

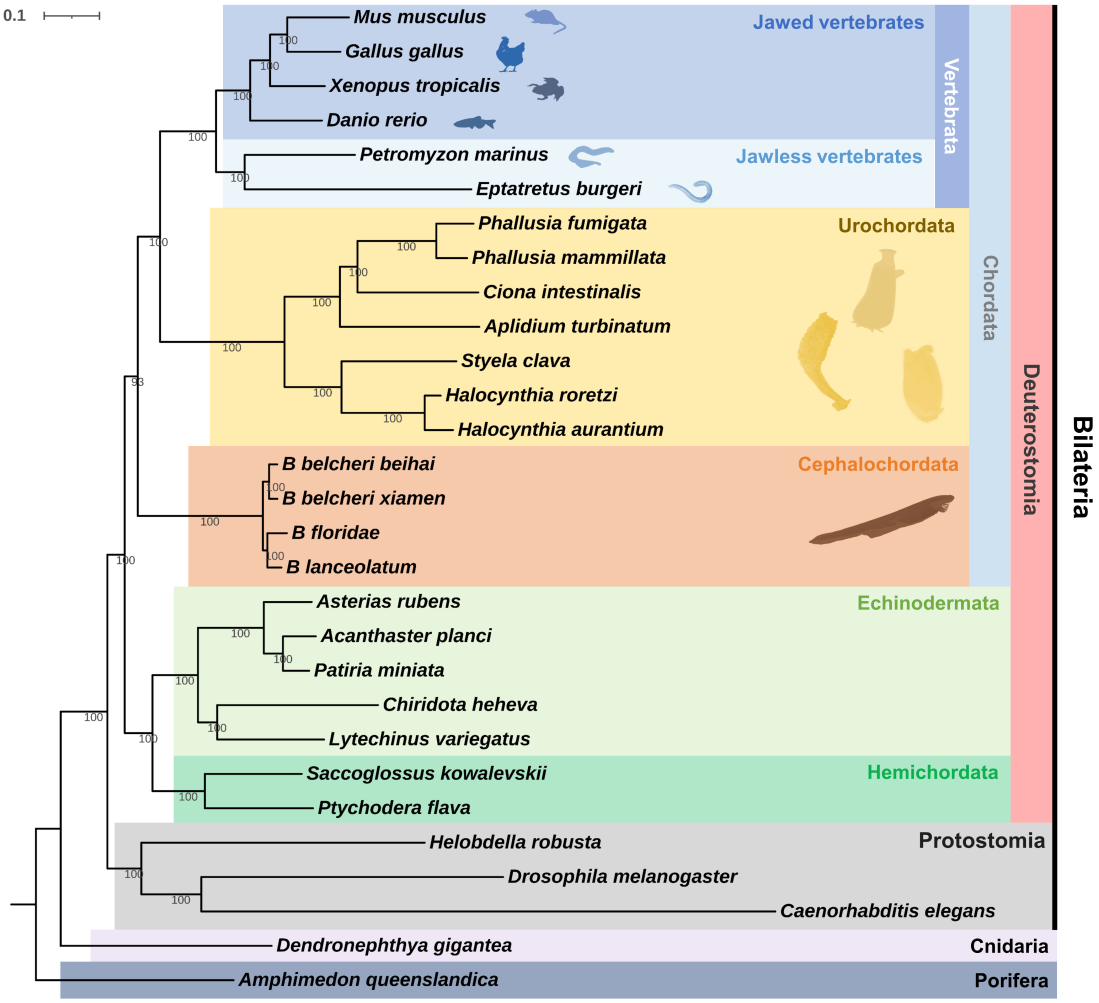

Fig. 1. Phylogenetic analysis of major deuterostome groups
The completeness of 29 genome assemblies (Table S2) was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO) using database metazoa_odb9. 112 single-copy BUSCOs overlapped by 29 genome assemblies were extracted in protein sequences and aligned by MAFFT for the phylogenetic analysis by RAxML (Supplementary Note 2.2).

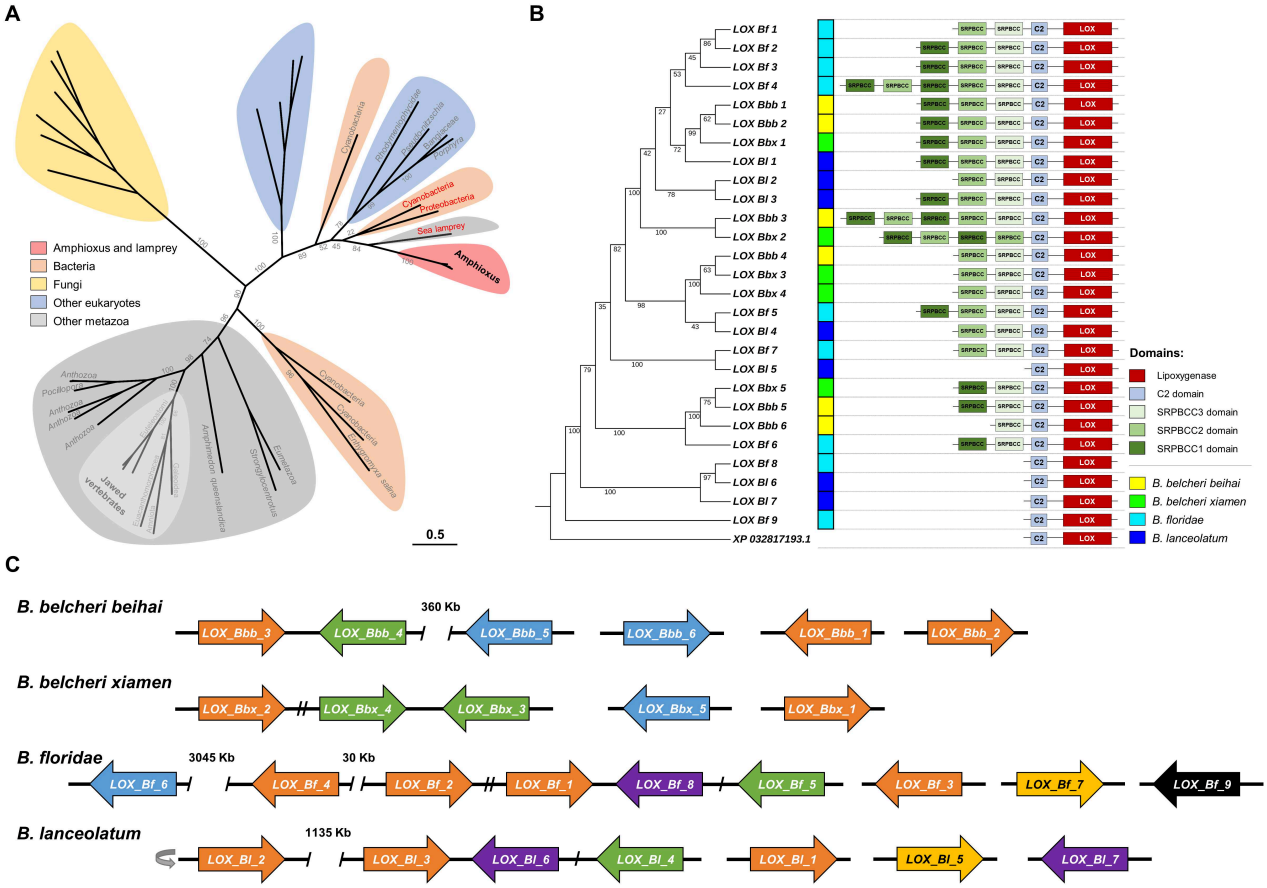

Fig. 2. HGT analysis of a group of lipoxygenases (LOXs) in amphioxi
(A) The amphioxus LOX genes gained via HGT and their closest genes in the UniRef50 database from other taxonomic categories, including bacteria, fungi, other eukaryotes (excluding metazoa, plants, and fungi) and other metazoa (excluding Amphioxiformes) were collected for phylogenetic analysis. **(B)** Phylogenetic analysis of all LOX genes of four amphioxi and their domain architecture. The LOX gene of the sea lamprey *Petromyzon marinus* (GenBank accession: XP_032817193.1) was used as outgroup. C2, C2-like domain (universal calcium-binding domain); SRPBCC, PYR/PYL/RCAR family domain in SRPBCC superfamily. **(C)** Gene synteny alignment of the LOX genes of four amphioxi. The distance precision was set as 5 Kb, e.g., 30 Kb represents for 30-35 Kb. If not noted, a slash (/) between two genes represents 5 Kb, e.g., one slash represents 5-10 Kb. Except the outlier (*LOX_Bf_9*), LOXs located on clades supported by high bootstrap value (>95) in the phylogenetic tree were painted in the same color. The suffixes stand for four amphioxi, *Bbb* for *B. belcheri beihai*, *Bbx* for *B. beicheri xiamen*, *Bf* for *B. floridae* and *Bl* for *B. lanceolatum*. The grey turnover arrow means after reverse complement.

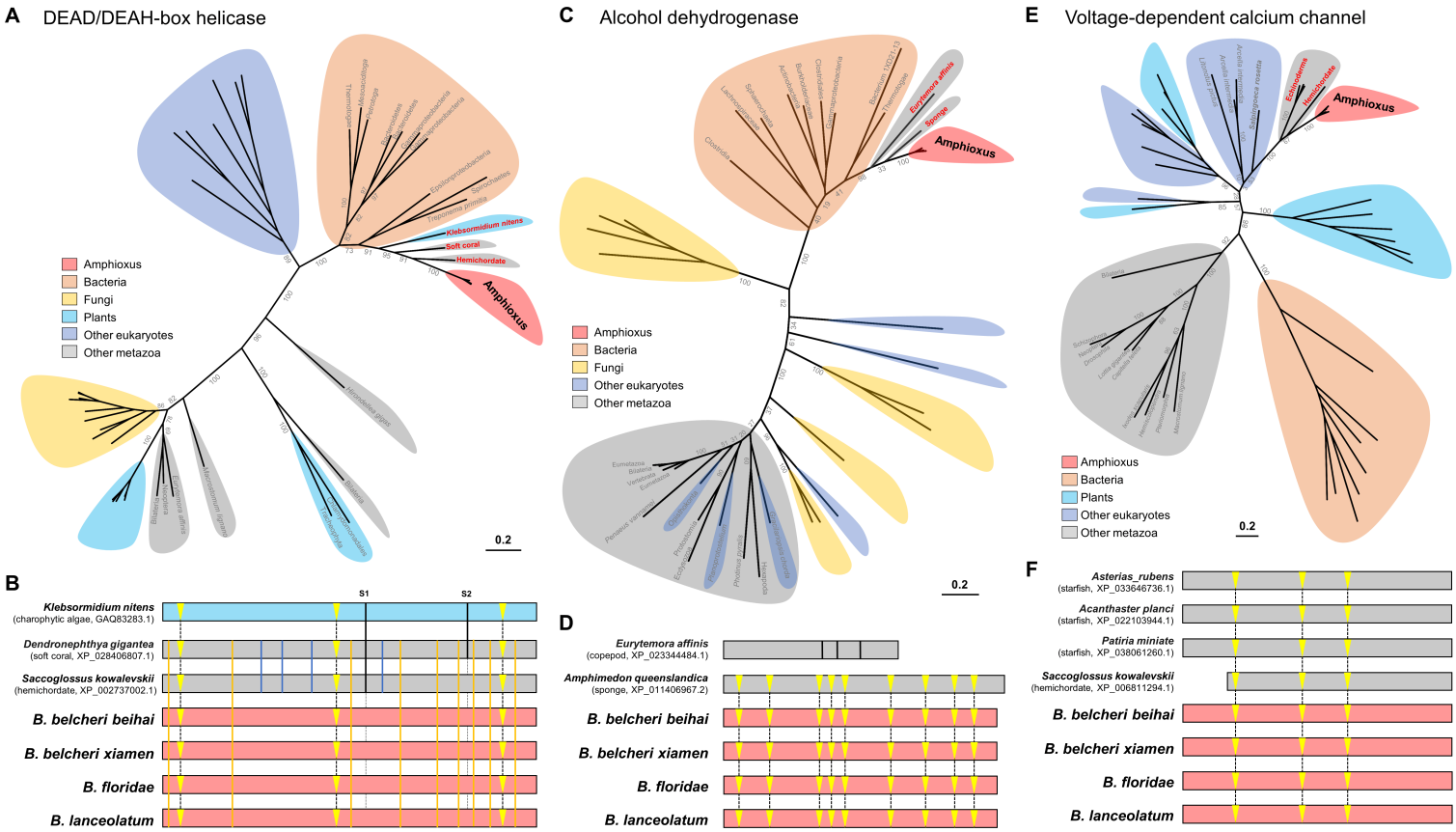

Fig. 3. HGT analysis of a DEAD/DEAH-box helicase (DBH), an alcohol dehydrogenase (ADH) and a voltage-dependent calcium channel (VDCC)

(A) The DBHs of amphioxi gained via HGT and their closest genes in the UniRef50 database from other taxonomic categories, including bacteria, fungi, other eukaryotes (excluding metazoa, plants, and fungi) and other metazoa (excluding Amphioxiformes) were collected for phylogenetic analysis. **(B)** Alignment of splice sites of the DBH genes of amphioxi, the charophytic algae *K. nitens* (GenBank accession: GAQ83283.1), the soft coral *Dendronephthya gigantea* (GenBank accession: XP_028406807.1) and the hemichordate *Saccoglossus kowalevskii* (GenBank accession: XP_002737002.1). Conserved splice sites were marked by vertical lines (Fig. S7), and three conserved splice sites of all seven sequences were highlighted in yellow triangles. Two unusual splice sites were labeled as S1 and S2. **(C)** The ADHs gained via HGT in amphioxi and their closest genes in the UniRef50 database from other taxonomic categories were collected for phylogenetic analysis. **(D)** Alignment of splice sites of the ADH genes of amphioxi, the sponge *Amphimedon queenslandica* (GenBank accession: XP_011406967.2) and the copepod *Eurytemora affinis* (UniProt accession: UPI000C762174) (Fig. S10). **(E)** The VDCCs gained via HGT in amphioxi and their closest genes in the UniRef50 database from other taxonomic categories were collected for phylogenetic analysis. **(F)** Alignment of splice sites of the VDCC genes of amphioxi, the hemichordate *Saccoglossus kowalevskii* (GenBank accession: XP_006811294.1) and three starfishes (GenBank accessions: XP_033646736.1, XP_022103944.1 and XP_038061260.1) (Fig. S12).

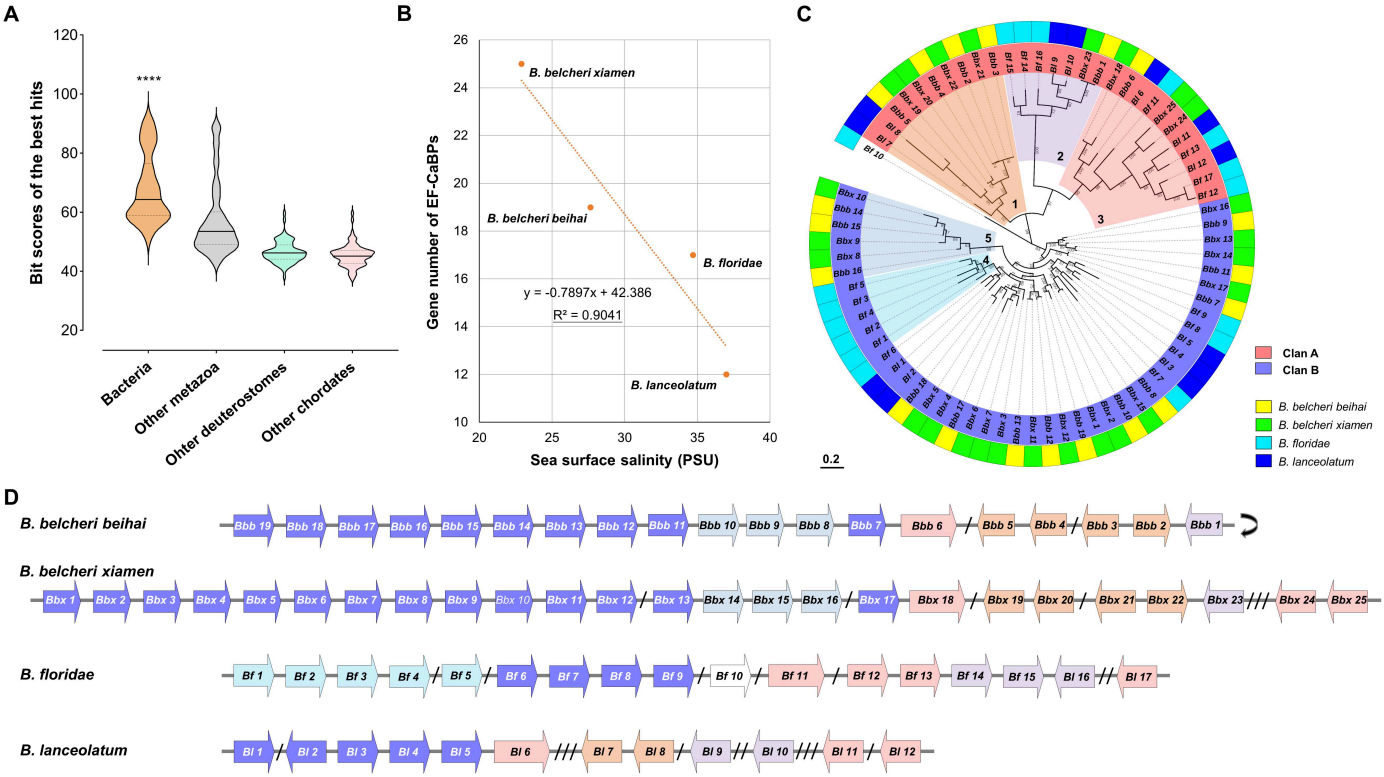

Fig. 4. A group of proximally arrayed EF-hand calcium-binding proteins (EF-CaBPs) **(A)** Of a group of EF-CaBP genes of amphioxus, bit scores of the best hits in the UniRef50 database to other metazoa, other deuterostomes and other chordates (excluding Amphioxiformes respectively) were compared with those to bacteria in violin plot with quartile lines. **** p <0.0001, when comparing the best hits to bacteria with other three groups, respectively. **(B)** Linear correlation of the gene number and the sea surface salinity of collection locations of four amphioxi. Linear regression was performed in $R^2 = 0.9041$. **(C)** Phylogenetic analysis of these EF-CaBPs of four amphioxi. Two large clans, Clan A and B, were highlighted and five specific clusters were marked. **(D)** Gene synteny alignment of the proximally arrayed EF-CaBP genes. A slash (/) represents 5 Kb, e.g., one slash represents 5-10 Kb, and three slashes represent 15-20 Kb, while none represents below 5 Kb. The grey turnover arrow means after reverse complement.