EST and transcriptome analysis of cephalochordate amphioxus—past, present and future
Yu-Bin Wang, Shu-Hwa Chen, Chun-Yen Lin and Jr-Kai Yu
Advance Access publication date 4 February 2012

Abstract
The cephalochordates, commonly known as amphioxus or lancelets, are now considered the most basal chordate group, and the studies of these organisms therefore offer important insights into various levels of evolutionary biology. In the past two decades, the investigation of amphioxus developmental biology has provided key knowledge for understanding the basic patterning mechanisms of chordates. Comparative genome studies of vertebrates and amphioxus have uncovered clear evidence supporting the hypothesis of two-round whole-genome duplication thought to have occurred early in vertebrate evolution and have shed light on the evolution of morphological novelties in the complex vertebrate body plan. Complementary to the amphioxus genome-sequencing project, a large collection of expressed sequence tags (ESTs) has been generated for amphioxus in recent years; this valuable collection represents a rich resource for gene discovery, expression profiling and molecular developmental studies in the amphioxus model. Here, we review previous EST analyses and available cDNA resources in amphioxus and discuss their value for use in evolutionary and developmental studies. We also discuss the potential advantages of applying high-throughput, next-generation sequencing (NGS) technologies to the field of amphioxus research.

Keywords: amphioxus; EST; evolution; development; transcriptome; genome

INTRODUCTION
Amphioxus (also known as lancelets) are a group of marine invertebrate animals belonging to the subphylum Cephalochordata. Together with Tunicata and Vertebrata, they constitute the phylum Chordata. All chordates share several key characteristics, including a dorsal nerve chord, a notochord, segmented somites and pharyngeal gill slits. Unlike most tunicates whose larvae undergo drastic metamorphosis and lose many chordate characteristics as they become sessile in adulthood, the adult body plan of amphioxus remains highly similar to that of vertebrates (Figure 1A). Furthermore, the embryology of amphioxus is more comparable to that of vertebrates [1]. Therefore, cephalochordates were traditionally considered the closest relatives to vertebrates. Recent molecular phylogenetic analyses, however, have indicated that tunicates are in fact the closest sister group to vertebrates and that cephalochordates actually represent the most basal group within the chordate lineage [2–4]. Therefore, amphioxus represents a key model system for understanding both conserved chordate developmental mechanisms and the evolutionary origins of the complex vertebrate body plan. For this reason, the genome of the amphioxus Branchiostoma floridae was sequenced and published in 2008 [5, 6], and the draft genome has become a rich source for comparative genomics and developmental studies in this organism.

To facilitate gene identification and verification of gene model predictions in the draft genome, a large collection of B. floridae expressed sequence tags...
ESTs was generated [6, 7]. ESTs are single-pass reads of several hundred base pairs generated from either the 5'- or 3'-end of randomly selected clones from cDNA libraries representing the transcript ends of the expressed portion of the genome [8]. In this review, we first provide a historical overview of EST studies that were performed prior to the completion of the *B. floridae* genome project, focusing on the biological insights that they provided. We next describe the large-scale EST analysis of *B. floridae* that was carried out along with the *B. floridae* genome-sequencing project and provide updated information on the currently available cDNA resources for this species. We also discuss the applications of the *B. floridae* cDNA resources for the use in evolutionary and developmental studies. Finally, we conclude with a discussion of the use of high-throughput next-generation DNA sequencing (NGS) technologies in the amphioxus model, focusing on the potential applications for future transcriptome analysis.

**Historical overview of amphioxus EST analyses**

Before the amphioxus genome-sequencing project, several studies had exploited EST surveys as a low-cost alternative to full genome sequencing to elucidate the tissue specific and developmental gene expression profiles of amphioxus and to investigate gene duplication events during vertebrate evolution. EST studies have been carried out by various research groups and with different amphioxus species, including the most widely used Florida amphioxus *B. floridae* and the Asian amphioxus species. It should be noted that the Asian amphioxus was originally named *Branchiostoma belcheri*, and later a subspecies status (*Branchiostomabelcheritsingtauense*) was recognized for the population distributed along the northern Chinese coast and in Japan [9, 10]. However, recent studies have clearly demonstrated that the amphioxus population of the Asian-Pacific coast is comprised of two morphologically and genetically distinct species, *B. belcheri* and *B. japonicum* [11, 12]. For clarity, the northern Asian-Pacific coast subspecies is now called *B. japonicum*, and its distribution extends to the Xiamen area near the southeast Chinese coast, where *B. japonicum* and *B. belcheri* co-exist [11]. Based on the location of the Asian amphioxus collection sites, we suspect that some of the EST studies [13, 14] did not actually collect samples from the southern species *B. belcheri* but actually...
collected samples of the northern species *B. japonicum*. To avoid confusion, henceforth will refer to *B. belcheri* and *B. japonicum* in this review to attribute the EST resources to the correct Asian amphioxus species; however, we still list the species name in parentheses as they appeared in the original literature for reference.

EST analysis was used to provide a first glimpse of the genes expressed in the amphioxus notochord [14]. The notochord is the defining characteristic of the Chordata phylum. In vertebrates, the notochord is a transient embryonic structure derived from the dorsal axial mesoderm that serves as an important signaling center, secreting various signals necessary for patterning the dorsal–ventral and left–right axes of the vertebrate body plan [15]. The notochord also functions as an axial skeletal structure in developing embryos before it is replaced by the vertebrae [15]. Unlike vertebrates, amphioxus retain the notochord throughout their lives and never develop vertebrae. Suzuki and Satoh [14] dissected the notochord tissues and isolated notochord cells from approximately 200 *B. japonicum* adults (called *Branchiostoma belcheri* in their paper) to construct a notochord cDNA library. They then sequenced both the 5′ and the 3′ ESTs from 257 randomly picked cDNA clones. After analyzing the amphioxus notochord ESTs, they found that 11% of the cDNA clones code for muscle-related proteins and 6% code for extracellular matrix proteins, suggesting that in addition to its structural function, the adult amphioxus notochord is a contractile tissue used in locomotion. Interestingly, with the exception of bFGF, they did not detect homologs of any known signaling molecules in their EST data. This finding is probably due to differences in the developmental stage assayed, as the Suzuki and Satoh library is derived from adult notochord cells, while most known signaling molecules primarily function during embryonic stages and are therefore not likely to be highly expressed in the adult stage. Additionally, the limited number of ESTs sequenced in this study may have been insufficient to detect low abundance transcripts. In fact, later large-scale EST analyses of different developmental stages of the Florida amphioxus *B. floridae* did identify classical signaling molecules and transcription factors involved in major axis patterning [16]. Moreover, the expression surveys with in situ hybridization indicated that many of these genes are expressed in either the notochord or in its precursor in amphioxus embryos [16], suggesting a conserved organizer function for amphioxus dorsal axial mesoderm.

In 2002, Mou *et al.* [13] published the sequences of 5235 ESTs from a cDNA library constructed from neurula-stage embryos of Asian amphioxus *B. japonicum* (called *Branchiostoma belcheri tsingtauense* in their paper). In addition to identifying many genes coding for common metazoan structural and enzymatic housekeeping proteins, Mao *et al.* also identified many transcription factors and regulatory proteins that are involved in developmental processes [13]. Thus, their EST analysis provided a preliminary global view of the genes expressed during this important developmental stage in amphioxus.

In another study, Panopoulou *et al.* [17] sequenced 14,189 5′ ESTs from two pre-normalized embryonic cDNA libraries of the Florida amphioxus *B. floridae*. Without whole-genome sequencing data, they attempted to compile a substantial amphioxus gene set to address the long-debated two rounds of whole-genome duplication (2R) hypothesis of early vertebrate genome evolution [18]. After assembling their amphioxus ESTs into 9,73 consensus sequences and assigning them to orthologous groups with genes from human, mouse, ascidian, *Caenorhabditis elegans*, *Drosophila*, and yeast, they estimated the extent of gene duplication at the transition from invertebrates to vertebrates. They found that on average, humans and mice have 2.6 times more gene copies per orthologous group than amphioxus. The authors therefore concluded at least one large genome duplication event occurred at the origin of vertebrates and that subsequent smaller scale duplications may have also occurred during the course of vertebrate evolution. However, this estimation of gene-copy ratio could be complicated by possible gene loss in vertebrate lineages after gene duplication events [19] or gene duplication events unique to the amphioxus lineage [20–23].

Later, the amphioxus genome-sequencing project provided more definitive information about the extent of whole-genome duplication during early vertebrate evolution. When whole-genome sequence data from the Florida amphioxus *B. floridae* became available [5, 6], a remarkably high level of conserved synteny between the amphioxus genome and human genome was recognized, leading to the identification of seventeen putative ancestral chordate linkage groups [6]. More interestingly, when mapped to human chromosomes, each putative ancestral chordate linkage group corresponded to large
segments on four different human chromosomes [6]. This quadruple conserved synteny pattern strongly suggests that two rounds of whole-genome duplication occurred in the vertebrate lineage.

**Large-scale EST analysis and the available cDNA resources of amphioxus**

In addition to the *B. floridae* genome project, a large-scale EST analysis of the amphioxus *B. floridae* has been carried out since 2004. Five non-normalized cDNA libraries were prepared from unfertilized eggs, gastrula, neurula, 36-h larvae (Figure 1B–E), and mixed male and female mature adults. Initially, 262,037 ESTs were sequenced from both the 5' and 3'-ends of approximately 140,000 cDNA clones at the National Institute of Genetics, Japan [16]. Since the first-strand cDNAs were synthesized with an oligo(dT) primer during library preparation, the 3' ESTs of this data set were subjected to cluster analysis to identify overlapping cDNA clones, and 21,229 unique transcript clusters were identified in this process. Representative cDNA clones from each cluster were chosen and re-arrayed into sixty-four 384-well plates to construct the *Branchiostoma floridae* Gene Collection Release 1’ [7]. A searchable online database for these ESTs and cDNA resources has been constructed at http://amphioxus.icob.sinica.edu.tw/. Users can perform BLAST searches against the EST database to identify their genes of interest and can also request cDNA clones through the website. At the same time, an additional set of 97,536 ESTs was sequenced from the same gastrula, neurula and 36-h larvae libraries by the Joint Genome Institute (USA) during the *B. floridae* genome-sequencing project [6]. After the removal of low-quality and contaminant EST sequences, 56,964 ESTs from this data set were also deposited in Genbank. Together with the aforementioned ESTs from *B. floridae*, these ESTs were used to predict protein-coding loci in the amphioxus draft genome. Because the previous clustering analysis did not include this second set of EST data [7], we reanalyzed the data and provided some updates on the status of the current *B. floridae* EST data.

To date, we have obtained 319,001 ESTs from these five *B. floridae* cDNA libraries (Table 1), and have found that 41.8% of the predicted amphioxus gene models on v.2.0 assembly are covered by the ESTs. This finding suggests that more than half of the predicted genes are either not expressed in the developmental stages analyzed in this EST analysis, or that they are expressed at such a low level that their transcripts could not be detected. We used 158,308 3' ESTs for the clustering analysis and detected 27,415 unique cDNA clusters (17,452 singleton clusters and 9,963 clusters of multiple ESTs). Of note, the number of cDNA clusters is larger than the number of protein-coding loci (21,900) predicted from the draft genome sequence [6]. This result is likely due in part to the existence of alternatively spliced isoforms of specific genes in amphioxus [24]. Unlike genome-sequencing coverage, it is not straightforward to extrapolate the coverage of the EST data to the entire transcriptome because the exact number of different transcript isoforms produced by each gene is not known in amphioxus. To estimate the coverage of this EST data, we plotted the relationship between the number of 3' ESTs and the unique cDNA clusters detected by the clustering analysis and found that the curve did not plateau (Figure 2A), suggesting that further EST sequencing of these cDNA libraries could still uncover many new cDNA clusters. We applied mathematical models to this plot to estimate the maximum diversity of amphioxus transcript species through extrapolation [25]. The model predicts that the amphioxus transcriptome contains a maximum of 79,100 unique transcripts, and many transcripts are predicted to be rare species that would only be identify by sequencing a considerable number of ESTs. A comprehensive analysis of mouse full-length cDNAs from the FANTOM project has identified more than 181,000 independent transcripts [26], and the data suggested that the total number of transcripts is at least one order of magnitude larger than the estimated number of protein-coding genes (22,000) in the mouse genome. A recent study using a transcript-tagging technique and the NGS platform identified more than 194,000 unique tag sequences from

---

**Table 1: Numbers of ESTs from five *B. floridae* cDNA libraries**

<table>
<thead>
<tr>
<th>Library</th>
<th>5' EST</th>
<th>3' EST</th>
<th>Total</th>
<th>3' Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egg</td>
<td>19,396</td>
<td>19,126</td>
<td>38,522</td>
<td>5,484</td>
</tr>
<tr>
<td>Gastrula</td>
<td>21,708</td>
<td>19,872</td>
<td>41,580</td>
<td>6,437</td>
</tr>
<tr>
<td>Neurula</td>
<td>69,058</td>
<td>68,582</td>
<td>137,640</td>
<td>14,193</td>
</tr>
<tr>
<td>36 h larvae</td>
<td>31,730</td>
<td>31,766</td>
<td>63,496</td>
<td>10,552</td>
</tr>
<tr>
<td>Adults</td>
<td>18,801</td>
<td>18,962</td>
<td>37,763</td>
<td>4,868</td>
</tr>
<tr>
<td>Total</td>
<td>160,693</td>
<td>158,308</td>
<td>319,001</td>
<td>27,415*</td>
</tr>
</tbody>
</table>

*The number of total clusters was calculated by combining all 3' EST data from the five libraries for the clustering analysis.
Figure 2: Clustering analysis of current B. floride ESTs from five cDNA libraries. (A) The plot of detected unique cDNA species with respect to the number of input 3' EST sequences from all five cDNA libraries. The curve has not reached its plateau due to the continuous discovery of many singleton clusters. (B) The corresponding plot with high confidence clusters, which are defined as clusters having at least two 3' EST reads. The number of unique high confidence cDNA clusters detected begins to plateau (99%) when the number of input 3' ESTs reaches approximately 116,000 sequences. (C) Relationships between the number of 3' ESTs and the unique cDNA clusters detected by clustering analysis at the five developmental stages indicate that adult amphioxus express a smaller number of transcripts. As embryonic development proceeds, the embryos and larvae begin to express additional transcript species.
the cheetah skin transcriptome [27]. Therefore, our estimation of that the amphioxus transcriptome contains 79 100 unique transcripts, which is ~40% of the mammalian transcriptome, is within a reasonable limits. It should be noted that non-coding RNAs (either polyadenylated or non-polyadenylated) are now recognized to contribute a significant portion of reads to transcriptome data [28, 29]. This finding may also explain in part the large difference between the number of predicted protein-coding genes in amphioxus and the number of transcripts. Further comprehensive analysis of amphioxus full-length transcripts and detailed annotation of different classes of non-coding RNAs in this organism will help determine the relative contribution of non-coding RNAs to the amphioxus transcriptome.

We also noted that when we focus on unique cDNA clusters composed of 2 or more reads (‘high confidence clusters’), the plot of sequenced 3’ ESTs versus the identified unique cDNA clusters shows a saturation plateau (99%) at approximately 116 000 3’ ESTs (Figure 2B); however, the number of transcript species (9862) determined by the high confidence cluster set is far less than the expected gene number (21 900 protein-coding loci). This result suggests that although the current coverage of amphioxus EST data is insufficient to identify transcripts that are expressed at relatively low levels, the current EST data can account for most of the highly expressed genes. The model calculates that an additional 880 000 ESTs would be needed to reach 90% transcriptome coverage and that approximately 2 million ESTs would be needed to detect >99% of the predicted transcript species (79 100). Therefore, traditional Sanger sequencing of these cDNA libraries does not represent a cost-effective way to increase our coverage of the amphioxus transcriptome. We will further discuss this issue in the last section of this review.

Comparisons of amphioxus and Ciona EST data
Since the five B. floridæ libraries prepared in this EST project were not amplified or normalized [7], the EST counts for each transcript cluster should be in proportion to their abundance in each particular library. Because quantitative data from more sophisticated methods, such as microarray analysis or NGS-based RNA-sequencing, are currently lacking in the amphioxus system, we therefore examined the relationship between the number of 3’ESTs and the number of detected clusters at each of the five developmental stages (Figure 2C) and used this as a proxy method to obtain a preliminary view of the global expression trends. Comparing the four embryonic stages shows that the ratio of detected clusters versus 3’ESTs increases during embryonic development, with the lowest ratio in the unfertilized eggs, comparable levels in the gastrula and the neurula, and the highest ratio in larvae. This pattern suggests that amphioxus express a relatively small population of transcripts and, as embryonic development proceeds, embryos and larvae begin to express additional transcript species. The ratio of detected clusters versus 3’ESTs is lowest in the adult stage, suggesting that amphioxus expresses the smallest number of transcripts in adulthood; presumably, many developmental genes are no longer expressed in adults or are expressed at such low levels that they were not detected in this EST analysis. Interestingly, this global expression pattern stands in stark contrast to the expression patterns found in similar large-scale EST analyses of the ascidian Ciona intestinalis [30, 31]. It has been demonstrated that Ciona eggs express the most diverse population of transcripts, with embryos and larvae gradually expressing fewer transcripts as they progress through development [31]. As the larvae undergo metamorphosis to develop an adult body plan, Ciona appears to express a slightly higher number of transcript species, suggesting that there is a major transition during metamorphosis. Recent microarray data from Ciona have confirmed this pattern, indicating that the largest expressed developmental gene set is the maternal gene cluster, which comprises 38.8% of the tested genes [32]. The sharp contrast between the developmental expression trends of Ciona and amphioxus may reflect fundamental differences in the early developmental mechanisms of ascidians and amphioxus. It is widely recognized that ascidian embryos rely heavily on early localized maternal factors for cell fate determination [33]. The most notable examples are the Postplasmic/PEM class of mRNAs [34, 35], such as the Macho-1 mRNA that initially localizes to the vegetal cytoplasm after fertilization and is subsequently inherited by specific blastomeres where it drives muscle differentiation [36], and the mRNA of Vasa homologue (G-VH) which is important in the development of primordial germ cells [37]. In contrast, amphioxus development is generally...
considered to be highly regulative [38]. Indeed, the global gene expression trends that have been identified in the amphioxus EST data seem to be in accordance with this concept.

However, it should be noted that mosaic and regulative development are simply two extremes of the same scale, and all embryos probably make use of both mechanisms [39]. For example, we now know that ascidian development does not completely depend on localized determinants. The regulative interactions mediated by secreted signaling molecules, such as Fgf, BMP and Nodal, play important functions in cell fate determination for mesoderm, endoderm and neural ectoderm tissues in ascidian embryos [40–42]. On the other hand, many vertebrates, such as *Xenopus* and zebrafish, display highly regulated development yet also possess high levels of maternal transcripts in their eggs [43–45]. Similar to vertebrates, the presence of maternal transcripts in amphioxus eggs should not be disregarded as their identities and functions have not yet been systematically studied. Using the *B. fluoridae* EST and cDNA resources, our group has recently demonstrated that the maternal transcripts of Vasa and Nanos localize asymmetrically near the vegetal pole of the amphioxus eggs, and our data suggest that they are the putative germ cell determinants of amphioxus [46]. More interestingly, we showed that the condensed vegetal pole cytoplasm is inherited asymmetrically by only one blastomere from the two-cell embryo to the early gastrula, suggesting that even the first two blastomeres of the amphioxus embryo are developmentally distinct [46]. This study highlights the potential mosaic property of amphioxus embryos and indicates a need for additional study of the molecular properties of maternal transcripts in amphioxus. Further detailed analysis of the amphioxus egg transcriptome and additional comparative studies between amphioxus and other chordate groups (namely, the ascidians and vertebrates) should provide more information on the evolution of early patterning mechanisms in chordates.

The use of amphioxus EST resources

Large-scale EST analysis and the cDNA resources of the amphioxus *B. fluoridae* have become valuable tools for research requiring comprehensive genomic information. Several phylogenomic studies have relied mainly on EST-based data from representative metazoan taxa (including amphioxus) to construct large data matrices for phylogenetic analyses [2, 3, 47]. This approach has been very effective and has provided many new insights that could not be easily discovered through single-gene analysis, such as the sister-group relationship between tunicates and vertebrates. The amphioxus EST data also allow researchers to systematically obtain a set of transcripts involved in a specific developmental process or in a specific metabolic pathway without the need for individual screening of the desired cDNAs. For example, Martinez et al. [48] used the amphioxus EST data along with other EST data and genome assembly data from representative organisms to trace the evolutionary history of neural crest-related genes. Using this method, they found that a considerable number of genes encoding signaling molecules emerged during vertebrate evolution, which may account for the cell type diversification of neural crest derivatives in early vertebrates.

In addition to data-mining research, the currently available *B. fluoridae* cDNA resource [7] is particularly useful for cDNA cloning-based studies. After initial identification of ESTs representing genes of interest, cDNA clones can be easily obtained through request via the *Branchiostoma fluoridae* cDNA Database (http://amphioxus.icob.sinica.edu.tw/), and their expression or function can then be analyzed. Using such approaches, researchers have identified amphioxus homologues of genes involved in the development of many important vertebrate features, and the subsequent comparative analyses have proved highly informative in inferring the evolutionary origins of these structures. For example, using the *B. fluoridae* EST database, researchers have identified most of the important signaling molecules and transcription factors involved in the development of the vertebrate organizer (Spemann’s organizer) in amphioxus [16]. They have also demonstrated that the expression patterns of these genes and the developmental mechanism for dorsal/ventral (D/V) patterning are conserved between amphioxus and vertebrates, suggesting that the Spemann’s organizer has deeper roots in the base of the chordate lineage. Similarly, to understand the evolutionary origin of the vertebrate neural crest, researchers have used genomic and EST databases to identify amphioxus genes whose homologues have been known to function in a putative neural crest gene regulatory network [49]. Using comparative expression analysis, they demonstrated that while amphioxus possess conserved upstream
signaling molecules and transcription factors for specifying the neural plate during embryogenesis, amphioxus do not express one group of important transcription factors (such as FoxD3, SoxE, Id and Twist) at the neural plate border [49]. They therefore proposed that this group of genes may have been co-opted by the neural plate border during early vertebrate evolution, contributing to the evolution of neural crest cells. In another study, researchers used the EST database to search for amphioxus homologues of developmental genes involved in cranial neural crest and cartilage development [50], and they found that those genes are predominantly expressed in the mesodermal tissues of amphioxus, suggesting that the deployment of the chondrogenic gene network of the mesoderm predates the evolution of neural crest-derived cartilages. Other studies have also used genomic and EST database to investigate the evolution of the somites [51] and of the lateral plate mesoderm [52]. These results demonstrated that amphioxus and vertebrates use a mostly conserved genetic toolkit in somite segmentation [51]; however, judging from marker gene expression, the ventral mesoderm of amphioxus seems to lack apparent regionalization compared to that of vertebrates [52]. EST data have also been used to confirm the existence of a large family of endogenous green fluorescent proteins (GFP) in *B. flordae* and other amphioxus species [53–55], making amphioxus the only deuterostome animal in which endogenous GFPs have been identified.

Similar approaches have also been used in *B. japonicum* to identify immune-relevant genes and liver-specific genes from amphioxus EST collections [56, 57]. Using EST data from lipopolysaccharide-challenged amphioxus, Liu *et al.* [56] identified 63 genes related to the immune response. Interestingly, some genes involved in histocompatibility and lymphocyte immune signaling were also identified in their study. Their results therefore support the previous notion that amphioxus possess some of the genetic components required for an adaptive immune system [58, 59]. This finding in turn suggests that amphioxus may possess a primitive adaptive immune system and that vertebrates may have co-opted these genetic components for the assemblage of a more sophisticated adaptive immune system. In another EST study using *B. japonicum*, Wang and Zhang [57] identified 69 homologues of vertebrate liver-specific genes and used quantitative real-time PCR to confirm that 58 of these genes are highly expressed in the hepatic cecum of amphioxus, supporting the hypothesis that the amphioxus hepatic cecum is homologous to the vertebrate liver. They also found that the majority of these hepatic cecum-specific genes responded to lipopolysaccharide challenges in a similar fashion to their zebrafish homologues, further suggesting that amphioxus also has liver-mediated innate immune response. Thus, the currently available EST data represent a rich source for both sequence-based analyses and molecular developmental studies in various amphioxus species.

**FUTURE PERSPECTIVES**

In recent years, the development of NGS technologies has revolutionized the scale of transcriptomic studies [60, 61]. The major advantage of NGS is the ability to generate large amounts of sequence data, ranging from hundreds of thousands to billions of short sequence reads (50–350 bp depending on the sequencing platform used) at a relatively low cost. Several commercially available NGS platforms, including Roche/454, Illumina/Solexa and ABI/SOLiD systems, have been successfully used for identifying small regulatory RNAs (microRNAs), characterizing alternative splicing isoforms, and mapping and quantifying transcriptomes in both traditional model organisms and in human cells [62, 63]. More recently, NGS applications have been expanded to include analysis of non-model organisms for genome/transcriptome studies [64–67]. Because of their high-throughput nature, the sequencing depth provided by the NGS platforms is usually sufficient for global transcriptome characterization from a single sample, or for surveying a great number of genes simultaneously from multiple samples [63]. Moreover, compared with hybridization-based methods such as cDNA microarrays or genomic tiling arrays, sequence-based NGS transcriptome profiling has less background noise and a much greater dynamic range for quantifying gene expression levels. The development of algorithms for *de novo* transcriptome assembly from short-read sequencing data and for the tagging approach for individual transcripts also make global gene expression analyses possible in organisms for which there are not yet high-quality reference genomes [27, 68, 69].
To date, the 454 and Solexa sequencing platforms have been used to characterize microRNA comple-
ments in the amphioxus system [70, 71] and have led to the discoveries of more novel microRNAs than were either computationally predicted based on the draft genome or identified via Sanger sequencing of small RNA libraries [72, 73]. We anticipate that further applications of NGS technologies will greatly enhance our current knowledge of the amphioxus transcriptome. For example, RNA sequencing (RNA-Seq) of libraries prepared from normal wild-type embryos at different developmental stages and from selected adult tissues would help us to catalogue the full complement of transcripts in amphioxus. The throughput provided by NGS platforms, such as Solexa, should be sufficient to fill in the gaps in our current amphioxus transcriptome coverage. Presumably, high-coverage NGS data could also help improve the current genome annotation by providing information on transcription start sites and exon/intron boundaries. Furthermore, it should also provide information on alternative transcript splicing patterns and help to identify non-coding RNAs in the amphioxus transcriptome.

Quantitative measures of transcript species from NGS data can be used for profiling global expression changes during embryonic development. Recently, various pharmacological reagents and recombinant proteins have been used to study the developmental functions of major signaling pathways in amphioxus [16, 49, 74–79]. Future NGS-based transcriptomic comparisons between wild-type embryos and reagent-treated embryos could not only reveal the global effects of manipulating these signaling pathways during embryogenesis but also help to identify novel target genes downstream of these signaling pathways. In the post-genomic era of amphioxus research, this approach will be highly effective and will provide global information for understanding the basic developmental mechanisms of chordates. With continued improvements in the current sequencing platforms and associated bioinformatic methods and simultaneous decreases in the cost of NGS-based analyses, this type of study is becoming economically feasible for many researchers. We anticipate that NGS-based transcriptome analyses will greatly increase the width and depth of evo-devo research and will further facilitate the use of non-model organisms for future comparative genomic studies.

**Key points**

- Amphioxus represents the most basal chordate group and is an emerging model organism for evolutionary developmental biology.
- The amphioxus draft genome and EST/cDNA resources have greatly increased our ability to use amphioxus in comparative genomics and developmental studies.
- The current B. floride EST data and its comparison with Ciona EST data provide interesting insights into the dynamic changes in gene expression occurring during embryonic development.
- The advances of high-throughput NGS technologies have dramatically expanded the scope of transcriptome studies; further applications of NGS in the amphioxus system will increase the width and depth of such transcriptomic analysis on a genome-wide scale.

**FUNDING**


**References**


