Enhancer activities of amphioxus Brachyury genes in embryos of the ascidian, Ciona intestinalis

Enhancer activities of amphioxus *Brachyury* genes in embryos of the ascidian, *Ciona intestinalis*

Hitoshi Tominaga[1,2,*], Noriyuki Satoh[3], Naoto Ueno[1,2], Hiroki Takahashi[1,2,*]

[1] Division of Morphogenesis, Department of Developmental Biology, National Institute for Basic Biology, National Institutes of Natural Sciences, Okazaki, Aichi 444-8585, Japan

[2] Department of Basic Biology, School of Life Science, the Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan

[3] Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan

*Correspondence

Hitoshi Tominaga and Hiroki Takahashi, Division of Morphogenesis, Department of Developmental Biology, National Institute for Basic Biology, National Institutes of Natural Sciences, Okazaki, Aichi 444-8585, Japan.

Email: tommy@nibb.ac.jp or taka@nibb.ac.jp

Grant Numbers: Grants-in-Aid from the Japan Society for the Japan Science (JSPS) KAKENHI 18K06370, 16H04824

A short title: Amphioxus *Brachyury* enhancers in *Ciona* system

KEYWORDS

Amphioxus *Brachyury*, enhancer activity, *Ciona* system, notochord, somites

Abstract

The notochord and somites are distinctive chordate structures. The T-box transcription factor gene, *Brachyury*, is expressed in notochord and plays a pivotal role in its formation. In the cephalochordate, *Branchiostoma floridae*, *Brachyury* is duplicated into *BfBra1* and *BfBra2*, which are expressed in the somite-formation region as well. In a series of experiments to elucidate the regulatory machinery of chordate *Brachyury* expression, we carried out a *lacZ* reporter assay of *BfBra* in embryos of the urochordate, *Ciona intestinalis*. Vista analyses suggest the presence of conserved non-coding sequences, not only in the 5'-upstream, but also in the 3'-downstream and in introns of *BfBra*. We found that: (1) 5'-upstream sequences of both *BfBra1* and *BfBra2* promote *lacZ* expression in muscle cells, (2) 3'-downstream sequences have enhancer activity that promotes *lacZ* expression in notochord cells, and (3) introns of *BfBra2* and *BfBra1* exhibit *lacZ* expression preferentially in muscle and notochord cells. These results suggest shared cephalochordate *Brachyury* enhancer machinery that also works in urochordates. We discuss the results in relation to evolutionary modification of

Brachyury expression in formation of chordate-specific organs characteristic of each lineage.

(181 words)

1. INTRODUCTION

Chordates comprise three taxa, cephalochordates, urochordates (tunicates), and vertebrates. Chordates are members of a super-phyletic group of deuterostomes, together with echinoderms and hemichordates. Recent molecular phylogeny showed two distinct clades in deuterostomes, ambulacrarians (echinoderms and hemichordates) and chordates (cephalochordates, urochordates, and vertebrates) (Halanych, 1995; Cameron et al., 2000; Delsuc et al., 2006; Bourlat et al., 2008). Among chordates, cephalochordates likely diverged first, more than 520 million year ago, leaving urochordates and vertebrates in a sister relationship (Delsuc et al., 2006; Bourlat et al., 2008; Putnam et al., 2008). This scheme of chordate origins suggests that the chordate ancestor was a free-living, cephalochordate-like animal that gave rise to urochordates and vertebrates. Comparative embryology and genetics suggest that urochordates became advanced filter-feeders (Satoh 2009) while vertebrates evolved more complex traits, presumably owing to two rounds of whole-genome duplication.

Evolutionary developmental biological mechanisms involved in the origin and evolution of chordates have been extensively studied since the commencement of evo-devo research field in the 1990s (Holland et al., 2015; Lowe et al., 2015; Satoh, 2016). Traditionally, it was thought that chordates share body plans that possess a notochord, a dorsal neural tube, pharyngeal gill slits, an endostyle, lateral somites, and a post-anal tail (Gee, 1999; Satoh, 2016). However, recent studies have shown that pharyngeal gill slits and a post-anal tail are shared by hemichordate acorn worms (Gerhart, 2006; Simakov et al., 2013). On the other hand, a mid-dorsal notochord and lateral somites are specific to chordates, although application of the term "somite" to bilateral muscle of the urochordate larval tail has been a matter of some debate. Therefore, the notochord and somites are the most significant organs in the origin and evolution of chordates. We have proposed that fish- or tadpole-like tail-beating larvae were deeply involved in innovation of these chordate-specific organs (Satoh et al., 2012; Satoh, 2016).

However, the developmental mode of notochord and somites is not identical in the three chordate lineages. For example, after gastrulation, cephalochordate embryos pinch out three folds at the mid-dorsal, left, and right sides of the archenteron (Conklin,

1932; Hirakow & Kajita, 1994). The mid-dorsal fold gives rise to notochord, and the left and right folds form somites in neurulae. In addition, cephalochordate notochord is composed of myofibrils, like the somites on both sides (Ruppert, 1997; Suzuki & Satoh, 2000, Inoue & Satoh, 2018). On the other hand, urochordate and vertebrate embryos form the notochord by convergent extension of its presumptive cells, which occur on both sides of early-stage embryos (Satoh, 2014). These notochords have no myofibril-like components and become occupied by vacuoles, which produce the characteristic stiffness of the organ.

Various transcription factors and signaling molecules involved in the formation of chordate body plans have been identified and characterized (Davidson, 2006; Holland et al., 2015; Lowe et al., 2015; Satoh, 2016). Among them, a T-box family transcription factor gene, *Brachyury*, has attracted special interest of evo-devo researchers (Holland, 2000; Technau, 2001; Satoh et al., 2012) because it has proven essential for notochord formation in urochordate (Yasuo & Satoh, 1993; Corbo et al., 1997) and vertebrate embryos (Wilkinson et al., 1990), although the function of *Brachyury* in cephalochordate embryos remains to be explored. In relation to the

aforementioned and other modes of notochord formation, the expression pattern of *Brachyury* differs among the three taxa (Figure 1). Specifically, *Brachyury* is first expressed in vertebrate embryos in the marginal zone of the blastopore, and then in notochord, with residual expression in the tailbud region (Wilkinson et al., 1990; Schulte-Merker et al., 1994). In urochordate embryos, *Brachyury* is expressed exclusively in the notochord (Yasuo & Satoh, 1993; Corbo et al., 1997), while in cephalochordate embryos, *Brachyury* is expressed in the marginal zone of blastopore, and then in notochord and somites (Holland et al., 1995; Terazawa & Satoh, 1995).

In addition, *Brachyury* is not exclusive to chordates, but is present in most metazoans, and the gene is commonly expressed in the archenteron invagination region (Holland, 2000; Technau, 2001; Satoh et al., 2012). A possible evo-devo explanation of *Brachyury* expression and function in metazoans is that non-chordate invertebrates employ *Brachyury* for its primary expression and function (PEF), which is associated with gastrulation and/or mesoderm formation (Figure 1) (Satoh et al., 2012). In contrast, chordates have acquired additional secondary expression and functions (SEF) of the gene, associated with notochord formation (Figure 1). Therefore, determining how the

chordate stem acquired SEF is essential to understanding genetic and molecular mechanisms involved in notochord formation. Its answer is profoundly associated evo-devo mechanisms involved in the origin and evolution of chordates (Holland et al., 2015; Lowe et al., 2015; Satoh et al., 2013; Satoh, 2016).

Beside vertebrate paralogs or ohnologs derived from ancient whole genome duplication, *Brachyury* is present as a single copy in most metazoans, with some exceptions, including cephalochordates (Bielen et al., 2007; Inoue et al., 2017). In the Florida lancelet, *Branchiostoma floridae,* two *Brachyury* genes (*BfBra1* and *BfBra2*) are present in the genome (Holland et al., 1995; Inoue et al., 2017). They encode very similar transcription factor proteins (Figure 2) (Holland et al., 1995). *BfBra1* is approximately 2,455 bp in length, encompassing a potential open reading frame of 1,350 nucleotides (nt) and a deduced protein of 449 amino acids. *BfBra2* is 1,862 bp, encompassing an ORF of 1,323 nt, a deduced protein of 440 amino acids. However, they differ markedly in their 3' untranslated regions (UTRs) (Holland et al., 1995). Lengths of the 3' UTRs are 834 nt in *BfBra1* and 381 nt in *BfBra2* (excluding stop codons and terminal A residues). Due to such high sequence similarity, no experiments

have succeeded in showing different and/or independent spatial and temporal expression of each gene. In addition, functional analysis of *BfBra* remains to be done.

Given this background, and in order to understand the evolutionary modification of *Brachyury* expression and function, we attempted to determine the regulatory mechanism of *Brachyury* expression in basal chordates, cephalochordates, and urochordates. To this end, at least four types of enhancer analysis were necessary: *lacZ* reporter assays of amphioxus *Brachyury* in amphioxus embryos, reporter assays of ascidian *Brachyury* in ascidian embryos, reporter assays of amphioxus *Brachyury* in ascidian embryos, and reporter assays of ascidian *Brachyury* in amphioxus embryos. Of these, a native assay of ascidian *Brachyury* enhancer has been performed in two ascidian species, *Ciona intestinalis* (Corbo et al., 1997; Takahashi et al., 1999) and *Halocynthia roretzi* (Matsumoto et al., 2007). Although the results of *Ci-Bra* and *Hr-Bra* are not strictly comparable due to a complicated combination of regulatory modules, these studies suggested significant roles of Zic-, Ets-, and Snail-binding sites for *Brachyury* expression control. On the other hand, the other three types of assay have

not been done yet. We report here *lacZ* reporter assays of amphioxus *Brachyury* in ascidian embryos.

2. Results

2.1 *BfBra1* and *BfBra2,* and conserved non-coding sequences

In *B. floridae,* the two *Brachyury* genes (*BfBra1* and *BfBra2*) align in different strands of genomic DNA, sharing approximately 8-kb of 3'-downstream sequences (Figure 2a). Both *BfBra1* and *BfBra2* consist of seven exons interrupted by six introns (Figure 2; Figure 3). As mentioned above, the two genes encode very similar transcription factor proteins (Figure 2b, c). However, they differ markedly in their 3' untranslated regions (UTRs) since the lengths of the 3' UTRs are 834 nt in *BfBra1* and 381 nt in *BfBra2* (excluding stop codons and terminal A residues) (Figure 2a).

A recent study demonstrates approximately 113,000 noncoding elements that have been conserved between two distantly related cephalochordate species,

Asymmetron lucayanum and *Branchiostoma floridae* (Yue et al., 2016). We performed two types of Vista analyses. First, we compared two strands of *B. floridae* genomic sequences that include *BfBra1* and *BfBra2* (Figure 3a), and second, we compared the *B. floridae* sequences with the *B. belcheri* sequences (Figure 3b). These analyses demonstrated the presence of conserved non-coding sequences (CNSs), not only in the 5'-upstream, but also in the 3'-downstream and intragenic regions of the genes as well (Figure 3). We therefore examined enhancer activity of all three genomic regions. Because the 3'-downstream and introns have no promoter sequences, we took advantage of the minimal promoter of *Ci-Bra* to construct *lacZ* reporters. We electroporated *BfBra-lacZ* constructs into fertilized *Ciona* eggs and detected reporter expression at 5 (gastrula), 7-8 (neurula) and 11-12 hours (tailbud embryos) of embryogenesis.

2.2 5' upstream sequences of *BfBra* have enhancer activity in muscle

5’ upstream sequences of both *BfBra1* and *BfBra2* contain two major CNSs, proximal (around -0.8 kbp; Figure 4a, blue) and distal (around -2 kbp; Figure 4a, orange). We first examined enhancer activity of *BfBla1-3kbp*, which includes both CNSs, and *BfBra2-2kbp*, which contains the proximal CNS only (Figure 4a). Approximately 45% (at 8h) and 36% (at 12h) of *BfBra2-2kbp*-electroporated embryos showed *lacZ* reporter in muscle cells, respectively (Figure 4b; Table 1). Similarly, approximately 46% (at 8 h) and 28% (at 12 h) of *BfBla1-3kbp*-electroporated embryos showed *lacZ* reporter in muscle cells (Figure 4c, Table 1). In general, signal intensity was higher and broader in *BfBra2* than *BfBra1,* and *lacZ* expression of the former expanded to posterior muscle cells (compare Figure 4b to c; Figure 4d). This result indicates enhancer activity of the 5’-upstream sequences of *BfBra* that promote *lacZ* expression specifically in muscle cells of *Ciona* embryos.

Next, we examined -5.5, -3, -2, and -1 kbp upstream constructs of *BfBra1* and -5.7, -2, and -1 kbp constructs of *BfBra2,* respectively, to determine the role of distal and proximal CNSs (Figure 4d). -5.5 and -3-kb upstream constructs of *BfBra1* and -5.7-kb of *BfBra2* contained both proximal and distal CNSs, while the others contained the

proximal CNS only. The three reporter constructs with both CNSs exhibited *lacZ* expression in muscle cells (Figure 4d). On the other hand, those with proximal CNS only showed ectopic expression in mesenchyme cells of several tailbud embryos, although the results of ectopic expression differed between *BfBra2* and *BfBra1* (Figure 4d). This suggests the presence of a module(s) in upstream sequences between -2kb and -5kb that represses mesenchyme expression.

Beaster-Jones et al. (2007) reported that when *lacZ*-reporter constructs of 5' upstream sequences of amphioxus *engrailed* ligated to the 0.7-kb *Ci-fkh* promoter were injected into *Ciona* eggs, reporter expression was observed ectopically in mesenchyme, notochord, and muscle cells. The constructs used in the present experiment did not contain *Ci-fkh* promoter, namely exclusively *BfBra* 5'-upstream sequences. In addition, no ectopic reporter expression was observed when constructs with the proximal and distal CNSs were electroporated into *Ciona* eggs, although a few electroporated with shorter (-1.0 and -0.5 kb) constructs showed ectopic expression in mesenchyme. This indicates that the vector itself did not significantly affect the major results.

Alignments of 600-bp sequences of the 5'-upstream region of *BfBra1* and *BfBra2* (Figure 5a) and those with *B. belcheri BbBra1* and *BbBra2* (Figure 5b) show the presence of possible sequence-specific transcription factor-binding motifs. These include two Snail motifs (shown by light green), three MyoD motifs (red boxes), four T-box motifs (yellow), and two E-box motifs (blue). Future studies should determine the role of these motifs in enhancer activity of the 5'-upstream region of *BfBra1* and *BfBra2*.

2.3 3'-downstream sequences of *BfBra* have enhancer activity in notochord

To date, no experiments have reported enhancer activity of the 3'-downstream sequences of *Brachyury*, including those of ascidians and vertebrates. *BfBra1* and *BfBra2* contain 834-bp and 381-bp 3' sequences, respectively (Figure 6a). In addition, Vista comparison of *Bra* from *B. floridae* and *B. belcheri* demonstrated the presence of several CNSs in the 3'-downstream sequences (Figure 3b). We made reporter constructs, *BfBra1+3kb* and *BfBra2+3kb*, using the *Ci-Bra* minimal promoter to

examine their enhancer activity. Interestingly, both constructs exhibited *lacZ* expression in notochord cells in *Ciona* embryos (Figure 6). In particular, *BfBra2+3kb* resulted in notochord-specific reporter expression in approximately 90% and 80% of electroporated embryos examined at 7 and 11 h, respectively (Figure 6b; Table 1). A similar result was obtained in *BfBra1+3kbp* constructs. Approximately 60% (at 7 h) and 50% (at 11 h) of the embryos showed *lacZ* expression in notochord cells (Figure 6c; Table 1). As in the 5'-upstream constructs, the intensity of the reporter was stronger in *BfBra2+3kb* than *BfBra1+3kb* (compare Figure 6b to c). These results clearly show that the 3'-downstream sequences of lancelet *Brachyury* have enhancer activity to promote reporter expression in notochord cells of urochordate embryos.

2.4 Intragenic sequences of *BfBra* have enhancer activity in muscle and/or notochord

As described above, both *BfBra1* and *BfBra2* are interrupted by six introns. From the 5' to 3' direction, the first introns (I1) of *BfBra1* and *BfBra2* are 339 and 353 bp long,

respectively (shown as 1 and 1' in Figure 7a, c). I2: 1,361 and 450 bp (shown as 2 and 2'). I3: 391 and 191 bp (shown as 3 and 3'). I4: 174 and 274 bp (shown as 4 and 4'). I5: 236 and 505 bp (shown as 5 and 5'), and I6: 1,174 and 348 bp (shown as 6 and 6' in Figure 7a, c), respectively. Vista comparison of *Bra1* and *Bra2* intron sequences from *B. floridae* and *B. belcheri* indicates the presence of CNSs (Figure 3b). We made reporter constructs that include each of the introns (Figure 7a, c). In contrast to the 5'-upstream and 3'-downstream sequences of *BfBra1* and *BfBra2*, which yielded similar spatial expression *lacZ* profiles, the *BfBra1* and *BfBra2* intragenic regions resulted in different profiles of spatial reporter-expression. *BfBra1* exhibited *lacZ* expression primarily in notochord cells (Figure 7b) and *BfBra2* primarily in muscle cells (Figure 7d). I5 and I6 of *BfBra2* showed enhancer activity (Figure 7d), but those of *BfBra1* lack activity almost entirely (Figure 7b).

Approximately half of *Ciona* embryos electroporated with *BfBra1* I1, I2, I3 and I4 constructs exhibited *lacZ* expression in 7-h-embryos, although none of the 143 and 86 electroporated embryos with I5 or I6 showed reporter expression (Figure 7b; Table 1). *lacZ* expression was seen in notochord cells when examined at 7 h of

development, although when examined at 11 h, a reporter of I1 and I4 was expressed in muscle cells of 5~10% embryos as well (Figure 7b; Table 1). This indicates that the first four introns of *BfBra1* have enhancer activity to promote notochord-specific expression of *lacZ* in the *Ciona* assay system.

Interestingly, spatial expression of *lacZ* in *BfBra2* intron constructs differed from that of *BfBra1* introns mentioned above. That is, *BfBra2* I1-I6 constructs exhibited *lacZ* expression mainly in muscle cells (Figure 7d; Table 1), although in some embryos examined at 11 h, reporter expression was seen in notochord cells as well (Figure 7d; Table 1).

These results clearly indicate the presence of enhancer activity in intron sequences of both *BfBra1* and *BfBra2*. As observed in the 5' and 3' sequences, *BfBra2* exhibited stronger enhancer activity compared to *BfBra1* in the intron sequences (compare Figure 7d with 7b). In addition, all introns of *BfBra2* showed this activity, primarily in muscle, but not in notochord cells. Also, the number of positive embryos is larger in *BfBra2* constructs, in which nearly 100% of manipulated embryos with *BfBra2* I-3 and I-4 showed reporter expression (Figure 7d).

In experiments to examine intron sequences and 3'-downstream sequences as well, we used *Ci-Bra* minimal promoter, since these amphioxus sequences have no minimal promoters for reporter expression. It has been reported that the *Ci-Bra* minimal promoter itself sometimes induces ectopic reporter expression in mesenchyme cells (Corbo et al., 1997). Although several embryos electroporated with the intron sequences showed ectopic expression, it is unlikely that the results obtained were caused by the empty vector construct itself.

2.5 Temporal control of lacZ reporter expression

Since both *BfBra1* and *BfBra2* are expressed first in the blastopore region of early gastrula embryos, we examined temporal control of the enhancer activity with the hope of better understanding the primary expression scheme of *Brachyury* (Figure 1a). The initial expression of reporter genes was examined in 5-h embryos, which correspond to mid gastrula stage of *Ciona*. Three reporter constructs, I-3 and I-5 of *BfBra2* and *BfBra1-3kbp,* were examined for this purpose (Figure 8). We found that nearly 100% of

BfBra2 I-3-electroporated embryos showed *lacZ* as early as 5 h in primordial cells of both notochord and muscle (Figure 8, left). In addition, nearly 75% of *BfBra2* I-5-electroporated embryos also showed early *lacZ* expression (Figure 8, right). Such early reporter expression in both notochord and muscle implies that the enhancers retain temporal control that appears to be associated with PEF expression of the genes.

3. Discussion

The present study showed that genomic regions around cephalochordate *BfBra1* and *BfBra2* have enhancer activity when examined in the urochordate *Ciona*-host system. The 5'-upstream sequences promote *LacZ* reporter expression in muscle cells, and the 3'-downstream sequences in notochord cells. *BfBra1* introns promote *LacZ* reporter expression primarily in notochord cells while *BfBra2* introns do so in muscle and notochord cells, respectively. Given recent findings, it is necessary to

examine the function of cephalochordate *Brachyury* and also native activity of the *BfBra* enhancer using amphioxus embryos, because the enhancer activity of cephalochordate genes has been examined in the 5'-upstream regions of *Forkhead* (Yu et al., 2004), *engrailed* (Beaster-Jones et al., 2007), and *actin* (Feng et al., 2006). Nevertheless, the present study offers several insights into the evolution of *Brachyury* expression and function relative to formation of the chordate-specific organs, the notochord and somites.

The first suggestion is that not only the 5'-upstream, but also the 3'-downstream and introns are involved in the enhancer activity of *Brachyury* genes. Enhancers of *Brachyury* have been studied in the urochordates, *Ciona intestinalis* (Corbo et al., 1997; Takahashi et al., 1999) and *Halocynthia roretzi* (Matsumoto et al., 2007), the zebrafish *Danio rerio* (Harvey et al., 2010) and the frog *Xenopus laevis* (Latinkic et al., 1997; Lerchner et al., 2000), although ambulacrarian *Brachyury* remains to be explored. However, the foregoing experiments examined only 5'-upstream sequences of the genes. In *Ciona* and *Halocynthia,* the proximal 5'-upstream sequences

that are requisite for reporter expression in notochord cells include Zic-, Ets-, and Snail-binding sites (Corbo et al., 1997; Takahashi et al., 1999). The Zic- and Ets-biding sites function as activator modules, the former for the element of ZicL transcription factor and the latter for the Ets-binding element through FGF-signaling cascade, both of which act upstream for *Brachyury* expression. On the other hand, the Snail-binding site functions as a repressor, because *Ci-snail* is expressed in both notochord and muscle, suppressing *Ci-Bra* in notochord cells (Fujiwara et al., 1998). The 5'-upstream sequences of zebrafish *Brachyury* or *no tail* (*ntl*) have enhancers required for expression in both blastopore and notochord (Harvey et al., 2010). The 5'-upstream sequences of *Xenopus Brachyury* contain a complex set of transcription factor binding sites required for gene expression in both blastopore and notochord, which include *paired*-type homeobox-related proteins and Smad-interacting protein (Latinkic et al., 1997; Lerchner et al., 2000). Because enhancer activity in 3'-downstream sequences and introns has not been examined in these chordates, the present study gives greater impetus to assess such activities in ascidian and vertebrate *Brachyury*.

The genomic arrangement of *Brachyury*, with seven exons interrupted by six introns, is conserved in most metazoans, although intron lengths differ by intron and by species (Inoue et al., unpublished). In *C. intestinalis*, intron lengths are less than 50 bp, suggesting that they have reduced enhancer activity. However, it is intriguing to ask whether the 3'-downstream of *Ci-Bra* has enhancer activity in *Ciona* embryos, and in vertebrate *Brachyury* as well. If they do, the regulatory enhancer machinery of *Brachyury* is more complex than we currently understand.

The present results suggest greater complexity in enhancer machinery responsible for *Brachyury* expression in basal chordates, than was previously supposed. Our data show that cephalochordate *Brachyury* enhancers function normally in the urochordate, *Ciona*. The *lacZ* expression of *BfBra* reporter constructs in the notochord of *Ciona* embryos seems reasonable, because *BfBra* is expressed in notochord-forming regions of amphioxus embryos and *Ci-Bra* is also expressed in notochord cells of urochordate embryos. Therefore, shared mechanisms might promote *BfBra* reporter expression in *Ciona*. It should be noted that notochord-specific enhancer activity is seen

in the 5'-upstream region of *Ci-Bra*, but in the 3'-downstream region of *BfBra* (Figure 6). In cephalochordate *Brachyury,* the 5'-upstream region seems responsible for expression in muscle (Figure 4). At present, we cannot readily explain this discrepancy. Hopefully, an examination of enhancer activity of the *Ci-Bra* 3'-downstream will help to answer this question.

However, a contradiction arises from the result that *Brbra* 5'-upstream constructs showed muscle-specific expression in the *Ciona* host-system. This is unexpected because urochordate embryos express *Brachyury* exclusively in notochord, not in muscle, and because 5'-upstream reporter constructs of *Ci-Bra* and *Hr-Bra* show no *lacZ* expression in muscle cells of ascidian embryos (Corbo et al., 1997; Matsumoto et al., 2007), because muscle expression of *Ci-Bra* is suppressed by Snail (Fujiwara et al., 1998). Possible binding sites of transcription factors in the 5'-upstream of both *BfBra1* and *BfBra2*, which include binding sites for Snail, MyoD, E-box, and T-box, respectively (Figure 5). MyoD and presumably E-box, as well, might be involved in muscle-specific expression of the gene. In addition, a T-box factor, Tbx6 is a key regulatory factor that mediates the function of the maternally-provided muscle-

determinant, Macho-1, and zygotically-expressed MyoD in *Ciona* embryos (Nishida & Sawada, 2001; Mitani et al., 2001; Yagi et al., 2005). A possible explanation is that *Ciona* Snail does not recognize the Snail sites of *BfBra.* Alternatively, due to three muscle-specific transcription-factor binding sites, the ability and/or grade of muscle-specific expression of *BfBra-lacZ* might overcome suppression by Snail. Although this has to be examined in amphioxus embryos, these results strongly suggest that the 5'-upstream of *BfBra* has enhancer activity for muscle-specific expression of the genes in amphioxus embryos.

Another important point is the change in the ratio of positive embryos with *lacZ* expression. In general, the ratio of positive embryos is higher at 7 h than at 11 h. In *Ciona* embryos, *Ci-Bra* expression is intense in the gastrula stage, but is reduced in tailbud embryos (Yasuo & Satoh, 1993; Corbo et al., 1997). *Ci-Bra* has an autonomous repression system to reduce expression activity in later stage embryos (Imai et al., 2006). Reduced numbers of *BfBra lacZ*-positive embryos might reflect endogenous *Ci-*

Bra expression control. In other words, results of the present reporter assay may not be artificial, but may reflect some feature of native *Ciona* embryogenesis.

Discrete expression of reporter constructs of amphioxus *Brachyury* in the *Ciona* system indicates that enhancer mechanism(s) of *Brachyury* expression are shared by cephalochordates and urochordates. When *Brachyury* expression profiles are compared between the two taxa, it appears that amphioxus *Brachyury* is more primitive or basal, whereas ascidian *Brachyury* is more derived and specific to this lineage (Satoh, 2016). While amphioxus *Brachyury* is expressed first in the blastopore of early gastrulae and then in enfolding regions of lateral somites and the mid-dorsal notochord (Holland et al., 1995; Terazawa & Satoh, 1995), ascidian *Brachyury* is expressed only in notochord (Yasuo & Satoh, 1993; Corbo et al., 1997). In other words, it has been argued that *Brachyury* PEF seems to have disappeared in urochordates. In ascidians, gastrulation takes place by ingression of endo-mesoderm cells at around the 110-cell stage, followed by coverage of the embryo by epidermal cells. The so-called marginal zone of the blastopore is not always visible during ascidian gastrulation. However, the

lacZ expression profiles of some *BfBra* constructs show some suggestion of PEF in *Ciona* embryos. For example, *BfBra lacZ* expression in both muscle and notochord in 5-h embryos (Figure 8) and 7-h embryos (Figure 7d) suggests some reminiscent PEF-like expression of genes around the blastopore. Of course, this discussion of PEF and SEF in ascidian embryos is not conclusive, but the present results provide cues for reconsideration or interpretation of the lack of PEF in ascidian embryos.

Another interesting issue regarding cephalochordate *Brachyury* is whether *BfBra1* or *BfBra2* is the ancestral gene (Figure 9). *Brachyury* is duplicated in cephalochordates. Recently Inoue et al. (2017) documented the monophyletic origin of tandemly arranged *Brachyury* genes of cephalochordates, which occur only in this lineage, suggesting a single copy of *Brachyury* in the chordate stem. The genomic region that includes amphioxus *Brachyury* genes is likely too naive to accept duplication of genes, since NOTUM and BVES, neighbors of *BfBra*, are also in a tandem duplicated state. The present study illustrates a difference between *BfBra1* and *BfBra2*. In the 5' and 3' regions and in introns, signal intensity was higher and broader

in *BfBra2* than *BfBra1*. In some cases of *BfBra2*, *lacZ* expression was expanded to posterior muscle cells. Intron enhancers displayed yet another difference. Primary expression of *BfBra1* occurs in notochord, whereas both muscle and notochord manifest expression of *BfBra2*. Together with other data, we prefer an evo-devo scenario in which *BfBra2* is more ancestral and *BfBra1* more derived (Figure 9). Originally, ancestral *BfBra2* is likely to have acquired the enhancer machinery for gene expression in muscle and notochord, as deduced from its spatial expression profile in the 5' region for somite expression, in the 3' region for notochord expression, and in introns for somite/notochord expression. In contrast, sister *BfBra1* became more specialized for notochord expression using the intron enhancers.

In conclusion, using the *Ciona* system, the present study demonstrated complex enhancer activity of cephalochordate *Brachyury* for notochord- and somite-specific expression. Since these mesodermal organs are essential for tail-beating larvae, which are thought to be essential for the origin and evolution of chordates (Satoh, 2016), the acquisition of such enhancer activity is critical to understanding chordate evolution.

4 METHODS

4.1 Animals and embryos

Adult *Branchiostoma floridae* were collected from Tampa Bay, Florida. Adult *Ciona intestinalis* were obtained through National Bio-Resource Projects of Japan or collected at Gamagori, Aichi Prefecture. *C. intestinalis* eggs and sperm were surgically collected and handled as described previously (Takahashi et al., 1999). After fertilization, embryos were cultured at 18℃.

4.2 Vista comparison of genomic region containing *Brachyury* genes

The genomes of *Branchiostoma floridae* (Putnam et al., 2008) and *B. belcheri* (Wang et al., 2014) have been decoded. Sequences of genomic regions that contain *Brachyury* genes of *B. floridae* were retrieved from *Branchiostoma floridae*-JGI Genome Portal

(scaffold 65: 2.5Mbp~2.6Mbp), and *B. belcheri* from LanceletDB (scaffold 45: 0.45Mbp~0.55Mbp). Sequence comparisons were carried out using VIST (*https://www.sanfrancisco.va.gov/services/VIST.asp*).

4.3 Plasmid construction

We constructed reporter plasmids from the *B. floridae Brachyury* locus that includes these regions. Genomic DNA was isolated from sperm of a single adult *B. floridae* using a DNeasy tissue kit (Qiagen). To construct the 5' upstream reporter plasmid, In-Fusion cloning technology (Clontech) was used to create pPD1.27 constructs. All PCR amplification was done using KOD plus DNA polymerase (Toyobo). Genomic fragments were amplified from adult amphioxus sperm DNA using PCR with a pair of genomic fragment-specific primers (Table 2) and were subcloned into pPD1.27.

3' downstream and introgenic sequences of *BfBra1* and *BfBra2* were amplified from adult amphioxus sperm DNA using PCR with the primers listed in Table 2. In preparing constructs, we used *C. intestinalis Brachyury* basal promoter, a device to

activate *LacZ* expression in *Ciona* embryos, which itself has no reporter expression activity and no influence from the spatial expression of *LacZ*. We subcloned them into the pSP1.72 *Ci-Bra* basal promoter>*LacZ* vector (Corbo et al., 1997; Wang & Christiaen, 2012), which includes -3 bp upstream of the TATA element.

4.4 Electroporation and lacZ detection

Electroporation was performed according to previously published protocols (Corbo et al., 1997; Wang & Christiaen, 2012). 80-µg aliquots of plasmid DNA 0.1 µg/mL were electroporated using a BioRad Gene Pulser. After electroporation, embryos were maintained in gelatin-coated dishes with filtered seawater containing 50 µg/mL streptomycin sulfate at 18℃. Embryos were fixed for 30 min at room temperature in 0.5 M NaCl, 27 mM KCl, 2 mM EDTA (pH 8.0) containing 1 % glutaraldehyde. Fixed embryos were rinsed in phosphate-buffered saline (PBS) and incubated in PBS containing 250 µM 5-bromo-4-chloro-3-indoly-b-D-galactopyranoside (X-gal), 0.1 % Triton X-100, 1 mM $MgCl_2$, 3 mM $K_4[Fe(CN)_6]$, and 3 mM $K_3[Fe(CN)_6]$ at 37 ℃ for

30 min. Stained embryos were washed in PBS to stop the reaction and observed under an Olympus stereomicroscope.

ACKNOWLEDGMENTS

We thank Dr. Yuuri Yasuoka for comments on the manuscript, Kanako Hisata for technical assistance, and Dr. Steven Aird for English editing. We thank the National Bio-Resource Project (MEXT, Japan) for providing the *Ciona* adults, and Dr. Yasuhisa Henmi for providing the Amphioxus adults. We thank Functional Genomics Facility, Spectrography and Bioimaging Facility, NIBB Core Research Facilities. This work was supported by Grants-in-Aid from the Japan Society for the Japan Science (JSPS) KAKENHI Grant Numbers 18K06370 to H. Takahashi and 16H04824 to N. Satoh.

Competing interests: The authors declare no competing financial interests.

References

Beaster-Jones, L., Schubert, M., & Holland, L.Z. (2007). Cis-regulation of the amphioxus *engrailed* gene: Insights into evolution of a muscle-specific enhancer. *Mech. Dev.* **124**, 532-542.

Bielen, H. et al. (2007). Divergent functions of two ancient Hydra *Brachyury* paralogues suggest specific roles for their C-terminal domains in tissue fate induction. *Development* **134,** 4187–4197.

Bourlat, S.J., Juliusdottir, T., Lowe, C.J., Freeman, R., Aronowicz, J., Kirchner, M., Lander, E.S., Thorndyke, M., Nakano, H., Kohn, A.B., Heyland, A., Copley, R.R., & Telford, M.J. (2006). Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85-88.

Cameron, C.B., Garey, J.R., & Swall, B.J. (2000). Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *Proc. Natl Acad. Sci., USA*, **97**, 4469-4474.

Conklin, E.G. (1932). The embryology of amphioxus. *J. Morphol*. **54**, 69-151.

Corbo, J. C., Levine, M., & Zeller, R. W. (1997). Characterization of a notochord-specific enhancer from the *Brachyury* promoter region of the ascidian, *Ciona intestinalis*. *Development* **124,** 589–602.

Davidson, E. H. (2006). The Regulatory Genome, Gene Regulatory Networks in Development and Evolution (Academic/Elsevier, London).

Delsuc, F., Brinkman, H., Chourrout, P., & Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439,** 965-968.

Feng, J., Li, G., Liu, X., Wang, J., & Wang, Y-Q. (2006). Functional analysis of the promoter region of amphioxus b-actin gene: a useful tool for driving gene expression in vivo. *Mol. Biol. Rep*. **41**, 6817-6826.

Fujiwara. S., Corbo, J. C., & Levine, M. (1998). The snail repressor establishes a muscle/notochord boundary in *Ciona* embryo. *Development* **125,** 2511-2520.

Gee, H. (1999). Before Backbone: Views on the Origin of the Vertebrates. Chapman & Hall, London.

Gerhart, J. (2006). The deuterostome ancestor. *J. Cell. Phyol.* **209**, 677-685.

Halanych, K.M. The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. *Mol. Phylogen. Evol*. **4**, 72-77 (1995).

Harvey, S. A., Tümple, S., Dubrulle, J., Schier, A. F., & Smith, J. C. (2010). *no tail* integrates two modes of mesoderm induction. *Development* **137,** 1127-1135.

Hirakow, R. & Kajita, N. (1994). Electron microscopic study of the development of amphioxus, Branchiostoma belcheri tsingtauense: neurula and larva. *Acta Anat. Nippon.* **69**, 1-13.

Holland, P. W., Koschorz, B., Holland, L. Z., & Herrmann, B. G. (1995). Conservation of *Brachyury* (*T*) genes in amphioxus and vertebrates: developmental and evolutionary implications. *Development* **121,** 4283–4291.

Holland, L. (2000). Body-plan evolution in the Birateria: early antero-posterior patterning and the deuterostome-protostome dichotomy. *Curr. Opin. Genet. Dev*. **10**, 434-442.

Holland, N. D., Holland, L. Z., & Holland, P. W. H. (2015). Scenarios for the making of vertebrates. *Nature* **520,** 450-455.

Imai, K., Levine, M., Satoh, N., & Satou, Y. (2006). Regulatory blueprint for a chordate embryo. *Science* **312**, 1183-1187.

Inoue, J., Yasuoka, Y., Takahashi, H., & Satoh, N. (2017). The chordate ancestor possessed a single copy of the *Brachyury* gene for notochord acquisition. *Zoological. Lett.* **3,** 4.

Inoue, J. & Satoh, N. (2018). Deuterostome genomics: Lineage-specific protein expansions that enabled chordate muscle formation. *Mol. Biol. Evol*. **35**, 914-924.

Latinkic, B. V., Umbhauer, M., Neal, K. A., Lerchner, W., Smith, J. C., & Cunliffe, V. (1997). The *Xenopus Brachyury* promoter is activated by FGF and low concentrations

of activin and suppressed by high concentrations of activin and by paired-type homeodomain protiens. *Genes. Dev.* **11,** 3265-3276.

Lerchner, W., Latinkic, B. V., Remacle, J. E., Huylebroeck, D., & Smith, J. C. (2000). Region-specific activation of *Xenopus Brachyury* promoter involves active repression in ectoderm and endoderm:a study using transgenic frog embryos. *Development* **127,** 2729-2739.

Lowe, C. J., Clarke, D. N., Medeiros, D. M., Rokhsar, D. S., & Gerhart, J. (2015). The deuterostome context of chordate origins. *Nature* **520,** 456-465.

Matumoto, J., Kumano, G., & Nishida, H. (2007). Direct activation by Ets and Zic is required for initial expression of the *Brachyury* gene in the ascidian notochord. *Dev. Biol.* **306,** 870-882.

Mitani, Y., Takahashi, H., & Satoh, N. (2001). Regulation of the muscle-specific expression and function of an ascidian T-box gene, As-T2. *Development* **128,** 3717-3728.

Nishida, H., & Sawada, K. (2001). *macho-1* encodes a localized mRNA in ascidian eggs that specifies muscle fate during embryogenesis. *Nature* **409,** 724-729.

Putnam, N. H. et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453,** 1064–1071.

Ruppert, E.E. (1997). Cephalochordata (Acrania). In Harrison F.W. & Ruppert, E.E. eds. Microscopic Anatomy of Invertebrates. Hemichordata, Chaetognatha, and the Invertebrate Chordates, vol.15 Wiley-Liss, New York, pp. 349-504.

Satoh, N. (2009). Advanced filter-feeder hypothesis for urochordate evolution. *Zool. Sci*. **26**, 97-111.

Satoh, N. (2014). Developmental Genomics of Ascidians. Wiley-Blackwell, New Jersey, USA.

Satoh, N. (2016). Chordate Origins and Evolution: the Molecular Evolutionary Road to Vertebrates. Boston: Elsevier.

Satoh, N., Tagawa, K. & Takahashi, H. (2012). How was the notochord born? *Evol. Dev.* **14,** 56–75.

Schulte-Merker, S., van Eeden, F.J., Halpern, M.E., Kimmel, C.B., & Nusslein-Volhard, C. (1994). *no tail (ntl)* is the zebrafish homolog of the mouse *T* (*Brachyury*) gene. *Development* **120**, 1009-1015.

Simakov, O., Kawashima, T., Malrletaz, F. et al. (2013). Hemichordate genomes and deuterostome origins. *Nature* **527**, 459-465.

Suzuki, M.M. & Satoh, N. (2000). Genes expressed in the amphioxus notochord revealed by EST analysis. Dev. Biol. 224, 168-177.

Takahashi, H., Mitani, Y., Satoh, G., & Satoh, N. (1999). Evolutionary alterations of the minimal promoter for notochord-specific *Brachyury* expression in ascidian embryos. *Development* **126,** 3725-3734.

Technau, U. (2001). *Brachyury*, the blastopore and the evolution of the mesoderm. *Bioessays* **23,** 788–794.

Terazawa, K. & Satoh, N. (1995). Spatial expression of the amphioxus homolog of *Brachyury* (*T*) gene during early embryogenesis of *Branchiostoma belcheri*. *Dev. Growth. Differ.* **37,** 395–401.

Wang, W. & Christiaen, L. (2012). Transcriptional enhancers in ascidian development. *Curr. Top. Dev. Biol.* **98,**147-172.

Wang, R., Chen, Z., Yan, X. et al. (2014). Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nature Comm*. **5**, 5896.

Wilkinson, D.G., Bhatt, S., & Hermann, B.G. (1990). Expression pattern of the mouse T-gene and its role in mesoderm formation. *Nature* **343**, 657-659.

Yagi, K., Takatori, N., Satou, Y., & Satoh, N. (2005). Ci-Tbx6b and Ci-Tbx6c are key mediators of the maternal effect gene *Ci-macho1* in muscle cell differentiation in *Ciona intestinalis* embryos. *Dev. Biol.* **282,** 535-549.

Yasuo, H. & Satoh, N. (1993). Function of vertebrate *T* gene. *Nature* **364,** 582–583.

Yu, J-K., Holland, N.D., & Holland, L.Z. (2004). Tissue-specific expression of *FoxD* reporter constructs in amphioxus embryos. *Dev. Biol*. **274**, 452-461.

Yue, J-X, Kozmikova, I., Hiroki Ono H. et al. (2016). Conserved noncoding elements in the most distant genera of cephalochordates: the goldilocks principle. *Genome Biol. Evol*. **8**, 2387–2405.

Figure Legends

FIGURE 1 Schematic drawing of *Brachyury* (*Bra*) expression in chordates. (a) *Bra* is expressed in blastopore (primary expression, shown in brown) and in dorsal-midline from which notochord is formed (secondary expression, shown in blue). (**b**) A cephalochordate (*Branchiostoma*) *Bra* is expressed both in notochord (blue) and myogenic somites (red) whereas a urochordate (*Ciona*) *Bra* is expressed in notochord and its expression is suppressed in myogenic tissue (red dots area).

FIGURE 2 Sequence similarity of *BfBra1* and *BfBra2.* (a) An alignment of *BfBra1* (green) and *BfBra2* (red) in the *B. floridae* genome. Vertical lines represent exons and thickened bars show 5'- and 3'-non-coding sequences. (**b)** Amino acid and (**c)** nucleotide sequences of *BfBra1* and *BfBra2* highly resemble each other. (**b)** Sequence identity is shown by asterisks and T-box is shown by dark yellow box. (**c)** Sequence identity is shown by vertical lines while differences are indicated by red. Black arrowheads show sites of intron insertion.

FIGURE 3 Sequence comparison of amphioxus *Brachyury* genes. (**a**) Vista alignment of *BfBra2* using a query of *BfBra1* nucleotide sequences. The Y axis shows percent conservation over a sliding 100-bp window (line is 50% identity). Blue areas show genomic regions with conserved coding sequences and brown areas show similar non-coding sequences (CNS). There are two peaks of CNS in the 5'-upstream sequence of *BfBra2* and *BfBra1*. (**b**) Vista alignment of *B. belcheri Brachyury* with a query from *B. floridae Brachyury* nucleotide sequences. CNSs are seen not only encountered in the 5' upstream, but also in the 3'-downstream and intron sequences. Black arrows indicate 5'and 3' sequences of the genes.

FIGURE 4 Enhancer activity of 5'-upstream regions of *BfBra1* and *BfBra2*. (a) *lacZ*-reporters (arrows) were constructed with 3-kb and 2-kb upstream sequences of *BfBra1* and *BfBra2,* respectively, which include conserved non-coding sequences (CNSs) shared by the two genes (see d). **(b)** Reporter expression of *BfBra2-2kbp* in *Ciona* embryos at neurula (8 h, upper) and tailbud stages (12 h, lower), and their

schematic drawings are shown at the left. The numbers of embryos examined (n=428 at 8 h and in n=362 at 12 h) and percentages of embryos with reporter expression are shown at the right. Green, expression in muscle; blue, no expression. (**c**)Reporter expression of *BfBra1-3kbp. lacZ* is expressed in muscle. (**d)** VISTA analysis shows two CNSs (proximal in blue and distal in orange). Expression of reporter constructs of -5.7, -2 and -1 kbp of *BfBra2,* including the proximal and distal CNS, while the second two include the CNS proximal only. *lacZ* expression of -5.5, -3, -2, -12 and -0.5 kbp of *BfBra1*. The first two include proximal and distal CNS while second two with proximal only and the third with no CNS. +, expression; -, no expression.

FIGURE 5 Possible sequence-specific DNA binding motifs seen in the 5'-upstream sequence of *Bra2* and *Bra1.* (a) *BfBra1* and *BfBra2* and (**b)** *BfBra1*, *BfBra2, BbBra1* and *BbBra2*. Those include Snail (green), MyoD (red), E-box (blue), and T-box (yellow), respectively. Sequence differences between the two genes are shown in red (**a, b**).

FIGURE 6 Enhancer activity of 3' downstream regions of *BfBra1* and *BfBra2*.

(**a)** *lacZ*-reporters were constructed with the 3-kb downstream sequences of *BfBra1* and *BfBra2,* respectively, which include several CNSs shared by the two lancelets (see Fig. S2b). (**b, c)** Reporter expression of *BfBra2+3kbp* (**b**) and *BfBra1+3kbp* (**c**) in *Ciona* embryos at neurula (7 h; upper) and tailbud stages (11 h; lower), and respective schematic drawings are shown left. Numbers of embryos examined and percentages of embryos with expression are shown at right. Red, expression in notochord; blue, no expression.

FIGURE 7 Enhancer activity of intragenic regions of *BfBra1* (a, b) and *BfBra2* (c, d). Both genes possess six introns (1-6 in *BfBra1* and 1'-6' in *BfBra2*). Enhancer activity was examined at each intron sequence mentioned above. Expression of reporter constructs in *Ciona* embryos was as previously described. All intragenic regions showed activity, except for I-5 and I-6 of *BfBra1*. Interestingly, *BfBra1* introns have enhancer activity primarily in *Ciona* notochord cells (**b**), whereas *BfBra2* introns have

activity primarily in muscle cells (**d**). In addition, *BfBra2* exhibits stronger activity than *BfBra1*.

FIGURE 8 Reporter activity of intron-3 and intron-5 of *BfBra2* is detected at early gastrula stage, 5h after fertilization. The expression is detected in primordium cells of both notochord and muscle.

FIGURE 9 Summary of results of the present reporter assay (a) and possible interpretation of results (b). Enhancer activity of 5' upstream sequences in muscle is shown in brown and that of the 3' downstream in notochord is in blue. The activity *BfBra2* introns in muscle and notochord is shown in purple.