〈Original Article〉

Population genetic structure of mitochondrial DNA in *Branchiostoma* species

Yuzo Takada[1), 2)] and Toshio Imai[1)]

Summary The *Branchiostoma* are thought to exhibit characteristics of their vertebrate ancestors of the vertebrates and are therefore important organisms in the research fields of phylogenetics, evolutionary biology, and phylogenetic systematics. With the aim of elucidating in detail the process of the evolution of *Branchiostoma*, we determined and compared the entire sequences of mitochondrial DNAs (mtDNAs) obtained from multiple individuals of *Branchiostoma belcheri* (*B. belcheri*; a Pacific lancelet), *Branchiostoma lanceolatum* (*B. lanceolatum*; an Atlantic lancelet), and *Branchiostoma floridae* (*B. floridae*; an Atlantic lancelet), which have frequently been used for research. Results have revealed substantial inter-species diversity in mtDNAs of these *Branchiostoma*. Some individuals of *B. floridae* had an *NADH1* whose initiation codon (ATA) included a base substitution that failed to function as an initiation codon, providing a reason for our selection of the GTG codon, which is located before ATA and is therefore well-conserved among species as the initiation codon for this species.

Key words: *Branchiostoma belcheri*, *Branchiostoma lanceolatum*, *Branchiostoma floridae*, Mitochondrial DNA, Initiation codon

1. Introduction

The *Branchiostoma*, which are classified under the phylum Chordata, subphylum Cephalochordata, exhibit the following characteristics. They are about 30-50 mm long and have a fish-like appearance. They have no head, and instead show a well-developed oral tegmentum covering the feeding organ in the anteroventral region. They obtain oxygen from water via a branchial cleft located in the anterior half of the body that is used for filtering food. They have a closed blood-vascular system without a heart and a photoreceptive organ but no organs equivalent to an eye. Although organs corresponding to the back and abdominal fins are not apparent, they swim well with a well-developed organ equivalent to a tail fin. They retain a rod-like notochord consisting of muscle tissue and running from the head to tail throughout life. Although, they have a nerve cord along the notochord, they lack neural ganglia or a brain. They possess simplified, vertebrate-like structures, such as muscle and neural tissue. Based on these characteristics, the

[1)]Laboratory of Physiological Chemistry, Department of Biology, Faculty of Science, Toho University, Funabashi 274-8510, Japan
[2)]Department of Forensic Medicine, National Defense Medical College, Tokorozawa 359-8513, Japan
Corresponding author: Yuzo Takada
Department of Forensic Medicine, National Defense Medical College
Namiki 3-2, Tokorozawa, Saitama 359-8513, Japan
Received for Publication December 12, 2008
Accepted for Publication January 15, 2009

Branchiostoma are believed to resemble their vertebrate ancestors[1]. *Branchiostoma* were first examined by P.S. Pallas in 1774[2] and have been very important organisms for exploring the ancestry and evolution of the vertebrates. Findings from both classic zoological studies as well as from recent genetic analyses suggest that *Branchiostoma* are closer to vertebrates than to ascidians[3, 4]. With this background, a genome-sequencing project for *Branchiostoma* as a laboratory animal was launched in 2005. Sequencing was completed in 2007 (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html), and the sequence has been published[5]. Meanwhile, the sequences of mitochondrial DNA (mtDNA) have been determined for *Branchiostoma belcheri* (*B. belcheri*)[6], *Branchiostoma lanceolatum* (*B. lanceolatum*)[7], and *Branchiostoma floridae* (*B. floridae*)[8]. In this study, with the aim of elucidating in detail the process of the evolution of *Branchiostoma*, we analyzed the sequences of mtDNAs obtained from multiple individuals of the above three species that have been studied by many investigators worldwide.

2. Materials and methods

1. Samples

Ten individuals of *B. belcheri* (T01-10), a Pacific species, were collected in waters close to Japan. Ten individuals of *B. lanceolatum* (M01-10), an Atlantic species, collected in Argeles-sur-Mer on the Mediterranean coast of France, were kindly provided by Dr. Michael Schubert from Institut de Génomique Fonctionnelle de Lyon (IGFL). Twenty individuals of *B. floridae* collected in Tampa Bay, Florida, in the USA were kindly provided by Dr. Michael Schubert from IGFL (M01-10) and Dr. Linda Holland from the Scripps Institution of Oceanography University of California-San Diego (H01-10).

2. DNA preparation, PCR amplification, sequencing, and data analysis

DNAs were obtained by crushing of *Branchiostoma* samples with a multi-bead shocker (YASUI KIKAI, Osaka, JPN), followed by purification with the DNeasy Tissue Kit (QIAGEN, Hilden, GER). PCR for sequencing mtDNAs was performed using previously-reported oligonucleotide primers for *B. belcheri*[6], or primers designed based on the data reported by Nohara *et al.*[9] for *B. lanceolatum* or data reported by Spruyt *et al.*[7] for *B. floridae*. A PCR mixture was prepared by adding 5 ng of a DNA sample, 400 nM of a primer set, and an appropriate amount of sterile distilled water to 10 μl of GeneAmp® Fast PCR Master Mix (2x) (Applied Biosystems, CA, USA) to produce a total volume of 20 μl. PCR was performed with an Applied Biosystems 9800 Fast PCR Thermal Cycler (Applied Biosystems) under the following conditions: preincubation at 95℃ for 20 seconds, thermal denaturation at 95℃ for 3 seconds, and 40 cycles of annealing and extension at 60℃ for 30 seconds. The PCR segments obtained were subjected to a sequencing reaction using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), followed by electrophoresis with an Applied Biosystems 3130xl Genetic Analyzer. The detection of electrophoretic patterns and an analysis of nucleotide sequences were performed using Sequencing Analysis Software v5.2 (Applied Biosystems). The entire sequence of the mtDNA was determined using GENETYX-MAC/ATSQ ver.4.0 Software (GENETYX, Tokyo, JPN). Nucleotide sequences were translated into amino acid sequences using EMBOOS Transeq (http://www. ebi.ac.uk/Tools/emboss/transeq/index. html) provided by EMB-EBI. The structures of transferRNA (tRNA) genes were analyzed using tRNAscan-SE 1.21 (http://lowelab. ucsc.edu/tRNAscan-SE/) provided by Todd M. Lowe[10]. Multiple sequence alignment analysis was performed using a CLUSTAL-W program (http://www.ddbj.nig.ac.jp/Welcome-j.html) provided by the DNA Data Bank of Japan (DDBJ)[11-13]. A molecular evolutionary tree was created using the NJplot (http://pbil.univ-lyon1.fr/software/njplot.html)[14], with the mtDNA sequence of *Lampetra fluviatilis* (Lam: Y18683)[15] as an outgroup.

3. Results

1. Mitochondrial DNA size

The sizes of mtDNAs analyzed were as follows:

Table 1 Localization of mtDNA genes and non-coding regions of Branchiostoma belcheri, Branchiostoma lanceolatum and Branchiostoma floridae

Branchiostoma belcheri (*Bb*-T01): 15,075 bp					
Gene	from	to	Size (bp)	Size (aa)	Strand
cytb	1	1141	1141	380	
tRNA-Thr	1143	1207	65		L
tRNA-Pro	1208	1272	65		
12S rRNA	1273	2119	847		
tRNA-Phe	2120	2182	63		
tRNA-Val	2186	2257	72		
16S rRNA	2258	3640	1383		
tRNA-Leu (TTR)	3641	3710	70		
NADH1	3711	4655	945	314	
tRNA-Ile	4656	4718	63		
tRAN-Met	4719	4785	67		
tRNA-Gln	4786	4853	68		L
NADH2	4855	5895	1041	346	
tRNA-Asn	5888	5959	72		L
tRNA-Trp	5960	6026	67		
tRAN-Ala	6025	6087	63		L
tRNA-Cys	6088	6146	59		L
tRNA-Tyr	6147	6211	65		L
COI	6216	7763	1548	515	
tRNA-Ser (TCN)	7761	7831	71		L
tRNA-Asp	7840	7902	63		
COII	7906	8596	691	230	
tRNA-Lys	8597	8661	65		
ATP8	8662	8826	165	54	
ATP6	8820	9502	683	227	
COIII	9503	10291	789	262	
NADH3	10292	10645	354	117	
tRNA-Arg	10646	10707	62		
NADH4L	10711	10986	276	91	
NADH4	10988	12346	1359	452	
tRNA-His	12347	12413	67		
tRNA-Ser (AGY)	12414	12480	67		
tRNA-Leu (CTN)	12481	12546	66		
NADH5	12547	14343	1797	598	
NC sequence	14344	14453	110		
tRNA-Gly	14454	14520	67		
NADH6	14502	15005	504	167	L
tRNA-Glu	15007	15074	68		L

Branchiostoma lanceolatum (*Bl*-M01): 15,139 bp					
Gene	from	to	Size (bp)	Size (aa)	Strand
cytb	1	1143	1143	380	
tRNA-Thr	1144	1208	65		L
tRNA-Pro	1209	1275	67		
12S rRNA	1276	2118	843		
tRNA-Phe	2119	2184	66		
tRNA-Val	2187	2253	67		
16S rRNA	2254	3629	1376		
tRNA-Leu (TTR)	3630	3700	71		
NADH1	3704	4643	940	313	
tRNA-Ile	4644	4707	64		
tRAN-Met	4708	4774	67		
tRNA-Gln	4774	4842	69		L
NADH2	4844	5884	1041	346	
tRNA-Asn	5877	5943	67		L
tRNA-Trp	5943	6008	66		
tRAN-Ala	6008	6069	62		L
tRNA-Cys	6070	6128	59		L
tRNA-Tyr	6129	6194	66		L
COI	6205	7752	1548	515	
tRNA-Ser (TCN)	7750	7820	71		L
tRNA-Asp	7828	7893	66		
COII	7894	8584	691	230	
tRNA-Lys	8585	8650	66		
ATP8	8651	8815	165	54	
ATP6	8809	9491	683	227	
COIII	9492	10280	789	262	
NADH3	10281	10634	354	117	
tRNA-Arg	10635	10697	63		
NADH4L	10700	10974	275	91	
NADH4	10975	12332	1358	452	
tRNA-His	12333	12397	65		
tRNA-Ser (AGY)	12398	12464	67		
tRNA-Leu (CTN)	12465	12532	68		
NADH5	12533	14329	1797	598	
NC sequence	14330	14519	190		
tRNA-Gly	14520	14586	67		
NADH6	14569	15071	503	167	L
tRNA-Glu	15073	15138	66		L

Branchiostoma floridae (*Bf*-M01): 15,076 bp					
Gene	from	to	Size (bp)	Size (aa)	Strand
cytb	1	1143	1143	380	
tRNA-Thr	1144	1207	64		L
tRNA-Pro	1208	1273	66		
12S rRNA	1274	2119	846		
tRNA-Phe	2120	2185	66		
tRNA-Val	2188	2254	67		
16S rRNA	2255	3621	1367		
tRNA-Leu (TTR)	3622	3692	71		
NADH1	3693	4635	940	314	
tRNA-Ile	4636	4698	63		
tRAN-Met	4699	4765	67		
tRNA-Gln	4765	4833	69		L
NADH2	4835	5875	1041	346	
tRNA-Asn	5868	5937	70		L
tRNA-Trp	5937	6005	69		
tRAN-Ala	6004	6066	63		L
tRNA-Cys	6067	6124	58		L
tRNA-Tyr	6125	6190	66		L
COI	6201	7748	1548	515	
tRNA-Ser (TCN)	7746	7816	70		L
tRNA-Asp	7825	7889	65		
COII	7890	8580	691	230	
tRNA-Lys	8581	8646	66		
ATP8	8647	8811	165	54	
ATP6	8805	9487	683	227	
COIII	9488	10276	788	262	
NADH3	10277	16028	352	117	
tRNA-Arg	10629	10695	67		
NADH4L	10698	10972	275	91	
NADH4	10973	12330	1358	452	
tRNA-His	12331	12396	66		
tRNA-Ser (AGY)	12397	12462	66		
tRNA-Leu (CTN)	12463	12530	68		
NADH5	12531	14327	1797	598	
NC sequence	14328	14456	129		
tRNA-Gly	14457	14524	68		
NADH6	14506	15009	504	167	L
tRNA-Glu	15011	15075	65		L

15,075 base pairs (bp) for *B. belcheri*, 15,139-15,141 bp for *B. lanceolatum*, and 15,075-15,077 bp for *B. floridae*. The mtDNAs of *Branchiostoma* species are the shortest among mtDNAs of the deuterostomes analyzed thus far. Nevertheless, the mtDNAs of *Branchiostoma* species contain as many genes as those of other larger genomes (13 protein-coding genes, 2 ribosomal RNA (rRNA) genes, and 22 tRNA genes) (Table 1).

The mean frequencies of each base in mtDNA were as follows: 26.8% for adenine (A), 15.4% for cytosine (C), 21.5% for guanine (G), and 36.2% for thymine (T), with purine bases accounting for 48.3% and pyrimidine bases for 51.7%. With respect to base frequency, no difference was observed within each species of *B. belcheri*, *B. lanceolatum*, and *B. floridae*, while a maximum 3.5% difference was observed inter-species. The mean GC content was 37.0%, with no intra-species difference and a maximum of 3.0% in inter-species differences (Table 2).

2. Protein-coding genes

With regard to the size of protein-coding genes, the number of bases in the *NADH1* differed between species, 945 bp for *B. belcheri*, 940 bp for *B. lanceolatum*, and 943 bp for *B. floridae*. The difference in the number of bases was reflected in a difference in the number of amino acids, 314 amino acids (aa) for *B. belcheri* and *B. floridae* and 313 aa for *B. lanceolatum*, with 1 aa shorter in *B. lanceolatum* than in the other two. The sizes of other genes were the same among all *Branchiostoma* species tested (Table 1).

Eleven of 13 protein-coding genes of *B. belcheri* and *B. floridae* and 12 of 13 protein-coding genes of *B. lanceolatum* have an ATG initiation codon. Termination codons are divided into a complete codon consisting of TAG or TAA and an incomplete codon ending with T or TA. Among the protein-coding genes of B. belcheri, the *COI*, *NADH1*, *NADH2*, *NADH4*, *NADH6*, and *ATP8* have a complete termination codon of TAG, while the *COIII*, *NADH3*, *NADH4L*, and *NADH5* have a complete termination codon of TAA, with other genes ending in an incomplete termination codon of T or TA. With regard to *B. lanceolatum*, the *COI*, *cytb*, *NADH2*, *NADH3* and *ATP8* have a complete termination codon of TAG, while the *COIII* and *NADH5* have a complete termination codon of TAA, with other genes ending with an incomplete termination codon of T or TA. In the case of *B. floridae*, the *COI*, *cytb*, *NADH2*, *NADH5*, *NADH6*, and *ATP8* have a complete termination codon of TAG, while the *COIII* has a complete termination codon of TAA, with other genes ending in an incomplete termination codon of T or TA (Table 3).

The mean frequencies of amino acids in protein-coding genes were as follows: 7.7% for glycine (G), 7.3% for alanine (A), 9.2% for valine (V), 16.4% for leucine (L), 6.1% for isoleucine (I), 5.3% for methionine (M), 6.9% for phenylalanine (F), 2.9% for tryptophan (W), 4.2% for proline (P), 7.7% for serine (S), 4.9% for threonine (T), 3.1% for asparagine (N), 2.3% for glutamine (Q), 1.0% for cysteine (C), 2.1% for aspartic acid (D), 2.7% for glutamic acid (E), 1.9% for lysine (K), 2.4% for histidine (H), 2.0% for arginine (R), and 4.0% for tyrosine (Y), with hydrophobic amino acids accounting for 66.0%, neutral amino acids for 19.0%, and hydrophilic amino acids for 15.0% (Table 4). No differences were observed in amino acid frequencies within each species of *B. belcheri*, *B. lanceolatum*, or *B. floridae*, while 0.1-0.7% inter-species differences were observed in frequencies of amino acids other than C, D, and H.

Table 2 Percentage base composition of mitochondrial DNA in *Branchiostoma* species

Base	*Bb*	*Bl*	*Bf*	Mean
Adenine (A)	27.6	26.0	26.9	26.8
Cytosine (C)	13.9	16.4	15.8	15.4
Guanine (G)	21.4	21.8	21.4	21.5
Thymine (T)	37.1	35.7	35.8	36.2
purine bases	48.9	47.8	48.3	48.3
pyrimidine bases	51.1	52.2	51.7	51.7
AT contents	47.0	61.7	62.7	57.1
GC contents	35.3	38.3	37.3	37.0

Data from: *Bb* (*B. belcheri*, T01-10, AB478559-AB478563), *Bl* (*B. lanceolatum*, M01-10, AB478564-AB478573), *Bf* (*B. floridae*, M01-10 and H01-10, AB478574-AB478593) in this study.

Table 3 Initiation and termination codons of the protein-coding genes in Branchiostoma species

Gene	*Bb*		*Bl*		*Bf*	
	Codon		Codon		Codon	
	Initiation	Termination	Initiation	Termination	Initiation	Termination
cytb	ATG	T*	ATG	TAG	ATG	TAG
NADH1	GTG	TAG	ATG	T*	GTG	T*
NADH2	ATG	TAG	ATG	TAG	ATG	TAG
COI	ATG	TAG	GTG	TAG	GTG	TAG
COII	ATG	T*	ATG	T*	ATG	T*
ATP8	ATG	TAG	ATG	TAG	ATG	TAG
ATP6	ATG	TA*	ATG	TA*	ATG	TA*
COIII	ATG	TAA	ATG	TAA	ATG	TAA
NADH3	ATG	TAA	ATG	TAG	ATG	T*
NADH4L	GTG	TAA	ATG	TA*	ATG	TA*
NADH4	ATG	TAG	ATG	TA*	ATG	TA*
NADH5	ATG	TAA	ATG	TAA	ATG	TAG
NADH6	ATG	TAG	ATG	TA*	ATG	TAG

Abbreviations are the same as in Table 2.

Asterisks indicate no such potential reasonably exists and that the termination codon is incomplete.

3. RNAs

The mean length of tRNA genes among the three *Branchiostoma* species was 66.2 nucleotides (nt), while those of tRNA genes varied among species, 66.0 nt for *B. belcheri*, 66.1 nt for *B. lanceolatum*, and 66.4 nt for *B. floridae*. Among tRNA genes, only the $tRNA^{M}$ exhibited the same size in all species, while other genes showed 1-5 nt inter-species differences (Table 1).

As for to rRNA genes, the size of the *12S rRNA* was strikingly similar in all *Branchiostoma* species: 847 bp for *B. belcheri*, 843 bp for *B. lanceolatum*, and 846 bp for *B. floridae*, with a mean of 845 bp. The size of the *16S rRNA* also showed inter-species similarities, 1,383 bp for *B. belcheri*, 1,376 bp for *B. lanceolatum*, and 1,367 bp for *B. floridae*, with a mean of 1,375 bp (Table 1).

4. Unassigned DNA

The size of unassigned DNAs was as small as 136-217 nt. In the mtDNA of *B. belcheri*, 110 nt (unassigned DNAs) were present in a single region between the *NADH5* and $tRNA^{G}$, 8 nt between the $tRNA^{S(TCN)}$ and $tRNA^{D}$, 4 nt between the $tRNA^{Y}$ and *COI*, 3 nt each between the $tRNA^{F}$ and $tRNA^{V}$, between the $tRNA^{D}$ and *COII*, and between the $tRNA^{R}$ and *NADH4L*, and 1 nt each in 5 other regions. With regard to *B. lanceolatum*, 190 nt were observed in a

Table 4 Percentage amino acid composition of protein-coding genes in Branchiostoma species

	Amino acid	*Bb*	*Bl*	*Bf*	Mean
hydrophobic residues	G	7.8	7.6	7.8	7.7
	A	7.4	7.4	7.2	7.3
	V	8.8	9.5	9.2	9.2
	L	16.1	16.6	16.5	16.4
	I	6.5	5.7	6.2	6.1
	M	5.5	4.8	5.4	5.3
	F	6.9	7.0	6.8	6.9
	W	2.9	2.9	2.8	2.9
	P	4.1	4.2	4.2	4.2
	total	66.0	65.7	66.1	66.0
neutral residues	S	7.4	7.9	7.8	7.7
	T	4.9	5.1	4.7	4.9
	N	3.2	3.1	3.0	3.1
	Q	2.3	2.2	2.3	2.3
	C	1.0	1.0	1.0	1.0
	total	18.8	19.3	18.8	19.0
hydrophilic residues	D	2.1	2.1	2.1	2.1
	E	2.6	2.7	2.7	2.7
	K	2.0	1.8	1.9	1.9
	H	2.4	2.4	2.4	2.4
	R	1.9	1.9	2.0	2.0
	Y	4.1	4.1	3.9	4.0
	total	15.1	15.0	15.0	15.0

Abbreviations are the same as in Table 2 and text.

Table 5 Lengths and sequences of unassigned DNA in *Branchiostoma species*

Distributed block			*Bb* Size (nt)	*Bb* Sequence	*Bl* Size (nt)	*Bl* Sequence	*Bf* Size (nt)	*Bf* Sequence
cytb	–	*tRNA-Thr*	1	G				
tRNA-Thr	–	*tRNA-Pro*						
tRNA-Pro	–	*12S rRNA*						
12S rRNA	–	*tRNA-Phe*						
tRNA-Phe	–	*tRNA-Val*	3	CTG	2	GA	2	GA
tRNA-Val	–	*16S rRNA*						
16S rRNA	–	*tRNA-Leu* (TTR)						
tRNA-Leu (TTR)	–	*NADH1*			3	GTG		
NADH1	–	*tRNA-Ile*						
tRNA-Ile	–	*tRAN-Met*						
tRAN-Met	–	*tRNA-Gln*						
tRNA-Gln	–	*NADH2*	1	T	1	T	1	T
NADH2	–	*tRNA-Asn*						
tRNA-Asn	–	*tRNA-Trp*						
tRNA-Trp	–	*tRAN-Ala*						
tRAN-Ala	–	*tRNA-Cys*						
tRNA-Cys	–	*tRNA-Tyr*						
tRNA-Tyr	–	*COI*	4	AATT	10	TGAAGAATTT	10	AGGATAGCTT
COI	–	*tRNA-Ser* (TCN)						
tRNA-Ser (TCN)	–	*tRNA-Asp*	8	TTTGTTTA	7	TAGATTA	8	TTTAAAYA
tRNA-Asp	–	*COII*	3	TCG				
COII	–	*tRNA-Lys*						
tRNA-Lys	–	*ATP8*						
ATP8	–	*ATP6*						
ATP6	–	*COIII*						
COIII	–	*NADH3*						
NADH3	–	*tRNA-Arg*						
tRNA-Arg	–	*NADH4L*	3	GTT	2	GT	2	GT
NADH4L	–	*NADH4*	1	A				
NADH4	–	*tRNA-His*						
tRNA-His	–	*tRNA-Ser* (AGY)						
tRNA-Ser (AGY)	–	*tRNA-Leu* (CTN)						
tRNA-Leu (CTN)	–	*NADH5*						
NADH5	–	*tRNA-Gly*	110	NC sequence	190	NC sequence	129	NC sequence
tRNA-Gly	–	*NADH6*						
NADH6	–	*tRNA-Glu*	1	T	1	A	1	A
tRNA-Glu	–	*cytb*	1	T	1	T	1	T

Abbreviations are the same as in Table 2.

single region between the *NADH5* and $tRNA^{G}$, 10 nt between the $tRNA^{Y}$ and *COI*, 7 nt between the $tRNA^{S(TCN)}$ and $tRNA^{D}$, 3 nt between the $tRNA^{L(TTR)}$ and *NADH1*, 2 nt each between the $tRNA^{F}$ and $tRNA^{V}$ and between the $tRNA^{R}$ and *NADH4L*, and 1 nt each in 3 other regions. As for *B. floridae*, 129 nt were found in a single region between the *NADH5* and $tRNA^{G}$, 10 nt between the $tRNA^{Y}$ and *COI*, 8 nt between the $tRNA^{S(TCN)}$ and $tRNA^{D}$, 2 nt each between the $tRNA^{F}$ and $tRNA^{V}$ and between the $tRNA^{R}$ and *NADH4L*, and 1 nt each in 3 other regions (Table 5).

5. Homology

A homology of protein-coding genes revealed a 94-100% homology within each species of *B. belcheri*, *B. lanceolatum*, and *B. floridae* and a 55-96% inter-species homology. Among protein-coding genes, the *COI*, *COII*, and *COIII* exhibited high degrees of homology (≧90%) both within and inter-species. The ATP8 exhibited a high intra-species homology (94-100%) but the lowest inter-species homology (55-68%) (Table 6).

Molecular phylogenic trees created using the NJplot were similar among protein-coding genes, tRNA genes, and rRNA genes. *B. belcheri*, *B. lanceolatum* and *B. floridae* were classified into separate clusters, respectively (Fig. 1).

4. Discussion

1. Base frequency

Base frequency exhibited no significant intra-species differences but did show significant inter-species differences ($P<0.001$)[16]. A comparison of purine and pyrimidine bases revealed that the latter were predominant, with no intra-species differences, and a maximum of only 1.1% inter-species difference. GC content exhibited no intra-species differences but did show a maximum 3.0% inter-species difference. Differences in basic structural factors, such as GC content and base frequency, are thought to be attributable to differences in mutation pressure[17]. Possession of sufficient levels of these basic structural factors is a requirement for the production of a functional protein[18]. The differences observed in such factors thus suggest that *Branchiostoma* species have long been under selective pressure due to differences in habitat.

2. Protein-coding genes

The sizes and sequences of the protein-coding genes of *Branchiostoma* species were very similar to those of other metazoans. While the numbers of amino acids were the same within each species, the mtDNA of *B. lanceolatum* was 1 aa shorter than those of *B. belcheri* and *B. floridae*. This inter-species difference in the numbers of amino acids was attributable to the *NADH1* of *B. lanceolatum* being 1 aa shorter than that of the other two species. This suggests that, if a reduction in the size of mtDNA is required for a rapid replication of mitochondria, *B. lanceolatum* is under stronger selective pressure to reduce its size than are the other *Branchiostoma* species.

The base frequency exhibited almost no intra-species variation and no significant differences, while the frequencies of most amino acids differed significantly inter-species ($P<0.001$). Inter-species differences were also seen in the distributions of hydrophobic and hydrophilic amino acids, as shown in Table 4. These findings indicate that *B. belcheri*, *B. lanceolatum*, and *B. floridae* differ distinctly from one anather.

The *COI*, *COII*, and *COIII* were best conserved inter-species, with 98-100% intra-species homology

Table 6 Percentage of identical base and amino acid sequences for protein-coding genes and rRNA genes in *Branchiostoma species*

ATP6	*Bb*	*Bl*	*Bf*
Bb	98-100	85-86	84-87
Bl		99-100	86-88
Bf			97-100

cytb	*Bb*	*Bl*	*Bf*
Bb	99-100	89-90	88-89
Bl		99-100	92-93
Bf			98-100

NADH4L	*Bb*	*Bl*	*Bf*
Bb	100	82	84-85
Bl		100	83-84
Bf			98-100

ATP8	*Bb*	*Bl*	*Bf*
Bb	94-100	66-68	55-59
Bl		100	61-62
Bf			96-100

NADH1	*Bb*	*Bl*	*Bf*
Bb	99-100	85-87	88-89
Bl		99-100	87-88
Bf			99-100

NADH5	*Bb*	*Bl*	*Bf*
Bb	99-100	77-78	74
Bl		99-100	80
Bf			98-100

COI	*Bb*	*Bl*	*Bf*
Bb	99-100	96-97	95-97
Bl		99-100	97-98
Bf			98-100

NADH2	*Bb*	*Bl*	*Bf*
Bb	98-100	73-74	67-68
Bl		99-100	72-73
Bf			97-100

NADH6	*Bb*	*Bl*	*Bf*
Bb	97-100	72-74	67-70
Bl		98-100	71-73
Bf			97-100

COII	*Bb*	*Bl*	*Bf*
Bb	99-100	93	93-94
Bl		99-100	95-96
Bf			99-100

NADH3	*Bb*	*Bl*	*Bf*
Bb	99-100	76-77	82-84
Bl		99-100	79-81
Bf			98-100

12S rRNA	*Bb*	*Bl*	*Bf*
Bb	99-100	78-79	74-76
Bl		98-100	77-78
Bf			98-100

COIII	*Bb*	*Bl*	*Bf*-L
Bb	99-100	95	90-91
Bl		99-100	93-94
Bf			99-100

NADH4	*Bb*	*Bl*	*Bf*
Bb	98-100	66-67	72-74
Bl		99-100	69-70
Bf			98-100

16S rRNA	*Bb*	*Bl*	*Bf*
Bb	99	79-80	77-78
Bl		99-100	75-76
Bf			98-99

Abbreviations are the same as in Table 2.

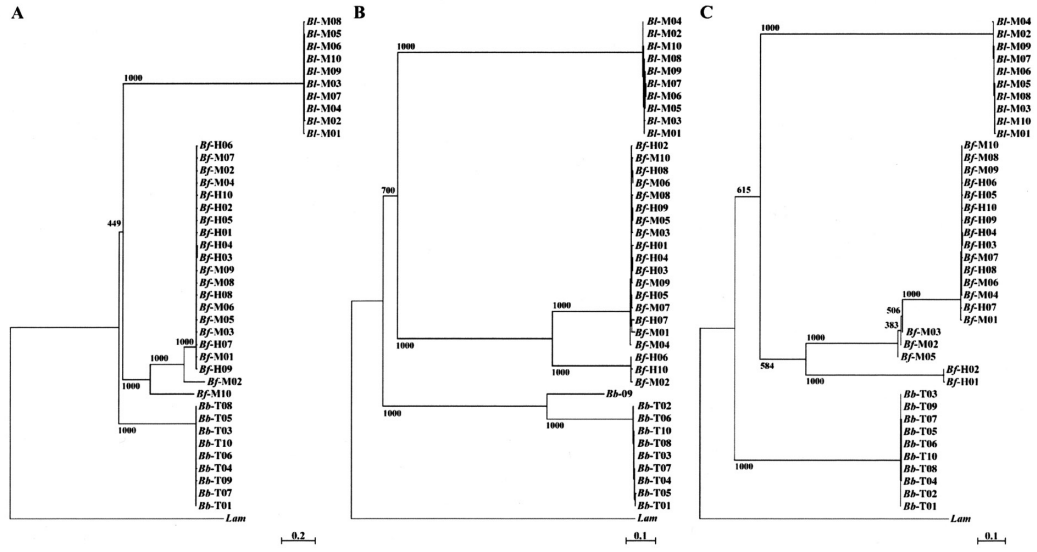

Fig. 1 Neighbor-joining analysis of the relationships between representative mitochondrial genes

An analysis based on 41 genotypes was conducted using CLUSTAL-W program (DDBJ). Bootstrap testing was performed (1,000 replicas). Numbers at forks represent bootstrap percentages. Tree is unrooted.

A) Amino acid sequence analysis of 13 protein-coding genes; B) Base sequence analysis of tRNA genes; C) Base sequence analysis of rRNA genes.

Abbreviations are same as in Table 2.

Lam (Lampetra fluviatilis: Y18683) was included as an outgroup.

and 90-98% inter-species homology. The *NADH2*, *NADH4*, and *NADH6* exhibited a 97-100% intra-species homology, but a low inter-species homology of 66-74%. The *ATP8* also exhibited an intra-species homology of 94-100%, but a very low inter-species homology of 55-68%. These genes thus appear to be more characteristic of inter-species differences than other protein-coding genes. The *ATP8*, which exhibited the greatest variation among *Branchiostoma* species, may thus be a useful marker for distinguishing habitat differences.

3. Initiation and termination codons

The initiation codons for the *NADH1* and *NADH4L* of *B. belcheri* appear to be GTG, since no ATG or other initiation codons in metazoan mitochondria are adjacent to these genes. The apparent initiation codons for the *COI* and *NADH1* of *B. floridae* are GTG and ATA, respectively, while that for the *COI* of *B. lanceolatum* is GTG, since, as in the case of *B. belcheri*, no ATG or other initiation codons in metazoan mitochondria are adjacent to these genes, and the codons in the initiation codon region are very similar to the amino terminal sequence of the *COI* in other metazoans. A GTG codon is thought to initiate the transcription of the *COI* or other genes in mtDNAs of many metazoans[19]. Any estimation of the initiation codon for the *NADH1* is somewhat unreliable since GTC and ATA, which are frequently-used initiation codons, are located in the estimated initiation codon region of the *NADH1*. The GTG codon is directly adjacent to the $tRNA^{L(TTR)}$ located upstream and is also next to the subsequent ATA codon. The initiation codon region of the *NADH1* of *B. lanceolatum* is also poorly conserved. Although Delarbre *et al.*[20] have determined that ATA is the initiation codon for the *NADH1,* the assumption of ATA as the initiation codon for the *NADH1* of *B. floridae* examined in this study resulted in some initiation codons being GTA, a finding that is inconsistent with the currently-

accepted initiation codon for that *Branchiostoma* . We thus considered it more reasonable to select GTG as the initiation codon, as it is located before the variable region and is therefore well conserved. Furthermore, no ATA but only GTG was found in the initiation codon region of the *NADH1* of *B. belcheri*. We thus determined GTG, which was directly adjacent to the $tRNA^{L(TTR)}$, to be the initiation codon. Although the initiation codon of *B. lanceolatum* may be estimated to be GTG, based on such factors as the nucleotide sequences and size reduction of other *Branchiostoma* species, ATG is adopted as the initiation codon for most protein-coding genes of *B. lanceolatum*. In the *NADH1* of *B. lanceolatum*, GTG is directly followed by ATG, and there currently appears to be no valid reason for replacing ATG with GTC. Consequently, we determined the initiation codon of the *NADH1* to be GTG in *B. belcheri*, and *B. floridae* and ATG in *B. lanceolatum*. The initiation codon for the *COI* was ATG in *B. belcheri* and *B. lanceolatum*, and GTG in *B. floridae*, while that for the *NADH4L* was GTG in *B. belcheri* and *B. lanceolatum*, and ATG in *B. floridae*.

A complete termination codon of TAG is commonly found in the *COI*, *NADH2*, and *ATP8*, among which the *COI* and *NADH2* do not overlap with downstream genes having the same transcription orientation. The TAG codon in the *ATP8* is followed by a well-conserved sequence for amino acids (WPW) at the 3'-terminal end of the *ATP8*, at which the *ATP8* and *ATP6* overlap. These genes, which exhibit an overlap in the mtDNA of the chordates, are translated via two cistronic mRNAs in the overlap region[21]. Incomplete termination codons of *Branchiostoma* can form complete termination codons if their transcriptional products overlap with downstream genes by as little as 1-2 nt. Delarbre *et al.*, who determined the sequence of a cDNA corresponding to the mRNA of the *NADH1* of *B. lanceolatum*, found that it ended with TAA[20], supporting the hypothesis that the *NADH1* has an incomplete codon. A comparison of termination codons between species revealed characteristic findings. The *cytb* of *B. belcheri* was found to have an incomplete termination codon, while that of *B. lanceolatum* and *B. floridae* was found to have a complete termination codon. The *NADH1*, *NADH4L*, and *NADH4* of *B. belcheri also* have a complete one, while those of *B. lanceolatum* and *B. floridae* are incomplete. The *NADH5* of *B. floridae* has a termination codon of TAG, while that of *B. belcheri* and *B. lanceolatum* has a termination codon of TAA. While the *NADH6* of *B. lanceolatum* has an incomplete termination codon, that of the other species is complete. A complete termination codon of TAG was found in the *COI*, *NADH2*, and *ATP8* of all *Branchiostoma* species.

These findings suggest that the initiation and termination codons vary among *Branchiostoma* species, indicating that *B. belcheri*, *B. lanceolatum*, and *B. floridae* are thus distinct from one another.

4. Unassigned DNA

Unassigned DNAs were found in 11 regions including non-coding (NC) sequences in *B. belcheri*, 9 in *B. lanceolatum*, and 8 in *B. floridae*, with sizes of 137, 217, and 154 nt, respectively. The sizes of unassigned DNAs, not including NC sequences in the corresponding species, were 27, 27, and 25 nt, with none significantly different. The difference in size of NC sequences is therefore reflected by a corresponding difference in size of unassigned DNAs. One characteristic finding about unassigned DNAs is that their insertion sites varied among *B. belcheri*, *B. lanceolatum*, and *B. floridae*. If a GTG codon was used as the initiation codon for the *NADH1*, unassigned DNAs would be inserted into the same 8 regions in *B. lanceolatum* and *B. floridae*. The only difference between the two species is that one has unassigned DNAs between the $tRNA^{Y}$ and *COI*, while the other has them between the $tRNA^{S(TCN)}$ and $tRNA^{D}$. *B. belcheri*, *B. lanceolatum*, and *B. floridae* differed with respect to their insertion sites, sequences, and sizes of their unassigned DNAs located in three regions. The difference in the makeup of unassigned DNAs appears to indicate an inter-species difference. *B. lanceolatum* and *B. floridae* exhibited differences in unassigned DNAs in two regions, excluding NC sequences. These unassigned DNAs are located between the $tRNA^{Y}$ and *COI*, and between the $tRNA^{S(TCN)}$ and $tRNA^{D}$. In particular, unassigned DNAs

in the latter region exhibited inter-species differences not only in sequence but also in size, while also exhibiting intra-species differences in *B. floridae*. These findings suggest that the difference in the composition of unassigned DNAs indicates an inter-species distinction.

The only NC sequence found in the mtDNA of *Branchiostoma* was located between the *NADH5* and *tRNA^G*, with a size of 110 nt in *B. belcheri*, 190-192 nt in *B. lanceolatum*, and 68 nt in *B. floridae*. This region is remarkably extended in *B. lanceolatum* compared to the other two species. The GC content in this region is 43% in *B. belcheri*, 44% in *B. lanceolatum*, and 40% in *B. floridae*, and is slightly higher than the GC content in the entire mtDNA (35, 38, and 37%, respectively). The longest NC sequence in the mtDNA of *Branchiostoma* is 191 nt in length, compared with 198 nt for *Petromyzon marinus*[22], 928 nt for *Cyprinus carpio*[23], and 1,183 nt for *Protopterus dollof*[24]. Upon comparing the NC sequences of various vertebrates, we found no apparent similarity. The degree of similarity in the NC sequence among the *Branchiostoma* species was also very low. The NC sequence in *Branchiostoma* is totally unassigned, suggesting that it may be a control region. However, the sequence in this region includes no signals currently known to be involved in mtDNA replication, such as a conserved sequence block (CSB)[25] or a termination-associated sequence (TAS)[26]. A repetitive sequence of 9 nt (TTTTTTGGG) can be found in *B. lanceolatum* and *B. floridae* as the only remarkable characteristic of this region, while it is not found in *B. belcheri*. Although possession of a relatively short genome appears to be an important factor enabling its rapid replication, there is no indication that *Branchiostoma* requires a quicker replication of mtDNA than do other organisms. However, the fact that the sizes of all tRNA genes, as well as two rRNA genes, are reduced in *Branchiostoma* suggests that selective pressure is exerted towards a reduction in the size of its mtDNA. The muscle fiber function in *Branchiostoma* relies mainly on fast muscles, since the movements required for propulsion, e. g., when avoiding danger or capturing food, are mainly produced by the rapid contraction of fast muscles, as are the movements of flatfish; moreover the muscle fibers of *Branchiostoma* develop a level of ATPase activity several times that in the muscle fibers of mammals within a short period of time[27]. It is thus clear that *Branchiostoma* species also require mtDNA replication. These characteristics of mammals are not apparent in *Branchiostoma*; they may show different characteristics or no such characteristics at all. Although it is possible that the NC sequence is involved in the initiation of mtDNA replication, the precise mechanism by which the replication of *Branchiostoma* mtDNA is initiated is unknown. The lack of control regions may be a characteristic of *Branchiostoma* species, although it is not known whether this lack is related to the reduction in genome size. The present findings still suggest that structural specificity has been determined either by the need to reduce genome size or by the necessity to form control regions of a specific type.

5. Homology

Molecular phylogenetic trees for protein-coding genes, tRNA genes, and rRNA genes have revealed that *B. belcheri*, *B. lanceolatum*, and *B. floridae* examined in the present study are phylogenetically disparate, with differences in the times of divergence among the three species. That time of divergence between *B. lanceolatum* and *B. floridae* (200 million years ago) was similar to the one between *B. belcheri* from the Pacific and *B. lanceolatum* or *B. floridae* from the Atlantic (approximately 180 million years ago). The time of divergence between *B. floridae* and *B. lanceolatum* was almost the same as that reported by Cañestro *et al.* (190 million years ago)[28]. These findings indicate that the three *Branchiostoma* species diverged at almost the same time but subsequently underwent evolution under different habitats.

5. Conclusion

Having examined the characteristics of different *Branchiostoma* species, we established that *B. belcheri*, *B. lanceolatum*, and *B. floridae* are distinct species.

In addition, we provided a reason for correcting the previously-reported initiation codon for the

NADH1 of *B. floridae* to GTG. The analytical result of present mtDNA involved a new comparative genome analysis of those above three species. *Branchiostoma* have been used in a number of studies, and their molecular biological findings are frequently cited. The findings we obtained should prove very useful for future studies in the fields of phylogenetics, evolutionary biology, and phylogenetic systematics.

6. Acknowledgments

We would like to express our deep gratitude to Dr. Michael Schubert and Prof. Vincent Laudet from the Institut de Génomique Fonctionnelle de Lyon (IGFL), and to Dr. Linda Holland and Prof. Nicholas Holland from the Scripps Institution of Oceanography, University of California-San Diego, for their cooperation in collecting the *Branchiostoma* samples used in this study.

References

1) Yasui K and Kubokawa K: Anatomy. Biology of Cephalochordate Lancelet. 30-83, University of Tokyo Press, Tokyo, (2005)
2) Pallas PS: Limax lanceolatus. Descriptio Limacis lanseolaris. Spicilegia Zoologica, quibus novae imprimus et obscurae animalium species iconibus, descriptionibus. 10. 19, Gottl. August. Lange., Berline, (1774)
3) Cameron CB, Garey JR and Swalla BJ: Evolution of the chordate body plan: New insights from phylogenetic analyses of deuterostome phyla. Proc. Natl. Acad. Sci. USA., 97: 4469-4474, 2000
4) Winchell CJ, Sullivan J, Cameron CB, Swalla BJ and Mallatt J: Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. Mol. Biol. Evol., 19: 762-776, 2002
5) Nicholas HP, Thomas B, David EKF, et al: The amphioxus genome and the evolution of the chordate karyotype. Nature, 453: 1064-1071, 2008
6) Takada MY, Mukaida M and Imai T: Molecular evolution of mitochondrial DNA of *Branchiostoma belicheri*. J. Anal. Bio-Sci., 27: 291-302, 2004
7) Spruyt N, Delarbe C, Gachelin G and Laudet V: Complete sequence of the amphioxus (*Branchiostoma lanceolatum*) mitochondrial genome: relations to vertebrates. Nucleic Acids Res., 26: 3279-3285, 1998
8) Boore JL, Daehler LL and Brown WM: Complete sequence, gene arrangement, and genetic code of mitochondrial DNA of the cephalochordate *Branchiostoma floridae* (Amphioxus). Mol. Biol. Evol., 16: 410-418, 1999
9) Nohara M, Nishida M and Nishikawa T: New complete mitochondrial DNA sequence of the lancelet *Branchiostoma lanceolatum* (Cephalochordata) and the identity of this species' sequences. Zool. Sci., 22: 671-674, 2005
10) Lowe TM and Eddy SR: tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res., 25: 955-964, 1997
11) Thompson JD, Higgins DG and Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22: 4673-4680, 1994
12) Kimura M: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol., 16: 111-120, 1980
13) Saitou N and Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol., 4: 406-425, 1987
14) Perriére G and Gouy M: WWW-Query: An on-line retrieval system for biological sequence banks. Biochimie, 78: 364-369, 1996
15) Delarbre C, Escriva H, Gallut,C, Barriel V, Kourilsky P, Janvier P, Laudet V and Gachelin G: The complete nucleotide sequence of the mitochondrial DNA of the agnathan *Lampetra fluviatilis*: bearings on the phylogeny of cyclostomes. Mol. Biol. Evol., 17: 519-529, 2000
16) Student: The probable error of a mean, Biometrika, 6: 1-25, 1908
17) Sueoka N: Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA., 85: 2653-2657, 1988
18) Ikehara K: Origins of gene, genetic code, protein and life (Comprehensive view of life systems from the GNC-SNS primitive genetic code hypothesis). Viva Origino, 29: 66-85, 2001
19) Wolstenholme DR: Animal mitochondrial DNA structure and evolution. Int. Rev. Cytol., 141: 173-216, 1992
20) Delarbre C, Barriel V, Tillier S, Janvier P and Gachelin G: The main features of the craniate mitochondrial DNA between the ND1 and the COI genes were established in the common ancestor with the lancelet. Mol. Bio. Evol., 14: 807-813, 1997

21) Fearnley IM and Walker JE: Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. EMBO J., 5: 2003-2008, 1986
22) Lee WJ, and Kocher TD: Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. Genetics, 139: 873-887, 1995
23) Chang YS, Huang FL and Lo TB: The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. J. Mol. Evol., 38: 138-155, 1994
24) Zardoya R and Meyer A: The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolloi*) supports its phylogenetic position as a close relative of land vertebrates. Genetics, 142: 1249-1263, 1996
25) Doda JN, Wright CT and Clayton DA: Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA., 78: 6116-6120, 1981
26) Chang DD and Clayton DA: Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Natl. Acad. Sci. USA., 82: 351-355, 1985
27) Takada Y, Mukaida M and Imai T: A histochemical evaluation of myofibrillar ATPase activity in the Branchiostoma somatic muscle. J. Anal. Bio-Sci., 31: 147-154, 2008
28) Cañestro C, Albalat R, Hjelmqvist L, Godoy L, Jörnvall H and González-Duarte R: Ascidian and amphioxus Adh genes correlate functional and molecular features of the ADH family expansion during vertebrate evolution. J. Mol. Evol., 54: 81-89, 2002