Analysis of ITS1 and ITS2 sequences in *Ensis* razor shells: suitability as molecular markers at the population and species levels, and evolution of these ribosomal DNA spacers[i].

Joaquín Vierna, Andrés Martínez-Lage, Ana M. González-Tizón[ii]

Department of Molecular and Cell Biology, Evolutionary Biology Group (GIBE), Universidade da Coruña, A Zapateira s/n, E-15071 La Coruña, Spain.

Genome, Volume 53**,** Issue 1, pages 23-34, January 2010
Received August 10, 2009, Accepted October 22, 2009, First published 15 December 2009.

How to cite:

J. Vierna, A. Martínez-Lage, A. M. González-Tizón (2010). Analysis of ITS1 and ITS2 sequences in *Ensis* razor shells: suitability as molecular markers at the population and species levels, and evolution of these ribosomal DNA spacers. *Genome* **53**(1): 23-34. https://doi.org/10.1139/G09-080

Abstract

Internal transcribed spacer 1 and 2 (ITS1 and ITS2) sequences were analysed in *Ensis* razor shells (Mollusca: Bivalvia: Pharidae). We aimed to (1) test ITS1 and ITS2 as molecular markers at the population level in the successful alien *E. directus* (Conrad, 1843); (2) test these spacers at the species level in *E. directus* and three other *Ensis* species, *E. siliqua* (L., 1758), *E. macha* (Molina, 1782), and *E. magnus* (Schumacher, 1817); and (3) analyse the evolutionary processes that may be shaping *Ensis* ITS1 and ITS2 extant variation. In *E. directus*, despite the intragenomic divergence detected, ITS1 and ITS2 were informative in differentiating the geographic areas considered (Denmark and Canada) by means of both the insertion-deletion polymorphism and the nucleotide polymorphism. In this species, the 5.8S ribosomal gene (5.8S) showed scarce polymorphism. At the species level, maximum parsimony and maximum likelihood analyses revealed that ITS1 and ITS2 may be suitable to reconstruct *Ensis* phylogenetic relationships. Finally, the evolutionary models that best fit the long-term evolution of *Ensis* ITS1–5.8S–ITS2 are discussed. A mixed process of concerted evolution, birth-and-death evolution, and selection is chosen as an option that may reconcile the long-term evolution of *Ensis* ITS1–5.8S–ITS2 and 5S ribosomal DNA.

Keywords: internal transcribed spacers; rDNA; concerted evolution; birth-and-death evolution; mixed process; indel polymorphism

Introduction

The evolution of nuclear ribosomal DNA (nrDNA) is traditionally explained by concerted evolution. Homogenizing mechanisms such as unequal crossovers and gene conversions eventually maintain the sequence similarity among all repeats of the array, and even among repeats belonging to different arrays (for a review see Eickbush and Eickbush 2007).

Nevertheless, it has recently been demostrated that birth-and-death processes may be responsible, at least in part, for the evolution of nrDNA in some organisms, including *Ensis* razor shells (Mollusca: Bivalvia: Pharidae) (Rooney 2004; Rooney and Ward 2005; Fujiwara *et al.* 2009; Vierna *et al.* 2009). Under a birth-and-death model, gene duplications generate new genes that can persist in the genome for long periods,

degenerate into pseudogenes, or become deleted. In a phylogenetic analysis of genes from several closely related taxa, sequences will show a between-species clustering pattern, in contrast to the within-species clustering pattern that is expected under concerted evolution. However, there are two exceptions to this rule: (1) recent gene duplicates and (2) multigene families that experience rapid gene turnover due to birth-and-death evolution (e.g., Zhang *et al.* 2000). In the latter case, rapid gene turnover can lead to the creation of species-specific gene clusters as a result of frequent gene duplication and loss (Rooney 2004).

Natural selection is expected to play a role in the long-term evolution of nuclear ribosomal genes, under either a concerted or birth-and-death evolution scenario, reducing sequence variation in the coding regions.

Under a strict concerted evolution scenario, intragenomic divergence remains very low, and all copies of a nrDNA gene or spacer from a single genome can be considered a single locus in phylogenetic or molecular population studies. Under birth-and-death evolution, substantially different copies may exist in the genome (except in the two cases stated above), therefore increasing the levels of intragenomic divergence. Since intragenomic variants produced by birth-and-death processes are paralogous, researchers must be very careful when using nrDNA genes or spacers as molecular markers at any taxonomic level. Comparisons between paralogous sequences may completely skew the results of the survey.

Even though concerted evolution functions extremely well on all nrDNA genes and spacers in the vast majority of organisms (Eickbush and Eickbush 2007), some authors have recommended analysing the intragenomic divergence levels before using the internal transcribed spacers of the major ribosomal genes (ITS1 and ITS2) as molecular markers (Leo and Barker 2002; Wörheide *et al.* 2004; Alvarez *et al.* 2007; Sword *et al.* 2007). Interestingly, according to Harris and Crandall (2000), in all cases reported so far where intragenomic divergence was looked for within ITS1 and ITS2, it was found to some degree.

In the present work we analysed sequence variation within ITS1 and ITS2 (hereafter, ITS) in *Ensis* razor shells. We studied their suitability as molecular markers, and the evolutionary processes that may be responsible for the long-term evolution of *Ensis* ITS1–5.8S–ITS2. We considered four *Ensis* species: *E. directus* (Conrad, 1843) (syn. *E. americanus* (Gould, 1870)), which is native to the Atlantic coast of North America but was introduced in European waters at the end of the 1970s (Cosel *et al.* 1982); *E. siliqua* (L., 1758) and *E. magnus* (Schumacher, 1817) (syn. *E. arcuatus* var. *ensoides* Van Urk, 1964 and *E. arcuatus* var. *norvegica* Van Urk, 1964), which are native to Atlantic Europe; and *E. macha* (Molina, 1782), which occurs along the southern coasts of Chile and Argentina. Taking into account the levels of intragenomic divergence detected, ITS were studied as molecular markers at the population level in *E. directus* and at the species level in all four species.

Material and methods

Ensis directus individuals were sampled at three Danish localities (Sillerslev, Sundsøre, and Juvre Deep) and one Canadian locality (Long Pond, Conception Bay, Newfoundland). All individuals were identified as *E. directus* according to morphological characters including animal size, shape and length of muscle scars, and shell colour and curvature. In the present article, all taxon names follow Cosel (2009) (when applicable). Animals were stored in 100% ethanol until DNA extraction.

Extraction of genomic DNA was performed from foot tissue using the NucleoSpin Tissue kit (Macherey-Nagel GmbH and Co. KG). For PCR we used a pair of primers that anneal at the 3′ end of the 18S ribosomal gene and the 5′ end of the 28S ribosomal gene (Heath *et al.* 1995) that were previously used on several bivalve species (Fernández *et al.* 2001; Insua *et al.* 2003; Toro *et al.* 2004; Mahidol *et al.* 2007). ITS1, the 5.8S ribosomal gene (hereafter, 5.8S), and ITS2 were amplified. Each reaction was performed in a total volume of 25 µL containing ~25 ng of genomic DNA, 0.625 U of *Taq* DNA polymerase (Roche Diagnostics), 5 nmol of each dNTP (Roche Diagnostics), 20 pmol of each primer, and the buffer recommended by the polymerase supplier. PCR conditions were as follows: an initial denaturation step at

94 °C for 3 min followed by 35 cycles of denaturation at 94 °C for 20 s, annealing at 54 °C for 20 s, and extension at 72 °C for 45 s, and a final extension at 72 °C for 5 min. PCR products were run on agarose gels, stained with ethidium bromide, and imaged under UV light. A single band of about 1000 bp was obtained from all individuals analysed, and it was cloned using the TOPO TA Cloning Kit (Invitrogen). Several transformant colonies (n = 2–10) from each individual were selected at random and grown in LB medium. A QIAprep Spin Miniprep Kit (QIAGEN) was used to purify the plasmids, which were sequenced using both M13 Forward and M13 Reverse primers (included in the cloning kit) in a capillary DNA sequencer (CEQ 8000 Genetic Analysis System, Beckman Coulter). Electropherograms were examined and assembled in BioEdit 7.0.9.0 (Hall 1999). A sequence similarity search was performed in BLAST (http://blast.ncbi.nlm.nih.gov) to determine the similarities of the sequences obtained to other ITS1–5.8S–ITS2 sequences from DDBJ/EMBL/GenBank. BLAST confirmed that all sequences obtained were similar to other mollusc ITS1–5.8S–ITS2 sequences, and permitted identification of the flanking regions of each ITS. Sequences were deposited in the DDBJ/EMBL/GenBank databases under the following accession numbers: FN391027–FN391081 for ITS1 and FN391082–FN391136 for ITS2.

To examine the potential of both ITS as phylogenetic markers within genus *Ensis*, we downloaded sequences from *E. siliqua* (accession Nos. AJ966667–AJ966681), *E. macha* (AM933624–AM933631), and *E. magnus* (AJ966682–AJ966697) available in DDBJ/EMBL/GenBank. No 5.8S sequences from these species were available.

Sequence alignments were performed in ClustalX 2.08 (Larkin *et al.* 2007) (penalties for gap opening = 7 and gap extension = 3) and manually adjusted for local optimization in MEGA 4.0.2 (Tamura *et al.* 2007). To use as many nucleotide positions as possible in the subsequent analyses, we performed three different alignments for each ITS: one with all *E. directus* sequences, one with *E. directus* and *E. macha* sequences, and one with all sequences from the four *Ensis* species. In this case, as some regions did not align properly, the alignment was corrected in the Gblocks Server (http://molevol.cmima.csic.es/castresana/Gblocks_server.html) (Castresana 2000) to increase the probability of considering only homologous positions in the analyses. All alignments are available as supplementary material (Figs. S1–S6).

Sequence lengths and mean nucleotide compositions were obtained from MEGA 4.0.2 (Tamura *et al.* 2007). Using DnaSP 4.9 (Rozas *et al.* 2003), we studied the insertiondeletion (indel) polymorphism of each marker. To analyse the potential of indel polymorphism to differentiate sequences from different sampling localities and geographic areas, a nexus file with the indel polymorphism of both ITS was generated in DnaSP 4.9 (Rozas *et al.* 2003) and run in Paup*4.0b10 (Swofford 2002) under maximum parsimony (MP) (settings were those used in the phylogenetic analysis of *E. siliqua*, *E. macha*, *E. directus*, and *E. magnus* ITS; see below). The number of "sequencetypes" and the haplotype diversities were also calculated in DnaSP 4.9 (Rozas *et al.* 2003), excluding nucleotide positions with gaps. In this study we use the word "sequencetype" to denote a single and unique type of ITS, following Wörheide *et al.* (2002). Mean sequence divergences within and between groups and the maximum divergence between sequences were obtained by means of the *p*-distance using MEGA 4.0.2 (Tamura *et al.* 2007). To study a possible clustering of *E. directus* ITS sequences according to sampling localities or geographic areas, maximum likelihood (ML) phylogenetic trees were reconstructed using the PhyML 3.0 Web server (http://www.atgc-montpellier.fr/phyml/) (Guindon and Gascuel 2003) for each ITS and for a concatenated data set of both ITS. The best-fit model of nucleotide substitution for each data set was selected by statistical comparison of 88 different models using jModelTest 0.1.1 (Posada 2008) and applying the Akaike information criterion. Starting trees were obtained using the BIONJ algorithm (Gascuel 1997) and gaps were treated as unknown characters. Tree topologies were estimated by a nearest neighbor interchange (Jarvis *et al.* 1983 and references therein). Node support was estimated by the bootstrap test (Felsenstein 1985) (100 replicates).

Ensis phylogenetic trees were reconstructed using Gblocks alignments (Figs. S3–S4) and two different phylogenetic methods, ML and MP. To save computational time, duplicate sequences were deleted. ML

analyses were performed as for *E. directus* sequences (see above). For MP analyses, a heuristic search was conducted in Paup*4.0b10 (Swofford 2002) using the tree-bisection-reconnection (TBR) branch-swapping algorithm, and assuming equal weights and unordered character states for all characters. Gaps were treated as a fifth character state or as missing information. Starting trees were obtained via stepwise addition with random addition of sequences (10 replicates). The number of trees held at each step during stepwise addition was one. In both ML and MP analyses, node support was estimated by the bootstrap test (Felsenstein 1985) with 1000 replicates.

All MP and ML phylogenetic trees were edited in FigTree 1.2.2 (Andrew Rambaut; http://tree.bio.ed.ac.uk/software/figtree/).

Results

Ensis directus ITS

A total of 55 ITS1–5.8S–ITS2 sequences were obtained experimentally from *E. directus* (the number of clones per individual and the number of individuals per geographic area or locality are summarized in Table 1). Average GC contents were 58.9% for ITS1, 63% for ITS2, and 57.4% for the 5.8S. The 5.8S displayed no indels and only 4 point mutations (positions 17A>G, 80A>G, 101T>C, and 135A>G).

The length of ITS1 ranged between 484 and 510 bp, whereas the length of ITS2 was less variable (295–299 bp). The 5.8S was 157 bp in all clones. ITS1 length from Danish sequences ranged between 497 and 510 bp, and for ITS2 the length was 295–299 bp. In Canadian sequences, ITS length variation was low: 484–486 bp for ITS1 and 296 bp (no length variation) for ITS2. Intragenomic length variation in ITS1 was present in 4 individuals from Denmark (Sillerslev 3, Sundsøre 53, Sundsøre 58, and Sundsøre 67) and 2 Canadian individuals (Long Pond 13 and Long Pond 17). Three Danish animals displayed intragenomic length variation in ITS2 (Sundsøre 58, Juvre Deep 117, and Juvre Deep 119).

The *E. directus* ITS1 alignment (Fig. S1) contained 16 indel events, with an average indel event length of 2.625 bp. In contrast with ITS1, the ITS2 alignment (Fig. S2) contained only 3 indel events, with an average indel event length of 1.667 bp. The analysis of indel polymorphism in each geographic area showed that for ITS1, in Danish sequences there were 13 indel events, with an average indel event length of 1.692 bp, whereas in Canadian sequences there were only 3 indel events (average indel event length = 2 bp). In Danish ITS2 sequences, there were 2 indel events, with an average indel event length of 2 bp, whereas Canadian ITS2 sequences displayed no indel polymorphism. A thorough examination of *E. directus* ITS alignments showed that some indels were characteristic of one of the two geographic areas under study. For example, in ITS1, a TTG insertion in position 178 and a CGAGACGGCGTTAAC deletion in position 236 characterized individuals from Canada. In ITS2, a single nucleotide deletion in position 275 also characterized all Canadian individuals. The MP cladogram constructed on the basis of ITS indel polymorphism (Fig. 1) showed that sequences belonging to each geographic area grouped together with a bootstrap support of 88%, but no resolution according to sampling localities was obtained.

Table 1. Number of sequences sampled (n), number of sequencetypes obtained (s), and haplotype diversities (Hd) of Ensis directus internal transcribed spacers (ITS1 and ITS2).

			ITS1		ITS2	
		n	**s**	**Hd (mean ± SE)**	**s**	**Hd (mean ± SE)**
sp	*Ensis directus*	55	25	0.921±0.024	19	0.931±0.013
area	Denmark	33	19	0.939±0.026	14	0.907±0.028
area/loc	Canada / Long Pond	22	6	0.632±0.104	7	0.771±0.045
loc	Sillerslev	6	5	0.933±0.122	3	0.733±0.155
loc	Sundsøre	19	9	0.848±0.068	8	0.836±0.065
loc	Juvre Deep	8	6	0.893±0.111	6	0.893±0.111
ind	Long Pond 12	7	1	0	1	0
ind	Long Pond 13	6	2	0.333±0.215	2	0.333±0.215
ind	Long Pond 15	3	2	0.667±0.314	1	0
ind	Long Pond 16	3	2	0.667±0.314	2	0.667±0.314
ind	Long Pond 17	3	2	0.667±0.314	2	0.667±0.314
ind	Sillerslev 3	3	3	1.000±0.272	2	0.667±0.314
ind	Sillerslev 8	3	2	0.667±0.314	1	0
ind	Sundsøre 53	3	2	0.667±0.314	1	0
ind	Sundsøre 58	10	4	0.533±0.180	4	0.533±0.180
ind	Sundsøre 66	3	1	0	2	0.667±0.314
ind	Sundsøre 67	3	2	0.667±0.314	1	0
ind	Juvre Deep 103	3	1	0	1	0
ind	Juvre Deep 117	3	3	1.000±0.272	3	0.074±0.272
ind	Juvre Deep 119	2	2	1.000±0.500	2	0.074±0.272

Note: Sequences were grouped in several categories according to the species (sp), area, locality (loc), and individual (ind) they belong to.

The 55 ITS1 and ITS2 sequences sampled from *E. directus* produced 25 ITS1 sequencetypes and 19 ITS2 sequencetypes. In both markers, haplotype diversity was higher in Denmark than in Canada (Table 1). Of the 14 individuals analysed, only 3 had a single ITS1 sequencetype and only 6 had a single ITS2 sequencetype. There were no shared sequencetypes between Denmark and Canada.

Some authors prefer to use gap information when calculating sequence divergences, whereas others do not. Therefore, we show all mean divergences in Tables 2–3, though *p*-distance values did not substantially differ when positions with gaps were considered or excluded. Within groups, *p*-distance values did not decrease from species to individuals in either ITS. In fact, in some cases the mean divergence within individuals was of the same order of magnitude as the mean divergence within areas. For example, for ITS1, individual Sundsøre 67 showed a mean intragenomic divergence of 0.018 ± 0.005, while the value for all Danish sequences was 0.013 ± 0.003 (Table 2). ITS1 mean intragenomic divergence was particularly high in individuals Sundsøre 67 and Juvre Deep 119.

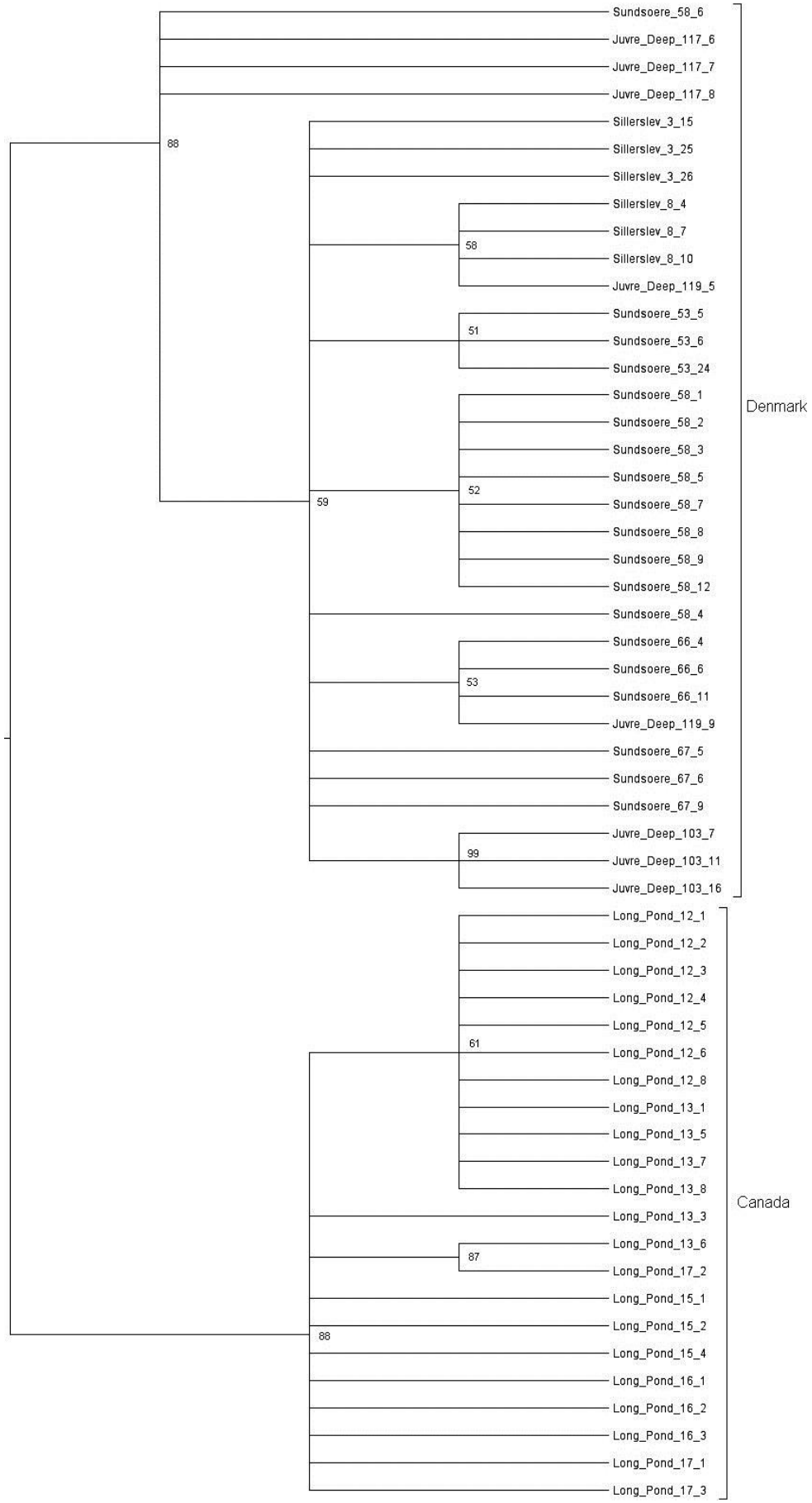

Figure 1. Maximum parsimony majority-rule consensus cladogram constructed on the basis of the insertion-deletion polymorphism of the internal transcribed spacers (ITS1 and ITS2) of *Ensis directus* (concatenated data set). Bootstrap values ≥ 50% are indicated at the nodes. Sampling localities, individuals, and clones are specified.

Table 2. Mean sequence divergence (*p*-distance, mean ±SE) within groups (species, areas, localities and individuals) for both internal transcribed spacers (ITS1 and ITS2) of *Ensis directus*.

			ITS1		ITS2	
		n	**Gaps considered**	**Gaps excluded**	**Gaps considered**	**Gaps excluded**
sp	*E. directus*	55	0.020±0.004	0.018±0.004	0.013±0.004	0.013±0.004
area	Denmark	33	0.014±0.003	0.013±0.003	0.014±0.004	0.013±0.004
area/loc	Canada / Long Pond	22	0.002±0.001	0.002±0.001	0.005±0.002	0.005±0.002
loc	Sillerslev	6	0.016±0.004	0.013±0.004	0.013±0.005	0.013±0.005
loc	Sundsøre	19	0.012±0.003	0.012±0.003	0.006±0.003	0.006±0.003
loc	Juvre Deep	8	0.008±0.003	0.009±0.002	0.018±0.006	0.017±0.005
ind	Long Pond 12	7	0	0	0	0
ind	Long Pond 13	6	0.002±0.001	0.002±0.001	0.001±0.001	0.001±0.001
ind	Long Pond 15	3	0.001±0.001	0.001±0.001	0	0
ind	Long Pond 16	3	0.004±0.002	0.004±0.002	0.002±0.002	0.002±0.002
ind	Long Pond 17	3	0.001±0.001	0.001±0.001	0.002±0.002	0.002±0.002
ind	Sillerslev 3	3	0.004±0.002	0.003±0.002	0.002±0.002	0.002±0.002
ind	Sillerslev 8	3	0.003±0.002	0.003±0.002	0	0
ind	Sundsøre 53	3	0.001±0.001	0.001±0.001	0	0
ind	Sundsøre 58	10	0.006±0.002	0.006±0.002	0.003±0.001	0.003±0.001
ind	Sundsøre 66	3	0	0	0.002±0.002	0.002±0.002
ind	Sundsøre 67	3	0.017±0.005	0.018±0.005	0	0
ind	Juvre Deep 103	3	0	0	0	0
ind	Juvre Deep 117	3	0.003±0.002	0.003±0.002	0.007±0	0.007±0
ind	Juvre Deep 119	2	0.012±0.005	0.012±0.005	0.023±0.008	0.024±0.009

Note: Mean sequence divergence was calculated with gaps considered and with gaps excluded; the later values are cited in the text.

Alignments revealed some divergent copies of ITS1 and ITS2 in the genome of *E. directus*. One ITS1 clone from individual Sundsøre 67 displayed 14 point mutations and 1 indel compared with the other clones of this individual. Similarly, one clone from Juvre Deep 119 displayed 5 point mutations and 3 indels. That clone from Sundsøre 67, however, did not show any variation in its ITS2 region, although variation would be expected because of the linkage of ITS1 and ITS2 sequences in the same clone. One ITS2 clone from individual Juvre Deep 119 displayed 7 point mutations and 1 indel, making its mean divergence quite high (0.024 ± 0.009). All ITS2 clones from individuals Juvre Deep 103 and Sillerslev 8 and one ITS2 clone from Juvre Deep 119 shared some point mutations between positions 269–276. This signal was strong enough to group all these sequences in one clade in the ITS phylogenetic tree (Fig. 2). In pairwise comparisons (matrix not shown), the maximum *p*distance obtained for ITS1 was 0.052 ± 0.010, between one clone from individual Sundsøre 67 and two clones belonging to individuals Sillerslev 8 and Long Pond 13. For ITS2, the maximum *p*-distance obtained was 0.034 ± 0.011, between one clone from Sundsøre 58 and one clone from Sillerslev 8.

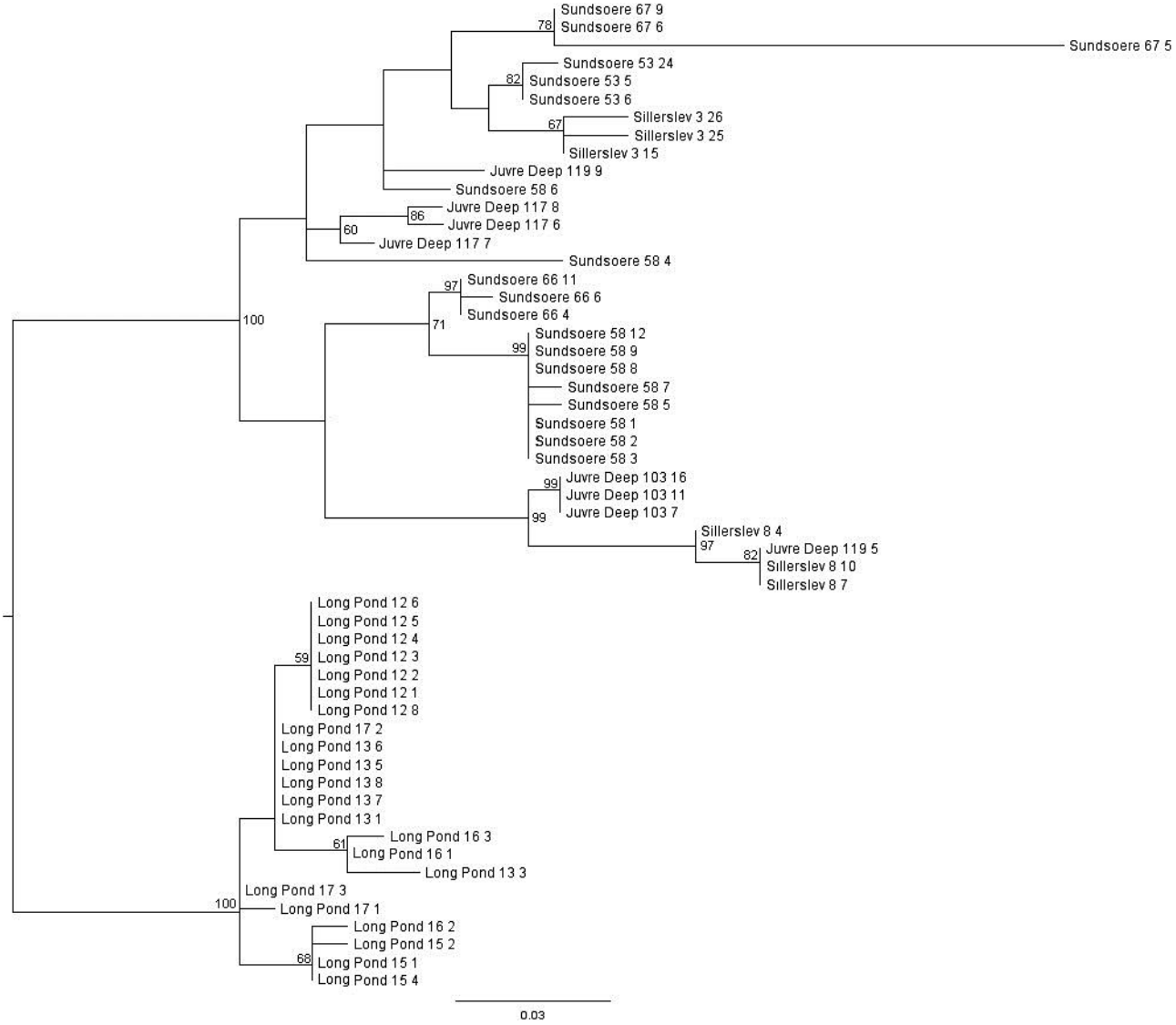

Figure 2. Maximum likelihood majority-rule consensus tree of *Ensis directus* internal transcribed spacers (ITS1 and ITS2) (concatenated data set) reconstructed following the GTR+I+G model. Bootstrap values ≥ 50% are indicated at the nodes. Sampling localities, individuals, and clones are specified.

ITS sequences did not cluster according to sampling localities in the area of Denmark in the ML trees. Instead, sequences from different localities were intermixed together in the same clades. However, a clustering by geographic areas was found in all trees. Higher bootstrap support (100%) for each geographic clade was obtained with the concatenated set of data (both ITS) (Fig. 2). ITS1 sequences alone were also informative in differentiating each geographic clade (bootstrap = 100%, tree not shown). ITS2 sequences were able to differentiate each geographic clade, but in this case bootstrap support was low (bootstrap = 43%, tree not shown).

The mean sequence divergence between *E. directus* localities was compared with the mean sequence divergence between different *Ensis* species (Table 3). In all cases except one, *p*-distances between localities were smaller than *p*-distances between species, even when the species alignment was corrected in the Gblocks Server (Castresana 2000) (Figs. S3–S4) (after alignment correction, mean divergence between species is somewhat reduced). The exception is the mean sequence divergence between *E. magnus* and *E. siliqua* (0.019 ± 0.006), which was somewhat smaller than the mean sequence divergence between Long Pond and the Danish localities in the ITS1 analysis. According to *p*distance, *E. macha* was the closest species to *E. directus* (for ITS1, 0.093 ± 0.013; for ITS2, 0.109 ± 0.018, using the Gblocks alignment). The *p*-distances between the Canadian locality and the Danish localities ranged between 0.025 ± 0.006 and 0.028 ± 0.006 for ITS1 and between 0.013 ± 0.005 and 0.021 ± 0.006 for ITS2.

Table 3. Mean sequence divergence (*p*-distance, mean ± SE) between species and between localities for both internal transcribed spacers (ITS1 and ITS2).

		ITS1		ITS2	
		Gaps considered	**Gaps excluded**	**Gaps considered**	**Gaps excluded**
sp	*E. directus* - *E. macha**	0.108±0.013	0.093±0.013	0.108±0.018	0.109±0.018
sp	*E. directus* - *E. siliqua**	0.208±0.018	0.195±0.019	0.181±0.023	0.179±0.023
sp	*E. directus* - *E. magnus**	0.218±0.019	0.203±0.019	0.170±0.022	0.170±0.022
sp	*E. macha* - *E. siliqua**	0.193±0.018	0.173±0.018	0.171±0.022	0.168±0.023
sp	*E. macha* - *E. magnus**	0.203±0.019	0.180±0.018	0.173±0.022	0.172±0.023
sp	*E. magnus* - *E. siliqua**	0.017±0.006	0.019±0.006	0.047±0.012	0.045±0.011
sp	*E. directus* - *E. macha*†	0.120±0.012	0.108±0.013	0.120±0.018	0.116±0.018
loc	Long Pond - Sillerslev	0.031±0.007	0.028±0.006	0.021±0.006	0.021±0.006
loc	Long Pond - Sundsøre	0.029±0.007	0.027±0.006	0.013±0.005	0.013±0.005
loc	Long Pond - Juvre Deep	0.027±0.007	0.025±0.006	0.019±0.006	0.019±0.006
loc	Sillerslev - Sundsøre	0.017±0.003	0.015±0.004	0.017±0.005	0.018±0.005
loc	Sillerslev - Juvre Deep	0.015±0.003	0.013±0.003	0.016±0.005	0.015±0.005
loc	Sundsøre - Juvre Deep	0.014±0.003	0.014±0.003	0.018±0.005	0.017±0.005

Note: Mean sequence divergence was calculated with gaps considered and with gaps excluded; the later values are cited in the text.

*Alignment was performed considering all species and was corrected in Gblocks. ITS1 alignment was reduced to 76% of its original length. ITS2 alignment was reduced to 81% of its original length.

†Alignment was performed considering only *E. directus* and *E. macha* sequences, with no correction in Gblocks (see main text)

Ensis ITS phylogenetic analyses

In the ITS1 ML analysis, all sequences belonging to the same species grouped together, with the exception of *E. siliqua*. In this species, ITS1 sequences did not form a single clade in the phylogenetic tree (Fig. 3*a*). However, in the ITS2 ML analysis, all sequences belonging to a single species were recovered as monophyletic, and species bootstrap support was higher than for ITS1 (Fig. 3*b*). Under MP, sequences belonging to a single species also grouped together, and bootstrap support for species nodes was higher with ITS2 sequences (≥97%). In MP analyses, gaps were treated as a fifth character state or as missing information. When gaps were treated as a fifth character state in the ITS1 MP analysis, the species node support increased in *E. siliqua*, decreased in *E. magnus*, and did not change in *E. directus* and *E. macha*. In the ITS2 MP analysis, species node support was higher when gaps were considered. The resulting topologies under MP criteria were identical regardless of the way in which gaps were treated.

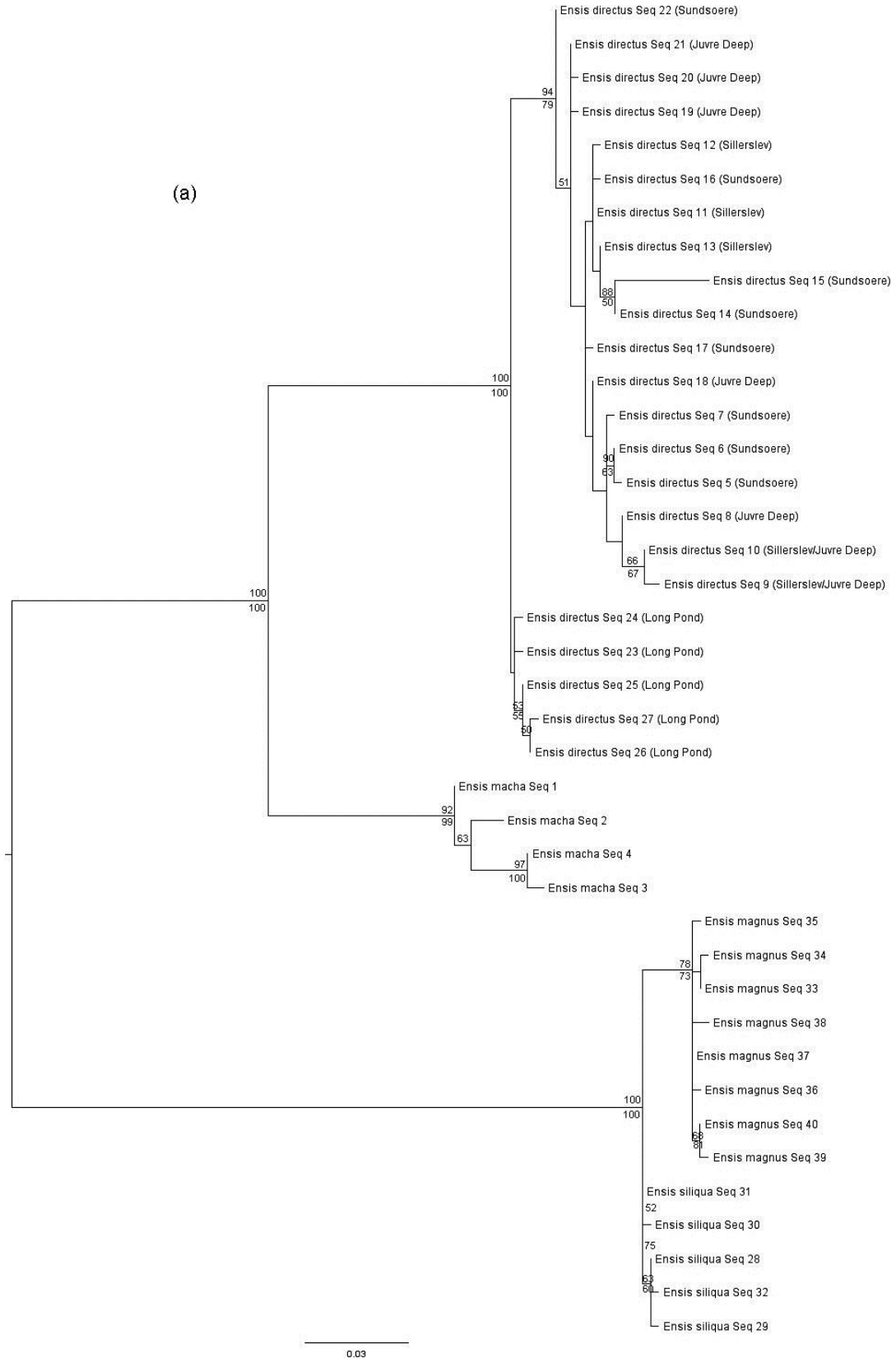

Figure 3. Maximum likelihood (ML) majority-rule consensus trees of *Ensis siliqua*, *E. macha*, *E. directus*, and *E. magnus* internal transcribed spacers (ITS1 and ITS2) reconstructed following the GTR+G model. Bootstrap values ≥ 50% obtained from the ML analysis are indicated above the nodes. When the same subtree was recovered in the maximum parsimony (MP) analysis, the corresponding bootstrap value (≥50%) is indicated below the node. *Ensis directus* sampling localities are indicated in parentheses. Gaps were treated as unknown characters in ML analyses and as fifth character states in MP analyses. (*a*) ITS1 phylogenetic tree. (*b*) ITS2 phylogenetic tree.

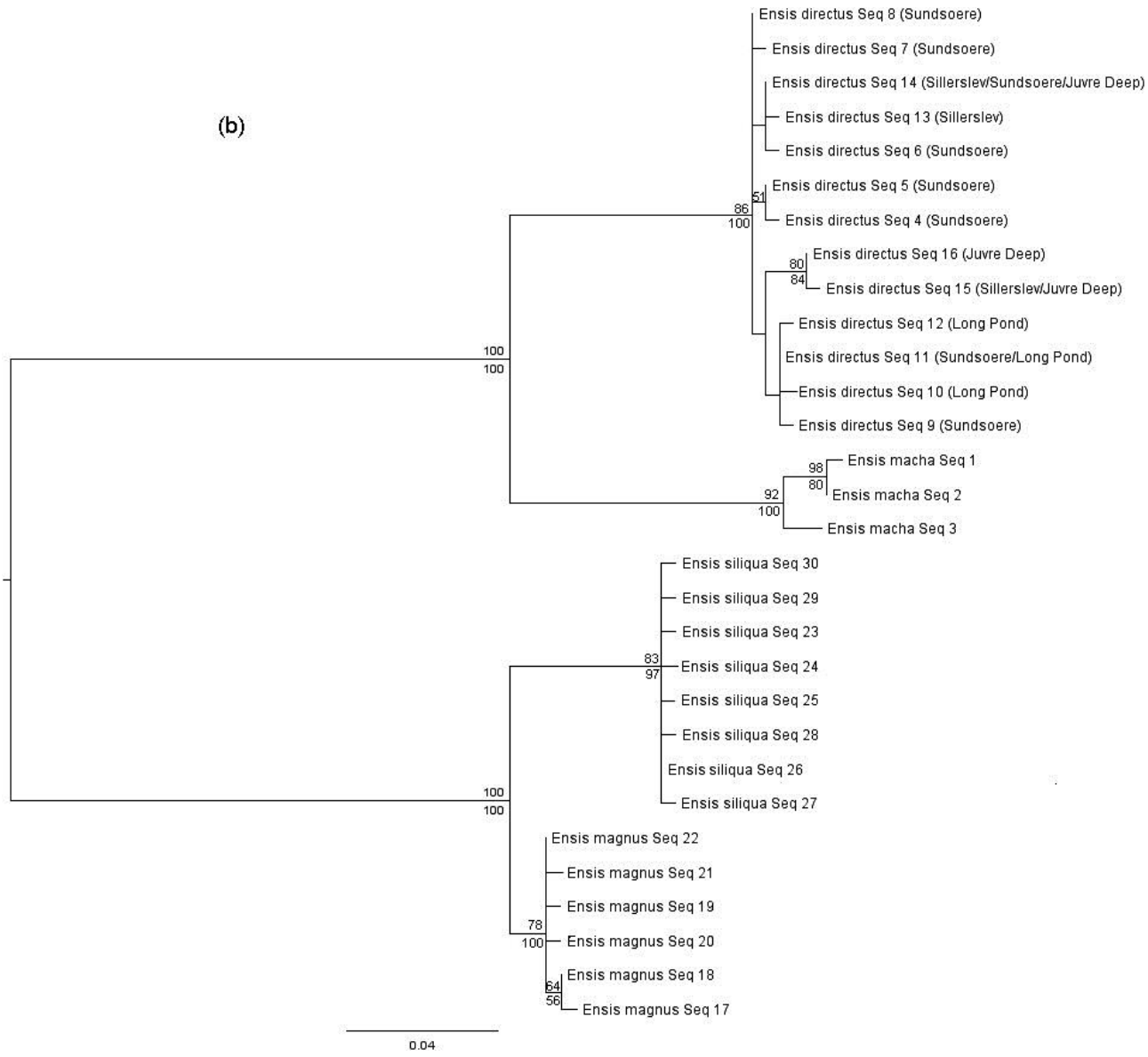

Figure 3. (concluded)

Discussion

Comparison of *Ensis directus* sequences with other mollusc ITS sequences

Marine molluscs are a diverse group according to the length and GC content of ITS1 (Chow *et al.* 2009). These authors studied the ITS1 of several marine animals, and extensive data regarding length and GC content for marine mollusc species are available. Average ITS1 length was 492.4 bp, and GC content was 55.9%. The values we obtained for *E. directus* ITS1 sequences are very similar to those values, so they are not an exception. Data on ITS2 length and GC content in Mollusca are not abundant, but some examples for bivalve species are available. ITS2 length was only 60 bp in the crocus giant clam, Tridacna crocea (Yu *et al.* 2000). It ranged between 317 and 446 bp in the Unionoidea species Unio pictorum, U. tumidus, U. crassus, Anodonta anatina, A. cygnea, Pseudanodonta complanata, and Margaritifera margaritifera (Källersjö *et al.* 2005). In the four scallop species studied by Insua *et al.* (2003), Aequipecten opercularis, Mimachlamys varia, Hinnites distortus, and Pecten maximus, ITS2 length ranged between 270 and 294 bp and GC content was 47%–50%. In the Veneridae species Meretrix meretrix, Cyclina sinensis, Mercenaria mercenaria, and Protothaca jedoensis, ITS2 length was 281–412 bp and GC content was 65.21%–67.87% (Cheng *et al.* 2006). Thus, *E. directus* ITS2 lengths were similar to those found in the scallop and venerid species. ITS2

GC content in *E. directus* is similar to that in venerids, as would be expected considering that scallops are Pteriomorphia bivalves, and venerids and *Ensis* species are Heteroconchia.

ITS as molecular markers in *Ensis directus*

Analysis of ITS variation in *E. directus* was performed by studying the length, indel polymorphism, haplotype diversity, and mean sequence divergence (*p*-distance) of each ITS at three different levels: individuals, localities, and geographic areas. Danish sequences were more variable than Canadian sequences in length, and they displayed a greater indel polymorphism, a greater haplotype diversity, and higher mean sequence divergence in both ITS1 and ITS2. The explanation of this geographic pattern should be addressed taking into consideration more localities from both native and introduced ranges and other data regarding the ecology and historical processes of *E. directus* (J. Vierna *et al.*, in preparation).

The detection of intragenomic divergence in both ITS in a majority of individuals let us conclude that ITS intragenomic divergence is widespread in *E. directus* populations. Interestingly, intragenomic mean sequence divergences in *E. directus* were in some cases higher than the mean sequence divergence for Danish or Canadian ITS sequences. Furthermore, the maximum *p*-distances obtained for ITS1 (0.052 ± 0.010) and ITS2 (0.034 ± 0.011) were considerably higher than the mean sequence divergence between Long Pond and the three Danish localities (Table 3). Similarly, Fairley *et al.* (2005) found that ITS1 sequence divergences within individuals were in some cases equal to or greater than those within localities in a study of the insect Anopheles aquasalis. In any nrDNA region under study, mean intragenomic divergence should be equivalent to mean divergence within populations (or localities forming a population). Nevertheless, estimates of mean intragenomic divergence (the values we obtain, as we cannot sample the entire population) may be much higher than "real" mean divergence if only a reduced number of clones are sampled per individual. This could have happened with individuals Sundsøre 67 and Juvre Deep 119. Thus, though intragenomic divergence occurs in *E. directus* individuals, it could have been overestimated in some of them.

However, the existence of some intragenomic divergence in ITS does not necessarily mean that these spacers are uninformative markers in molecular population studies. For example, Rodriguez-Lanetty and Hoegh-Guldberg (2002) compared the levels of intragenomic divergence in both ITS in the scleractinian coral Plesiastrea versipora with the divergence among populations. They found low levels of intragenomic divergence that were always lower than levels of divergence among populations. Therefore, they concluded that ITS were suitable molecular markers for that phylogeographic survey. In the insect Pediculus humanus, ITS2 displayed very high intragenomic divergence, so Leo and Barker (2002) concluded that it was an unsuitable marker for molecular population studies in that species.

Our results clearly show that both ITS were informative molecular markers at the population level regardless of the intragenomic divergence detected. Remarkably, although mean intragenomic divergences or maximum *p*-distances were sometimes higher than divergences between localities, ITS1 and ITS2 were able to differentiate each geographic area in both the MP cladogram — based on indel polymorphism — and the ML tree (Figs. 1–2). According to ITS1 and ITS2, both geographic areas are different populations, and this means that individuals from Denmark may have come from a source population other than Long Pond.

However, neither ITS nucleotide polymorphism nor indel polymorphism were informative in differentiating individuals belonging to each Danish locality. The most plausible explanation for this is that individuals from the three Danish localities form a single population, but this should be confirmed by other molecular markers. *Ensis* razor shells have external fertilization and undergo indirect development, so it is not unexpected that sampling localities that are separated by many kilometres are actually the same population.

ITS as molecular markers at the species level

In *Ensis* phylogenetic analyses, species node support was higher under MP than under ML criteria, with the exception of *E. magnus* (ITS1 tree). Similarly to studies done in the Unionidae bivalves (Källersjö *et al.* 2005), we found that the use of gaps as character states increased the node support for MP phylogenies in the

majority of cases (the exception is *E. magnus* in the ITS1 analysis). All ITS1 and ITS2 sequences belonging to each species were monophyletic (except *E. siliqua* sequences in the ITS1 ML tree), and both ITS recovered American and European species as reciprocally monophyletic. This is in accordance with morphological characters (Cosel 2009) and with the mean sequence divergence obtained between species (Table 2). Our results support the conclusion that both ITS are informative phylogenetic markers within genus *Ensis*, but the real ability of these spacers to resolve the phylogeny should be confirmed by other nuclear and mitochondrial sequences, and including all extant species. It should be taken into account that it is possible to get a species-specific clustering pattern if multiple divergences have occurred since the species last shared a common ancestor. That means that the pattern obtained could be due to the fact that the species may not be very closely related. If this were true, ITS sequences could show a paraphyletic pattern when more closely related species are considered. Though to date we cannot be sure whether *E. siliqua–E. magnus* and *E. directus–E. macha* are in fact sister species, we think they may be quite closely related, as they share some morphological features that distinguish them from other *Ensis* species. However, this point will not be clarified until more data on the phylogeny of genus *Ensis* become available.

Long-term evolution of Ensis ITS1–5.8S–ITS2

Considering the levels of intragenomic divergence detected in E. directus individuals, it seems that concerted evolution is not completely efficient. We sampled some ITS1 and ITS2 clones from individuals Sundsøre 67, Juvre Deep 103, Sillerslev 8, and Juvre Deep 119 that displayed some point mutations and indels that differentiated them from the other E. directus sequences. Though we cannot be completely sure, these clones do not appear to be pseudogenes because their corresponding 5.8S did not display any point mutation or indel. They may be copies that are not yet homogenized but that will eventually undergo homogenization through the mechanisms leading to concerted evolution. Considering all four Ensis species, tree topologies support a concerted evolution scenario, as ITS variants from each Ensis species are monophyletic. However, the long-term evolution of Ensis ITS1–5.8S–ITS2 can also be explained by other evolutionary models. Under a birth-and-death model, genes are shared for prolonged periods between species, except in exceptional cases when rapid turnover occurs. In these cases, rapid gene turnover can lead to the creation of species-specific gene clusters as a result of frequent gene duplication and loss. Consequently, few or no genes are shared between species (Rooney 2004). In the light of our results, birth-and-death processes and purifying selection may explain Ensis ITS1–5.8S–ITS2 long-term evolution. The monophyly of ITS sequences within each species may be the result of these rapid gene duplications and losses, whereas the evolution of the 5.8S may be driven by purifying selection. Sequence clusters observed within both ITS in E. directus (Fig. 2) may represent new copies arising by gene duplication along the evolutionary history of this species. In addition, the clear separation of ITS sequences belonging to each geographic area may be a consequence of this rapid gene turnover. Finally, a third possibility that may explain the long-term evolution of Ensis ITS1–5.8S–ITS2 is a mixed process of concerted and birth-and-death evolution, as described by Nei and Rooney (2005). According to this model, homogenizing mechanisms (unequal crossovers and gene conversions) and birth-and-death processes (gene duplication and loss) drive the long-term evolution of a given multigene family. Even though our results do not preferentially support any of the three models discussed, we believe that the most logical explanation is that homogenizing mechanisms, birth-and-death processes, and selection are all acting and shaping Ensis ITS1–5.8S–ITS2 extant variation.

The nontranscribed spacer region (NTS) of *Ensis* 5S ribosomal DNA was found to evolve under a birth-and-death model, but it was suggested that homogenizing mechanisms may be also taking part within each 5S ribosomal DNA variant in each species (Vierna *et al.* 2009). The organisation of 5S ribosomal DNA is more flexible than the organisation of the major ribosomal genes (and spacers), as it can be dispersed throughout the genome, found in its own tandem arrays, or found in both types of arrangement (Rooney and Ward 2005). Taking all this together, the long-term evolution of these two ribosomal families in *Ensis* could be reconciled under a mixed process of concerted evolution, birth-and-death evolution, and purifying selection (in the case of the 5S and 5.8S genes). Under this model, homogenizing mechanisms are more efficient within ITS1–5.8S–ITS2, as the major ribosomal genes may not be as dispersed as the 5S ribosomal DNA in

Ensis genomes, and they may be organised in a smaller number of arrays. Eickbush and Eickbush (2007) pointed out that unequal crossovers are more frequent between sister chromatids than between chromosomes (homologous and nonhomologous). This could be the reason why intragenomic divergence is much higher in *Ensis* NTS than in ITS. Nevertheless, more analyses should be performed to understand the chromosomal organisation of these ribosomal families in *Ensis* razor shells.

Conclusions

From this work, we conclude that (1) ITS1 and ITS2 are suitable molecular markers in *E. directus* despite the intragenomic divergence detected. Both indel and nucleotide polymorphisms are informative at the population level. (2) ITS1 and ITS2 may also be informative markers at the species level. Gaps are useful in reconstructing the phylogenetic relationships among *Ensis* species under MP. (3) The long-term evolution of *Ensis* ITS1–5.8S–ITS2 can be reconciled with the long-term evolution of *Ensis* 5S ribosomal DNA under a mixed process of concerted evolution, birth-and-death evolution, and selection in which the homogenizing mechanisms are less efficient within the 5S ribosomal DNA.

Finally, taking into consideration this and other studies, we recommend analysing the levels of intragenomic divergence before using any nrDNA region as a molecular marker.

Acknowledgements

We are grateful to K. Thomas Jensen, Anne S. Lousdal, and Ray J. Thompson for providing us with the *E. directus* samples, and to Marta Duyos Míguez for reviewing the English grammar. We give sincere thanks to Rudo von Cosel for identifying some of the specimens used in this work, and his comments on *Ensis* taxonomy. Finally, we would like to thank two anonymous reviewers that greatly improved the quality of this article with their comments. J.V. is supported by a "María Barbeito" fellowship from Xunta de Galicia (Spain).

References

Alvarez B, Krishnan M, Gibb K. 2007. Analysis of intragenomic variation of the rDNA internal transcribed spacers (ITS) in *Halichondrida* (Porifera: Demospongiae*).* J. Mar. Biol. Assoc. U.K. **87**(6): 1599-1605.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17**(4): 540-552

Cheng HL, Xia DQ, Wu TT, Meng XP, Ji HJ, Dong ZG. 2006. Study on sequences of ribosomal DNA internal transcribed spacers of clams belonging to the Veneridae family (Mollusca: Bivalvia). Acta Genet. Sin. **33**(8): 702-710

Chow S, Ueno Y, Toyokawa M, Oohara I, Takeyama H. 2009. Preliminary analysis of length and GC content variation in the ribosomal first internal transcribed spacer (ITS1) of marine animals. Mar. Biotechnol. (NY) **11**(3): 301-306

Cosel R von. 2009. The razor shells of the eastern Atlantic, part 2. Pharidae II: the genus *Ensis* Schumacher, 1817 (Bivalvia, Solenoidea). Basteria **73**: 9-56

Cosel R von, Dörjes J, Mühlenhardt-Siegel U. 1982. Die amerikanische schwertmuschel *Ensis directus* (Conrad) in der Deutschen Bucht. I. Zoogeographie und taxonomie mi vergleich mit den einheimischen schwertmuschel-Arten. Senckenb. Marit. **14**: 147-173

Eickbush TH, Eickbush DG. 2007. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics **175**(2): 477-485

Fairley TL, Kilpatrick CW, Conn JE. 2005. Intragenomic heterogeneity of internal transcribed spacer rDNA in neotropical malaria vector *Anopheles aquasalis* (Diptera: Culicidae). J. Med. Entomol. **42**(5): 795-800

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**(4): 783-791

Fernández A, García T, Asensio I, Rodríguez MA, González I, Hernández PE, Martín R. 2001. PCR-RFLP analysis of the internal transcribed spacer (ITS) region for identification of 3 clam species. J. Food Sci. **66**(5): 657-661

Fujiwara M, Inafuku J, Takeda A, Watanabe A, Fujiwara A, Kohno S, Kubota S. 2009. Molecular organization of 5S rDNA in bitterlings (Cyprinidae). Genetica **135**(3): 355-365

Gascuel O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. **14**(7): 685-695

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**(5): 696-704

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41**: 95-98

Harris DJ, Crandall KA. 2000. Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): implications for phylogenetic and microsatellite studies. Mol. Biol. Evol. **17**(2): 284-291

Heath DD, Rawson PD, Hilbish TJ. 1995. PCR-based nuclear markers identify alien blue mussel (*Mytilus* spp.) genotypes on the west coast of Canada. Can. J. Fish. Aquat. Sci. **52**(12): 2621-2627

Insua A, López-Piñón MJ, Freire R, Méndez J. 2003. Sequence analysis of the ribosomal DNA internal transcribed spacer region in some scallop species (Mollusca: Bivalvia: Pectinidae). Genome **46**(4): 595-604

Jarvis JP, Luedeman JK, Shier DR. 1983. Comments on computing the similarity of binary trees. J. Theor. Biol. **100**: 427-433

Källersjö M, von Proschwitz T, Lundberg S, Eldenäs P, Erséus C. 2005. Evaluation of *ITS* rDNA as a complement to mitochondrial gene sequences for phylogenetic studies in freshwater mussels: an example using Unionidae from north-western Europe. Zool. Scr. **34**(4): 415-424

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al.*. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**(21): 2947-2948

Leo NP, Barker SC. 2002. Intragenomic variation in ITS2 rDNA in the louse of humans, *Pediculus humanus*: ITS2 is not a suitable marker for population studies in this species. Insect Mol. Biol. **11**(6): 651-657

Mahidol C, Na-Nakorn U, Sukmanomon S, Yoosuk W, Taniguchi N, Nguyen TTT. 2007. Phylogenetic relationships among nine scallop species (Bivalvia: Pectinidae) inferred from nucleotide sequences of one mitochondrial and three nuclear gene regions. J. Shellfish Res. **26**(1): 25-32

Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. **39**(1): 121-152

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. **25**(7): 1253-1256

Rodriguez-Lanetty M, Hoegh-Guldberg O. 2002. The phylogeography and connectivity of the latitudinally widespread scleractinian coral *Plesiastrea versipora* in the Western Pacific. Mol. Ecol. **11**(7): 1177-1189

Rooney AP. 2004. Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in apicomplexans. Mol. Biol. Evol. **21**(9): 1704-1711

Rooney AP, Ward TJ. 2005. Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth and death of a concerted evolution paradigm. Proc. Natl. Acad. Sci. U.S.A. **102**(14): 5084-5089

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**(18): 2496-2497

Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Mass.

Sword GA, Senior LB, Gaskin JF, Joern A. 2007. Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera: Acrididae). Syst. Entomol. **32**(3): 420-428

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**(8): 1596-1599

Toro J, Innes DJ, Thompson RJ. 2004. Genetic variation among life-history stages of mussels in a *Mytilus edulis–M. trossulus* hybrid zone. Mar. Biol. (Berl.) **145**: 713-725

Vierna J, González-Tizón AM, Martínez-Lage A. 2009. Long-term evolution of 5S ribosomal DNA seems to be driven by birth-and-death processes and selection in *Ensis* razor shells (Mollusca: Bivalvia). Biochem. Genet. **47**(9–10): 635-644

Wörheide G, Hooper JNA, Degnan BM. 2002. Phylogeography of western Pacific *Leucetta 'chagosensis'* (Porifera: Calcarea) from ribosomal DNA sequences: implications for population history and conservation of the Great Barrier Reef World Heritage Area (Australia). Mol. Ecol. **11**(9): 1753-1768

Wörheide G, Nichols SA, Goldberg J. 2004. Intragenomic variation of the rDNA internal transcribed spacers in sponges (phylum Porifera): implications for phylogenetic studies. Mol. Phylogenet. Evol. **33**(3): 816-830

Yu ET, Juinio-Meñez MA, Monje VD. 2000. Sequence variation in the ribosomal DNA internal transcribed spacer of *Tridacna crocea.* Mar. Biotechnol. (NY) **2**(6): 511-516

Zhang J, Dyer KD, Rosenberg HF. 2000. Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. Proc. Natl. Acad. Sci. U.S.A. **97**(9): 4701-4706

[i] Supplementary data for this article are available on the journal Web site (http://genome.nrc.ca) or may be purchased from the Depository of Unpublished Data, Document Delivery, CISTI, National Research Council Canada, Building M-55, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada. DUD 5311. For more information on obtaining material refer to http://cisti-icist.nrc-cnrc.gc.ca/eng/ibp/cisti/collection/unpublished-data.html.

[ii]hakuna@udc.es