ORIGINAL ARTICLE

Recent and old duplications in crustaceans "Internal Transcribed Spacer 1": structural and phylogenetic implications

Alain Van Wormhoudt[1] · Mehdi Adjeroud[2] · Heloise Rouzé[3] · Matthieu Leray[4]

Abstract

Internal Transcribed Spacer structures are important in preserving accessibility to specific enzymes for the maturation of rRNAs. ITS1 sequences reported in the literature in Crustaceans range between 182 and 820 bp and are characterized by the absence of repeats or the presence of only a limited number of microsatellites. Here, we sequenced ITS1 for a range of shrimp families (infraorder Caridea) and show that most taxa have much larger ITS1 sequences. We find a high number of microsatellites in *Alpheus hebes* and *Crangon crangon* and we report repeat units in *Pandalidae*, *Palaemonidae* and mainly in *Alpheidae* species. Up to four repeats were found in *A. vanderbilti* (1915 bp), *A. rostratus* (1635 bp) and *A. lottini* (1625 bp). In general, four helices were found in ITS1. Repeat units led to extra hairpins and loops. No conserved positions occurred except in helix 4. Three clades were defined in *A. lottini* for the first time. We estimated the ITS1 divergence rate for the three clades of *A. lottini* collected in French Polynesia using existing calibrations of substitution rates. Rates of sequence evolution are largely influenced by repeat units, which likely evolve separately. By comparison with COI marker, we estimated the divergence rate of the whole ITS1 sequence to range from 0.5 to 1.4% Pmy and between 0.12 and 0.5% for the 3′ end of ITS1 located outside the repeat units. Given the degree of identity between repeats, we suggest that a duplication event recently occurred in *A. floridanus* (98% identity) whereas an ancient duplication happened in *A. sulcatus* (50% identity) early at the origination of the group *Alpheidae*, approximately 50 mya ago. In conclusion, our results highlight an over representation of shorter ITS1 sequences in public repositories, and underlines the importance to further understand patterns of molecular evolution of this functionally important gene.

Keywords ITS1 · Repeat units · Crustacean · Decapods · Secondary structure

✉ Alain Van Wormhoudt
  avw@mnhn.fr; alainvw29@gmail.com

1 Station de Biologie Marine du Muséum National d'Histoire Naturelle EPHE, Laboratoire Evolution Moléculaire et Adaptation, 29900 Concarneau, France

2 UMR Entropie, IRD, Université de Perpignan, Via Domitia, 66860 Perpignan, France

3 PSL Research University, EPHE-UPVD-CNRS, Laboratoire d' Excellence 'CORAIL', USR 3278 CRIOBE, BP 1013, 98729 Papetoai, Mooréa, French Polynesia

4 Smithsonian Tropical Research Institute, Smithsonian Institution, Panama City, Republic of Panama

Introduction

Ribosomal DNA (rDNA) is encoded by genes present in multiple copies organized in tandemly repeated clusters. In the genome of eukaryotes, each cluster consists of the 18S, 5.8S and 28S coding regions, separated by two internal transcribed spacers, ITS1 and ITS2. "Concerted evolution" is believed to maintain intragenomic similarities among copies of these sequences by mechanisms such as gene conversion and crossing over [22]. However, the gain or loss of repeats and the accumulation of mutations with the continual turnover of DNA results in high levels of nucleotide and insertion-deletion polymorphisms within and among populations [10]. As a result, most rDNA genes have proven useful markers for phylogenetic inferences at the species, genus and family levels.

The ITS1 gene has in general been excluded from phylogenetic applications due to its potential for high variations

in length [3]. Nonetheless, its rapid divergence rate may help delineate cryptic and incipient species. In arthropods, ITS1 size is highly variable. The largest sizes were reported in insects. For example, *Anopheles punctulatus* individuals contain multiple ITS1 variants that range from 1.2 to 8.0 kb [1]. In some other taxa, ITS1 length variation was observed due to three to six short internal repeats [2, 14, 35], or to longer repeated sequences [28]. In crustaceans, ITS1 lengths have been reported to range from 182 to 820 bp [3]. One exception, are the ITS1 sequences in crayfishes, the longest found thus far, which are about 1300 bp long due to a large number of single sequence repeats [7, 12].

The secondary structure of ITS1 plays an important role in defining the split sites, which allow the release of the rRNA molecules during the maturation process [23]. These structures are functionally important and, as a result, are thought to be well conserved [4]. Due to these evolutionary constraints on the split sites, it was postulated that the remainder of the sequences might evolve near the neutral level [34]. In most eukaryotes, ITS1 consists of an open multibranch loop with several helices [8]. Long hairpins are formed by tandem repeats paired with one another with a conserved motif in the terminal loop.

In this paper we first characterize the size range of ITS1 in Alpheidae and in other Caridea. We discuss the consequence of the presence or absence of repeat elements on the ITS1 structure. We then compare the evolutionary rate of COI and ITS1 among sympatric cryptic species of the ecologically important *Alpheus lottini* species complex explicitly including the effect of differences in the number of repeat units. Finally, we estimate the date of apparition of these events of divergence between cryptic clades and species taking into account the nucleotide substitution rate.

Materials and methods

Material

Crangon crangon, Athanas nitescens, Palaemon elegans, Palaemon serratus, Hippolyte varians and *Atyaephyra desmarestii* were collected in 2004 in Concarneau, Brittany. *Chorocaris chacei, Mirocaris fortunata* and *Rimicaris exoculata* were collected by D. Desbruyeres in 2007 on the Mid Atlantic Ridge (2400 m depth). *Plesionika hsuehuyi* was collected by B. Richer de Forges in 1993 in New Caledonia (camp. Bathus3). *Sergia robusta* and *Alpheus macrocheles* were collected by P. Noel on the Atlantic coast of Spain in 2000 and *Stylodactylus libratus* by R. Cleva in the Marquesas (France) in 1983. *Acanthephyra pelagica* and *Ephyrina figuerai* were collected by S. Iglesias in 2000 in the Bay of Biscay (depth 1250 m, La Croix Morand). We collected *Alpheus lottini* in the Indian Ocean (La Réunion) and in the Pacific Ocean, around the French Polynesian islands. Other samplings sources were indicated in Table 1. DNA extracts from *Alpheus cristulifrons*, *Alpheus cylindricus* (eastern Pacific), *Alpheus vanderbilti*, *Alpheus utriensis* (Caribbean) *Alpheus dentipes*, *Alpheus sulcatus*, *Alpheus floridanus*, *Alpheus rostratus* and *Alpheus hebes* were provided by N Knowlton and ST Williams.

Table 1 Collecting locality and provenance of *Alpheus lottini* specimens used in this study

	Latitude	Longitude	Source	No COI	No ITS1
Tulear	23.21.13 S	43.40.34 E	Tomassin, 1972, MNHN	1	0
Nosy Be	13.20 S	48.15 E	Grosnier, 1959, MNHN	1	0
La Réunion	21.06.52 S	55.31.57 E	Van Wormhoudt, Iglesias, 2008, Saline les Bains, MNHN	5	2
Djibouti	11.36 N	43.10 E	Guerin, Banner, Récif Ambouli, 1933, MNHN	1	0 (1)
Jubal Strait	27.40 N	33.55 E	Fehlmann, 1965, SMITHSONIAN INSTITUTE USNM 173985	1	0
Seychelles	4.37 S	55.27 E	Banner, 1980, MNHN	1	0 (1)
Maldives	4.10 N	72.30 E	Gardiner, 1900, récif Narfare, MNHN	1	0 (1)
Malaysia	3.05 N	101.40 E	Serene, 1958, MNHN	1	0
Marquesas	9.30 S	140 W	Odinez, 1996, MNHN	1	0
Bora- Bora	16.29.21 N	151.44.11 W	Van Wormhoudt,2005, MNHN	4	9
Rangiroa	15.07.30 N	147.38.42 W	Van Wormhoudt, Malpot, 2005, MNHN	6	5
Moorea	17.29.31 S	149.50.08 W	Van Wormhoudt, Adjeroud, 2005, MNHN	14	26
Vairao	17.49.09 S	149.17.32 W	Van Wormhoudt, Ifremer COP, 2002, MNHN	1	5
Fangataufa	22.15 S	138.45 W	Poupin, 1996, MNHN	2	1
Clipperton	10.18.14 N	109.13.04 W	Poupin, 2007, MNHN	2	4
Colombia	4.36.36 N	74.04.55 W	Wicksen, 1935, SMITHSONIAN INSTITUTE USNM 237168	1	0

Sampling sources, number of analysed COI and ITS1 sequences and estimated GPS coordinates of sampling locations are indicated. Numbers between parentheses indicate the number of specimens for which only the more conserved 3′ end of ITS1 was sequenced

DNA extractions, PCR and sequencing

DNA was extracted from 1 pleopod using the CTAB method and polymerase chain reactions (PCR) were performed on 0.1 µg of DNA. Two primers were designed for ITS1 amplification: ITS1FW (5′-CACACCGCCCGTCGCTACTA-3′) located at the 3′end of 18S and ITS3R (5′-TCGACSCAC GAGCCRAGTGATC-3′) located at the 5′ end of 5.8S rDNA. Three internal primers were used to finalize the longest ITS1 sequences: ITS5 (5′- GCACCTCAGAAGAGAACC ATG-3′), ITS24R (5′-GAAGCGGGGGTTCCCTCACAC-3′) and ITS30R (5′-CTGTGGTGGGCTCCAACCCT-3′). Two additional primers, ITS9 (5′GCCAATGCCCCAGGTGGG GTCA-3′) and ITS3R, were used to amplify the end of the molecule. Regarding COI, the primer combination COIF (5′-CCAGCTGGAGGAGGAGAYCC-3′) and H7188 (5′-CAT TTAGGCCTAAGAAGTGTTG-3′) was used [36]. All PCR reactions were done according to the GE Healthcare protocol (Ready to Go PCR) at 52 °C. Sequencing reactions were performed on purified PCR products with BigDye® sequencing reagents (Applied Biosystems™). The initial phase of denaturation (2 min at 96 °C) was followed by 40 cycles at 96 °C for 30 s, 50 °C for 30 s and 72 °C for up to 4 min. The DNA sequence was determined on an automated ABI3130 sequencer. In one case, ITS1 fragments were ligated into pGEM®-T Easy Vector (Promega) and transformed into *Escherichia coli* JM109 competent cells before being cloned and sequenced. Four colonies were extracted and the insert sequenced.

Data analysis

DNA sequences were aligned using Bioedit's ClustalW accessory application [11] and edited manually for ITS1. The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura-2-parameter distance [15] for COI and the Tamura-Nei method [27] for ITS1. This best model was selected using MEGA7 [18] with gamma distribution $G = 0.65$ and estimated invariable sites $= 0.50$. Bayesian posterior probabilities were also assessed using Bayesian Markov Chain Monte Carlo (MCMC) analysis in MrBayes (vers. 3.2.4). The phylogenetic tree reconstruction was computed with branch lengths proportional to the number of substitutions per site and was represented with both ML bootstrap support (BS) and Bayesian posterior probabilities (PP) values. The phylogenetic analysis involved 52 sequences for ITS1. All positions containing either gaps or missing data were eliminated.

Prediction of structural domains and motifs

Structural RNA folding elements were recognized with the help of mFOLD [13, 39] that screens for thermodynamically

optimal secondary structures. Default values were chosen to fold the ITS1 sequences and the folding procedure was reiterated to obtain optimal energies using default conditions. Change in temperature setting ($T = 25$ °C) did not affect the general architecture but did result in lower energy levels for secondary structure as suggested by Ki and Han [14].

Results

Variation of ITS1 size in Caridea

ITS1 sequences from 27 caridean species belonging to 9 different families were analysed (Table 2). The simplest organization was shown in Alvinocaridae, Stylodactylidae, Hippolithidae, some Alpheidae (i.e., *Athanas nitescens* and *Alpheus macrocheles*), Penaeidae (i.e., *Litopenaeus vannamei*) and in most Caridea who presented no repeat units, minisatellites or microsatellites. In other shrimps (*Crangon*, *Acanthephyra*, *Ephyrina*), microsatellites were detected and the size of the ITS1 ranged between 619 bp and 1179 bp. Although less common, repeat units were identified in the Caridae, Pandalidae and Palaemonidae. Two repeats were present, either at the 5′end of ITS1 (*Palaemon elegans*) or at the 3′ end (*P. serratus* or *Macrobrachium rosenbergii*). They were more common among Alpheidae, with up to four repeats in *Alpheus sulcatus, A. cristulifrons, A. lottini* and *A. vanderbilti* with a size ranging from 995 to 1915 bp. Sequence similarity among repeats was highly variable, ranging from 50 to 98%. The highest levels were reported in *A. rostratus* (91%) and in *A. floridanus* with up to 98% identity between repeat 1 and 2, whereas *A. cristulifrons* and *A. sulcatus* showed the lowest levels of similarity (Table 2). In *A. lottini* species complex, the size of ITS1 ranged from 1597 to 1624 bp with numerous gaps or insertions for specimens from Fangataufa (Fig. 1).

Repeat units and ITS1 structure in Caridea

All ITS1 sequences present helices and stems. The simplest structures were reported in *A. macrocheles* (Fig. 2a) or in *Ephyrina, Stylodactylus* and in *Rimicaris* (not shown), with four helices and stems. No sequence identities are reported within these stems and loops. Repeat units give either extra loops or mirror loops. These additional helices are present at different positions depending on thermodynamic constraints. Regarding the location of repeat units, no fixed position for repeat units could be determined. Two to three loops (D1-1 to D1-3), corresponding to two or three repeats were present at the beginning of ITS1 in *A. rostratus* (Fig. 2e), in *A. cylindricus* (Fig. 2f) or in *A. floridanus* (Fig. 2b). In *A. cristulifrons* (Fig. 2c), the repeat units, present in the middle of the spacer, give two extra loops at the same positions. In

Table 2 Characteristics of ITS1 across *Caridea*

Sub-families	Taxon	NCBI AC	Size	Repeat 1	Repeat 2	Repeat 3	Repeat 4	% 1/2	% 1/3	% 2/3	% 3/4	% 2/4
Alvinocarididae	*Chorocaris chacei*	KJ155574	601									
	Rimicaris exoculata	KJ155581	402	nd								
	Mirocaris fortunata	KJ155577	628									
Crangonidae	*Crangon crangon*	KJ155575	1179	231 (GA)7, 340 (GA)6, 610 (CA)7, 750 (GA)6								
Stylodactylidae	*Stylodactylus libratus*	KJ155583	352	nd								
Acanthephridae	*Acanthephyra pelagica*	KJ155571	619	255 (GA)10								
	Ephyrina figueirai	KJ155576	637	430 (GTT)7								
Atyidae	*Atyaephyra desmarestii*	KJ155573	597	nd								
Pandalidae	*Plesionika hsuehyui*	KJ155579	1490	49–359	410–740			57				
Palaemonidae	*Macrobrachium rosenbergii*	HM804252	1087	178–286		761–876		84				
	Palaemon serratus	KJ155580	1113	160–310		708–850		66				
	Palaemon elegans	KJ155578	1166	230–360	428–545			59				
Hippolytidae	*Hippolyte varians*	KJ155584	305	nd								
	Athanas nitescens	KJ155572	309	nd								
Alpheidae	*Alpheus macrocheles*	KJ155560	478	nd								
	Alpheus hebes	KJ155561	1522	245 (TAC)13, 301 (GAG)6, 922 (TGC)7								
	Alpheus rostratus	KJ155569	1635	200–490	510–750			91				
	Alpheus lottini (clade A)	KJ155550	1625	471–638	689–961	1029–1190	1215–1487		70			72
	Alpheus lottini (clade B)	KJ155549	1604	457–615	665–922	1007–1166	1169–1432		67			71
	Alpheus lottini sp (clade C)	KJ155559	1603	512–639	700–961	1038–1176	1178–1440		72			71
	Alpheus cristullifrons	KJ155562	995	235–315	355–450	545–657	775–882	64			50	
	Alpheus utriensis	KJ155563	1021	235–315	355–450	545–657	775–882	64			50	
	Alpheus vanderbilti	KJ155564	1915	220–329		780–887			89			
	Alpheus cylindricus	KJ155565	1820	224–333		805–917			89			
	Alpheus floridanus	KJ155567	1031	10–130	130–245	246–355		98	66	67		
	Alpheus sulcatus	KJ155570	1229	90–275	276–460	490–680	720–910	50			54	
	Alpheus dentipes	KJ155566	964	68–180	250–360			56				

Estimation of sizes, number of microsatellites, number and percent identity of repeat elements (over 50%) are indicated. % indicated the different identities between the different repeats

Fig. 1 Alignment of ITS1 sequences of the five *Alpheus lottini* cryptic species: Reunion 2 corresponds to Clade A2, Rangiroa 4 to clade A1, Bora Bora M1 to clade B1 Fangataufa to clade C and Moorea A11 to clade B2. Gaps are bordered

Fig. 2 Analysis of the secondary structure of ITS1 in selected crusta-▸
cean: Hairpin structures were formed at the lowest free energy values:
the optimal folding was selected according to the revised energy rules
that tend to contain more "correct" base pairs. **a** *Alpheus macroche-*
les, **b** *A. floridanus*, **c** *A. cristulifrons*, **d** *A. lottini*, **e** *A. rostratus*, **f** *A.*
cylindricus, **g** *A. sulcatus*, **h** *Palaemon serratus*

A. lottini, duplication occurred in the middle of the spacer
with two alternate extra loops (D2-2 and D3-2) presenting
some kind of symmetrical structure (Fig. 2d). In *A. sulcatus*
(Fig. 2g), although repeats presented very low identities,
two extra loops were found. In *P. serratus* (Fig. 2h), where
the repeats are located at two different positions, we noticed
the presence of two extra loops at two different positions
(D2-1 and D2-2).

Conserved last ITS1 stems and phylogeny of Alpheidae

The 3′end ITS1 stems were aligned and compared between
Alpheidae. Pairwise sequence identities, ranging from 0.602
to 0.976%, were measured between all species except with
the more divergent *A. sulcatus* (Table 3). These last stems
showed splicing sites and conserved sequences in the last
helix 4 (Table 4).

Importance of ITS1 for delineation of *A. lottini* cryptic species and comparison of evolution rates with that of COI

The comparison of divergence rates between ITS1 repeats
and COI within the *A. lottini* species complex (Table 5)
highlighted two distinct clades (A and B; following nomen-
clature of [36] with two well-supported sub-clades within
each major clade (A1, A2 and B1, B2; Fig. 3). In clade
A, 2.1% divergence was measured between the samples
of the Indian Ocean (A2) and those from the Pacific (A1)
(Table 5a). Within clade B, 3% divergence was measured
among Pacific sub-clades (Table 5a and Sup. 1). Specimens
of both sub-clades were collected in Moorea and Bora Bora,
occasionally together in the same colony of the coral *Pocil-*
lopora damicornis. Levels of sequence divergence between
clades A and B and specimens of the cryptic *A. lottini* spe-
cies collected in Fangataufa (clade C) ranged from 11.5 to
12.9%.

These values varied from 1 (B1 vs. B2) to 3.5% (A1 vs.
A2) when considering the 3′ end of ITS1 (Table 5a over
diagonal). With regards to COI, levels of sequence diver-
gence between sub-clades of clades A and B ranged from
3.1 to 8.5%. Clade C from Fangataufa was 12.6 to 13.8%
divergent from clades B and A, respectively (Table 5b and
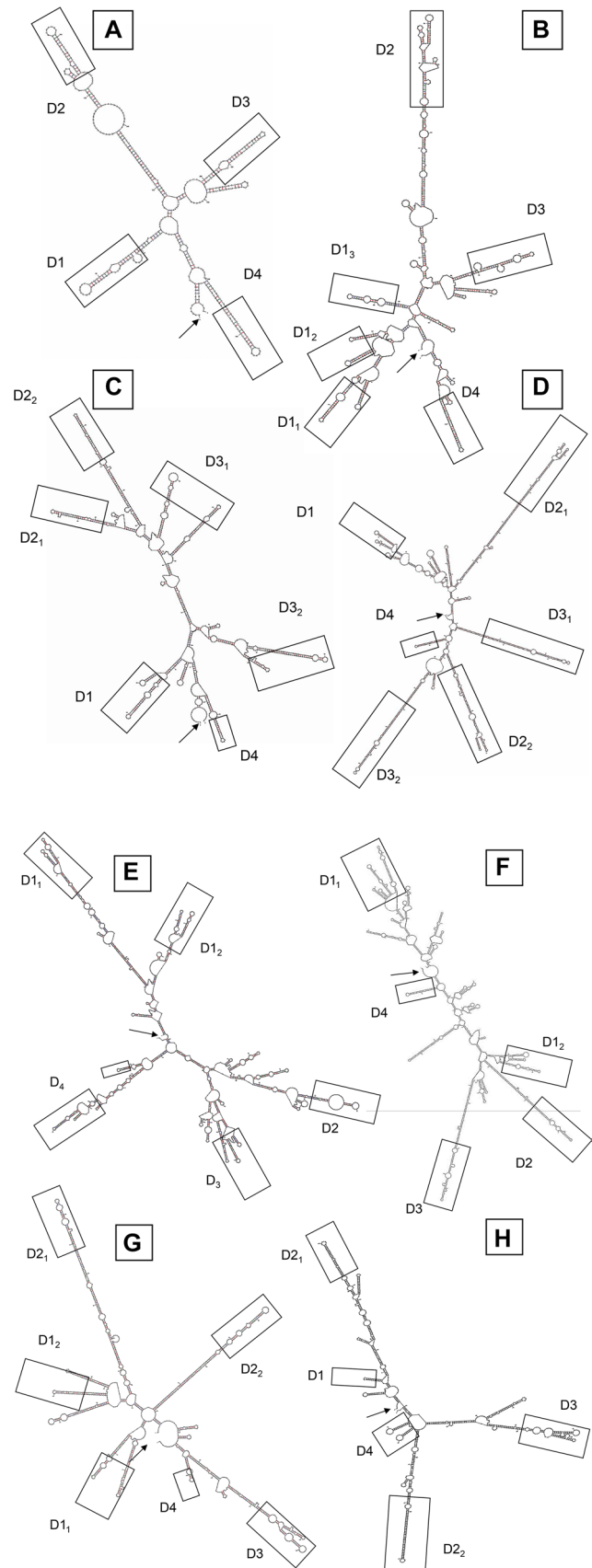Sup. 2).

Table 3 Sequence identity of the last ITS1 stems (conserved 3′ ends) between species of Alpheidae

Species	lottini A	lottini B	Fangataufa	macrocheles	cristullifrons	utriensis	cylindricus	vanderbilti	dentipes	floridanus	hebes	rostratus	sulcatus
A. lottini A	ID												
A. lottini B	0.976	ID											
A. lottini C	0.929	0.952	ID										
A. macrocheles	0.661	0.671	0.679	ID									
A. cristulifrons	0.666	0.676	0.661	0.733	ID								
A. utriensis	0.666	0.676	0.653	0.725	0.95	ID							
A. cylindricus	0.746	0.75	0.772	0.604	0.62	0.613	ID						
A. vanderbilti	0.753	0.757	0.779	0.611	0.623	0.615	0.977	ID					
A. dentipes	0.744	0.736	0.729	0.613	0.646	0.653	0.805	0.794	ID				
A. floridanus	0.639	0.648	0.641	0.688	0.747	0.731	0.627	0.623	0.669	ID			
A. hebes	0.664	0.674	0.681	0.565	0.605	0.627	0.618	0.618	0.615	0.602	ID		
A. rostratus	0.707	0.717	0.695	0.574	0.628	0.628	0.633	0.633	0.617	0.621	0.722	ID	
A. sulcatus	0.361	0.361	0.361	0.355	0.315	0.315	0.35	0.356	0.361	0.353	0.303	0.322	ID

Discussion

A majority of ITS1 sequences reported for crustacean in NCBI library are short. Yet, our results for a broader range of Caridea show that most taxa have long ITS1 sequences due to the presence of microsatellites or repeat units. Repeat units were very common in Alpheidae. Recent duplication (84–98% identity) occurred in *A. floridanus* and *A. rostratus*. Given the low levels of similarity between repeat units in *A. sulcatus,* this species may represent one of the oldest Alpheidae as suggested by Williams et al. [37]. "Old duplications" may suggest that the gene duplication arose before members underwent subsequent speciation and in the case of *A. sulcatus* occurred at the beginning of radiation of Alpheoidea while "recent duplications" suggest a continuous evolution. The absence of repeats over a long time period: 400–437 mya for Penaeidae and around 150–236 mya for Caridae [6] may be explained by the utmost importance of maintaining their physical structure due to higher specificity of nucleases involved in the processing of pre-ARN, opening a new field in evolution.

In inferred secondary structures, repeats form generally long hairpins, with a conserved motif in the terminal loop, and tandem repeats pair with one another over most of their length. This variability excludes in general ITS1 from phylogenetic applications [3], but suggests the general feasibility of population studies based on this marker. In most eukaryotes where this has been investigated, ITS1 consists of an open multibranch loop with several helices [8]. The presence of repeat units affects largely the number of loops and their positions. Only three loops have been detected in mosquitoes [2]. In our cases, the minimum is four (*A. macrocheles*) as in molluscan Pectinidae [34] and Haliotidae [30]. Up to seven ITS1 structural domains have been reported in a Trematoda [32] and six in our case. Some authors proposed some lack of constraints in the middle of the spacer [14], but it appears from our results that this lack of constraint should be extended to other domains. Either, D1 loops may be duplicated as in *A. rostratus, A. cylindricus* or *A. floridanus* or D2 loop as in *P. serratus*. In *A. sulcatus*, the two different repeats gave rise to two extra D1 and D2 loops. In *A. lottini* two extra loops correspond to repeats 2 and 3. In ten ladybird species (Coccinellidae) [33] a repeat was present in the middle and another one at the 5′end and the ITS1 size range from 791 to 2572 bp. In our case, only the stem and helix 4 can properly be aligned and are conserved across different families as in Boroginales [8], which suggest an important role in processing. Meanwhile, processes of rRNA cleavage involved in the release of mature rRNA remain poorly understood as well as the specificity of the

Table 4 ITS1 splicing sites and conserved sequences (in bold) in helix 4 of the *Alpheidae*

	Splice 5'	helix 4 and splice 3'
Alpheus lottini A	TAACGA	GTGCCAGAAAAGAAA------CCCCAACCTTTAAAACCCG-------TCACTCTA--GCGCTTGCGGA-GGCGCCCTGGAGCGGC-G------------------ GGCCCACGTGGGT--GTGGG--TCTGCTGCGCCGGGGTGCTGAGGCATGAC---AACTTACAA
Alpheus lottini B	TAACGA	GTGCCAGAAAAGAAA------CCCTAACCTTTAAAACCCG-------TCACTCTA--GCGCTTGCGGA-GGCGCCCTGGAGCGGC-G------------------ GGCCCACGTGGGT--GTGGG--TCTGCTGCGCCGGGGTGCTGAGGCATGAC---AACTTACAA
Alpheus.sp Fanga	TAACAA	GTGCCAGAAAAGAAA------CCTTACCCTTTAAAACCTG-------TCACTCTA--GCGCTTGCGGA-GGTGCCCTGGAGCGGC-G------------------ GGCCCACGTGGGT--GTGGG--TCTGCTGCGCCGGGGGGCTGAGGCTTGAC---AACTTACAA
A. macrocheles	TATGAA	GTGCCAAAAAAA------------CCTCTTTGAAACCCT------TGATTCCTTTGTGCTTGCCTTTGGTGCTCTGGCGTGGC-C------------------ GGCCCCCGTGGAG--GTGCGGTCTTGGTGCGCTGGGGTGCTAATGCTTGAC---AACTTACAA
A. cristulifrons	TAAAGA	GTGCCAGAGAAA------------AAACTTT-AAACCCT-------TAATTCTTTTCTGCTTGCGGA-GGTGCTCCGGAATGGC-A------------------ GGCCCGTGTGGAC--GTGCGGGCTTGCATGGCCGGGGTGCTGAG-GCTGCA---AACTTACAA
A. utriensis	TACTGA	GTGCCAGAGAAA-------------AAACTTT-AAACCCT------TAATTCTTTTCTGCTTGCGGA-GGTGCTCCGGAATGGC-A------------------ GGCCCGTGTGGAC--GTGCGGGCTTGCATGGCCGGGGTGCTGAGAGATGTC---AAACTACAA
A. cylindricus	TACGTC	GTGCCAGAGAATGAAACCTTTACTCTCTTTT--AAACCT--------TTACTCTA--GCGCTTGCGGA-GGTGCTCTGGAGCGGC-G------------------ GGCCCATGTGGGT--GTGGGAC--TGCTGCCCCGGGGTGCTGAGGCCTGAC-AAACTTTACAA
A. vanderbilti	TACGTC	GTGCCAGAGAACGAAACCTTTACTCTCTTTTTTTAAACCT-------TTACTCTA--GCGCTTGCGGA-GGTGCTCTGGAGCGGC-G------------------ GGCCCATGTGGGT--GTGGGAC--TGCTGCCCCGGGGTGCTGAGGCCTGAC-AAACTTTACAA
A. dentipes	TAACGA	GTGCCAGAGAA---------ACCCATTTTT--AAACCT--------TTACTCTTT-GCGCTTGCGGA-GGTGCTCTGGAGCGGC-G------------------ GTCCCACGTGGGTTAGTGGGAC--TGCTGTGACGGGGTACTGAGGGATGAC-AAACTT-ACAA
A. floridanus	TAAAGA	GTGCCAGAGAAA------------CC-TTTC-AAACCCA-------TTATTCTTTACCGCTTGCGGA-GGTGCCTTGCCGTGGC-T------------------ TGCTCATGTGGAT--GTGTGGGCTGGCTGTGGTGGGGTGCTGAGGGATGCA--AAACATACAA
A. hebes	TAACGA	GTGCCAGAAAGACCT----TCTTTA---CCCAAAAACCAACCC-----CAATCTT--GCGCTTGCGGA-GGTGCCCCGGAGCGGC-G------------------ GGCCCACGTGGAT--GTGAGGGCCTGCTGCGACGAG-TGCTGAGAG--GAG-ACAAACTACAA
A. rostratus	TACCGT	GTCCCAGAAAAAAT----GAGCAAAAACCCAAAACCCCTTTT----TCATTCTA--GCGCTTGCGGA-GGTGCCCCGGAGCGGC-GTCG-------------- GGCCTATGTGGAT--GTGTGGGCCGGCTGCAACGGG-TGCTGA-GGAAGAC-AAACTTACAA
A. sulcatus	TACGTT	GGGCCGGAGGACG-----------ACGTCACAGGTGCTACTG----TTGCTGGCT-CTGAAGCCGTGTGGCAGGTTTGAGCGTCCCTGGTGTGCGAGGGACCCC- GCTTCTCGTGGGACGGCT---------TGC---GGGGTCCTCCTGTGCG-----AAATTACAA

The last conserved loop is enclosed

Table 5 Computed pair wise distance between sisters' clades of *Alpheus lottini*. Clades are defined in Fig. 3. (a) Up the diagonal are indicated the results obtained with the 3′ end of the ITS1, last 493 bp out of duplicated sequences, and down the diagonal with the total ITS1 sequences. (b) Pair wise distances for COI are given in comparison for the same individuals collected at the same places

(a) ITS1	nb	% distance	Clade C	Clade A1	Clade A2	Clade B1	Clade B2	% distance	nb
Clade C	1	nd	_	7.9	8.4	6	6.2	nd	1
Clade A1	29	1.2	11.5	_	3.5	4	4.3	1.7	30
Clade A2	2	1.2	11.8	2.1	_	4.7	4.9	3.8	4
Clade B1	11	1.8	12.9	5.8	6.1	_	1	0.7	11
Clade B2	9	1.9	11.8	4.6	5.2	3	_	0.5	9

(b) COI	nb	% distance	Clade C	Clade A1	Clade A2	Clade B1	CladeB2
Clade C	2	nd	_				
Clade A1	12	2.2	12.6	_			
Clade A2	12	0.8	13.5	3.1	_		
Clade B1	12	8.2	13.8	10	10.8	_	
Clade B2	5	6.5	13.6	7.2	8.9	8.5	_

different nucleases that are implicated in the processing of the 5′end of ITS1 (Lamama and Karbstein, [19] or the 3′end of ITS1 [9].

The existence of two major clades within the *Alpheus lottini* species complex (A and B) was confirmed with ITS1 sequences. These clades had previously been characterized based on COI with a divergence ranging from 10 to 13% (Knowlton and Weight, [16, 26, 29, 38] using samples collected in different localities (Sup. 2). Furthermore, we believe that the use of a combination of different mitochondrial and nuclear markers will allow the detection of past (due to a longer persistence of mtDNA) and recent hybrids (unpublished data on the geographical mode of speciation in *A. lottini* complex) as observed in other species [25].
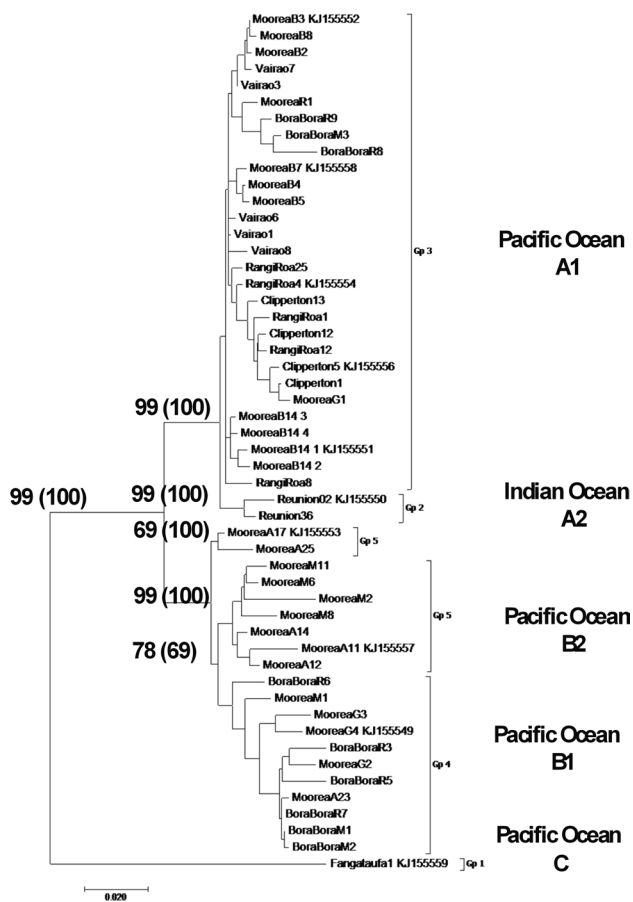
Fig. 3 Phylogeny of *Alpheus lottini* subspecies inferred using ITS1 sequences from maximum likelihood and Bayesian approaches. Gen-Bank accession numbers are indicated when nucleotide sequences were downloaded from NCBI. Maximum Likelihood bootstrap values (BS; first value) and Bayesian posterior probabilities (PP; in brackets) are indicated as well as the major clades A, B and C delineated in the *A. lottini* species complex

the absence of a physical separation between these Oceans at some geological periods [38]. However, it does not exclude a more recent isolation of the Indian Ocean during the Pleistocene glacial events (~ 700,000 years ago) that resulted in a 120 m drop in sea level [24]. The limited time since the physical separation has not allowed the genetic differentiation of these genes. In addition to previous results [16, 38] we characterized the existence of the two clades A and B in Moorea and Bora Bora, confirming results obtained by Williams et al. [38] using nuclear genes coding for myosin. By comparing pairwise distances between COI and ITS1 and a rate of COI evolution of 1.5% substitutions Pmy as proposed by Knowlton and Weigt [17], ITS1 substitution rate may range from 0.8 to 1.2% Pmy for *A. lottini*. This rate is high compared to other species [20]. For example, it ranges from 0.22 to 0.3% Pmy for *Haliotis* species [5] to 0.775% for bivalves [21]. These values are lower if we consider the 3′ end of ITS1 (493 bp) located outside the repeat units. We obtained a substitution rate around 0.12 to 0.5% Pmy, which suggests that the different units have evolved independently. These values are comparable to those found in other invertebrates [31]. Overall, duplication events may have occurred 2 mya for the most recent ones (98% identity) to nearly 100 mya for the oldest ones (50 to 56% identity). This indicates that divergence rates reported here for ITS1 repeat units may be overestimated and more studies are necessary to better understand their origination and their functional role in the ITS1 structure. By comparisons of the structure of the different ITS1, complex of species, recently described as sister species can be well defined such as *A. vanderbilti*/*A. cylindricus* or *A. cristulifrons*/*A. utriensis* which presented the same repeat units at a same position.

Recently, the complexity of coral-Alpheides symbionts has been characterized by mitochondrial analysis by two of us [26]. In particular, it revealed that each of two genetically analysed lineages of *A. lottini* had different affinities with trapeziid crabs. Identifying cryptic species involved in these associations and their ecological roles remain challenging by using classical markers. ITS1, with its high degree of polymorphism will be very useful.

Our results confirm the presence of a complex of species that evolved separately. *Alpheus* sp. from Fangataufa, which is morphologically very similar to *A. lottini* correspond to a new clade to be studied in detail (Rouzé et al. in prep) even if it appears to be 11.5–12.9% divergent from *A. lottini* clades A and B based on ITS1 sequence alignments. Specimens of clade A were present in both the Indian and the Pacific Ocean (A1 and A2) and the difference between the two subclades is not significant [26] even if the presence of several insertions or deletions is reported (Sup. 1). It may confirm

Conclusion

Repeat units should be taken into consideration when estimating the ITS1 evolution rate. Our results confirm that consensus motifs are not universal while the conservative structures, even repeated, remain important for the action of nucleases. More studies are necessary to understand the role of repeat units, which are in variable numbers and in constant evolution. ITS1 could be very useful to trace the evolution of a complex of species and establish the bonds between different species. Moreover, they can be used to identify cryptic species or hybrids involved in coral-Alpheides symbionts.

in the manuscript and specially reviewer #2 for providing constructive comments on this manuscript.

Compliance with ethical standards

References

1. Beebe NW, Cooper RD (2002) Distribution and evolution of the *Anopheles punctulatus* group (Diptera: Culicidae) in Australia and Papua New Guinea. Int J Parasitol 32:563–574

2. Bower JE, Cooper RD, Beebe NW (2008) Internal repetition and intra individual variation in the rDNA ITS1 of the *Anopheles punctulatus* group (Diptera: Culicidae): multiple units and rates of turnover. J Mol Evol 67(4):397–411

3. Chu KH, Li CP, Ho HY (2001) The first internal transcribed spacer (ITS-1) of ribosomal DNA as a molecular marker for phylogenetic and population analyses of crustacean. Mar Biotechnol 3:355–361

4. Coleman AW, Preparata RM, Mehrotra B, Mai JC (1998) Derivation of the secondary structure of the ITS-1 transcript in Volvocales and its taxonomical correlations. Protist 149(2):135–146

5. Coleman AW, Vacquier VD (2002) Exploring the phylogenetic utility of sequences for animals: a test case for abalone (*Haliotis*). J Mol Evol 54:246–257

6. De Grave S, Pentcheff ND, Ahyong ST, Chan TY, Crandall KA, Dworschak PC, Felder DL, Feldmann RM, Fransen CHJM, Goulding LYD, Lemaitre R, Low MEY, Martin JW, Ng PKLN, Schweitzer CE, Tan SH, Tshudy D, Wetzer R (2009) A classification of living and fossil genera of decapod crustaceans. Raffles Bull Zool 21(Supplement):1–109

7. Edsman L, Farris JS, Kallersjö Prestegaard T (2002) Genetic differentiation between noble crayfish *Astacus astacus* (L), populations detected by microsatellite length variation in the rDNA ITS1 region. Bull Fr Pêche Piscic 367:691–706

8. Gottschling M, Hilger HH, Wolf M, Diane N (2001) Secondary structure of the ITS1 transcript and its application in a reconstruction of the phylogeny of Boroginales. Plant Biol 3:629–636

9. Granneman S, Petfalski E, Tollervey D (2011) A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. EMBO J 30(19):4006–4019

10. Graur D, Li WH (2000) Fundamentals of molecular evolution. Sinauer Associates, Sunderland

11. 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

12. Harris DJ, Crandall K (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

13. Jaeger JA, Turner DH, Zucker M (1990) Predicting optimal and sub-optimal secondary structure for mRNA. Methods Enzymol 183:281–306

14. Ki JS, Han MS (2007) Cryptic long internal repeat sequences in the ribosomal DNA ITS1 gene of the dinoflagellate *Cochlodinium polykrikoides* (dinophycea): a 101 nucleotide six-repeat track with a palindrome-like structure. Genes Genet Syst 82:161–166

15. Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

16. Knowlton N, Weigt LA (1997) Species of marine invertebrates: a comparison of the biological and phylogenetic species concepts. In: Claridge MF, Dawah HA, Wilson MR (eds) Species: the units of Biodiversity. Chapman and Hall, London, pp 199–219

17. Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. Proc R Soc Lond B Biol Sci 265:2257–2263

18. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

19. Lamana AC, Karbstein K (2009) Nob1 binds the single-stranded cleavage site D at the 3′-end of 18S rRNA with its PIN domain. Proc Natl Acad Sci 106(34):14259–14264

20. Marko PB (2002) Fossil calibration clocks and the divergence time of germinate species pairs separated by the Isthmus of Panama. Mol Biol Evol 19(11):2005–2021

21. Ni G, Li Q, Kong L, Zheng X (2012) Phylogeography of bivalve *Cyclina sinensis*: testing the historical glaciations and Changjiang river outflow hypotheses in Northwestern Pacific. PLoS ONE 7(11):1–12

22. Ohta T (2000) Evolution of gene families. Gene 259(1–2):45–52

23. Phipps KR, Charette JM, Baserga SJ (2011) The SSU processome in ribosome biogenesis: progress and prospects. WIREs RNA 2(1):1–21

24. Randall JE (1998) Zoogeography of shore fishes of the Indo-Pacific Region. Zool Stud 37(4):227–268

25. Roussel V, Van Wormhoudt A (2016) The effect of Pleistocene climate fluctuations on distribution of European abalone (*Haliotis tuberculata*). Biochem Genet 55:124–154

26. Rouzé H, Leray M, Magalon H, Penin L, Gelin P, Knowlton N, Fauvelot C (2017) Molecular characterization reveals the complexity of previously overlooked coral-exosymbiont interactions and the implications for coral-guild ecology. Sci Rep 7:44923

27. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

28. Van Herwerden L, Blair D, Agatsuma T (1999) Intra- and inter-individual variation in ITS1 of *Paragonimus westermani* (Trematoda/Diagenae) and related species: implication for phylogenetic studies. Mol Phylogenet Evol 12(1):67–73

29. Van Wormhoudt A (2009) Existence de sous-espèces d' *Alpheus lottini* à Clipperton. In: Charpy L (Coord) Clipperton: environnement et biodiversité d'un microcosme océanique. IRD, Marseille, Muséum National d'Histoire Naturelle, Paris, Patrimoines Naturels, 420 pp, 68:172–173

30. Van Wormhoudt A, Gaume B, Le Bras Y, Roussel V, Huchette S (2011) Two different and functional nuclear rDNA genes in the abalone *Haliotis tuberculata*: tissue differential expression. Genetica 139(10):1217–1227

31. Vilas R, Criscione CD, Blouin MS (2005) A comparison between mitochondrial DNA and the ribosomal internal transcribed regions in prospecting for cryptic species of platyhelminth parasites. Parasitology 131:1–8

32. Von der Schulenburg JHG, Englisch U, Wagele JW (1999) Evolution of ITS1 rDNA in the Digenea (Platyhelminthes: Trematoda): 3′ end sequence conservation and its phylogenetic utility. J Mol Evol 48:2–12

33. Von der Schulenburg JHG, Hancock JM, Pagnamenta A, Sloggett H, Majerus MEN, Hurts GD (2011) Extreme length variation in the first ribosomal internal transcribed spacer of ladybird beetles (Coleoptera; Coccinellidae). Mol Biol Evol 18(4):648–660

34. Wang S, Bao Z, Li N, Zhang L, Hu J (2007) Analysis of the secondary structure of ITS1 in Pectinidae: implications for phylogenetic reconstruction and structural evolution. Mar Biotechnol 9:231–242

35. Warberg R, Jensen T, Frydenberg J (2005) Repetitive sequences in the ITS1 region of ribosomal DNA in congeneric microphallid species (Trematoda: Digenea). Parasitol Res 97(5):420–423

36. Williams ST, Knowlton N, Weigt LA (1999) Indo-Pacific molecular biogeography of the coral-dwelling snapping shrimp *Alpheus lottini* (Decapoda: Caridea: Alpheidae). In: Sheppard CRC, Seaward MRD (eds) Ecology of the Chagos Archipelago, Occasional publications of the Linnean Society, vol 2. Westbury for the Linnean Society of London, Otley, pp 195–206

37. Williams ST, Knowlton N, Weigt LA, Jara JA (2001) Evidence for three major clades within the snapping shrimp genus *Alpheus* inferred from nuclear and mitochondrial gene sequence data. Mol Phylogenet Evol 20(3):375–389

38. Williams DM, Jara J, Gomez E, Knowlton N (2002) The marine Indo-West Pacific break: contrasting the resolving power of mitochondrial and nuclear genes. Integr Comp Biol 42(5):941–952

39. Zuber M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31(13):3406–3415