HAL open science

Widespread occurence of mariner transposons in coastal crabs

Q-T Bui, N Casse, V Leignel, V Nicolas, B Chénais

To cite this version:

Q-T Bui, N Casse, V Leignel, V Nicolas, B Chénais. Widespread occurence of mariner transposons in coastal crabs. Molecular Phylogenetics and Evolution, 2008, 47 (3), pp.1181-1189. 10.1016/j.ympev.2008.03.029 . hal-04412414

HAL Id: hal-04412414

https://hal.inrae.fr/hal-04412414v1

Submitted on 23 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Molecular Phylogenetics and Evolution 47 (2008) 1181–1189

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Widespread occurence of *mariner* transposons in coastal crabs

Q.-T. Bui [a], N. Casse [a,*], V. Leignel [a], V. Nicolas [b], B. Chénais [a,*]

[a] Université du Maine, EA3265, Laboratoire de Biologie et Génétique Evolutive, UFR Sciences et Techniques, Avenue Olivier Messiaen, F-72085 Le Mans cedex 9, France
[b] Muséum National d'Histoire Naturelle, UMR 5202, USM 601, Département de Systématique et Evolution, F-75005 Paris, France

ARTICLE INFO

Article history:
Received 28 September 2007
Revised 12 March 2008
Accepted 17 March 2008
Available online 28 March 2008

Keywords:
16S phylogeny
Crab
Horizontal transfer
Transposable element

ABSTRACT

Mariner-like elements (*MLEs*) are ubiquitous DNA mobile elements found in almost all eukaryote genomes. Nevertheless most of the known copies are inactive and the question of the genome invasion by *MLEs* remains largely hypothetical. We have previously reported the presence of highly homologous copies of *MLEs* in the genome of phylogenetically distant crustacea living in the same hydrothermal environment suggesting the possibility of horizontal transfer. In order to further support the hypothesis that horizontal transmission of *MLEs* might occur between crustacean sympatric species, we described here 85 *MLE* sequences found in the genome of a large spectrum of coastal crab species. The number of the *MLEs* copies in genomes was variable. Half of these *MLEs* fit with the *irritans* subfamily of *MLEs* whereas the second half grouped in a new subfamily called *marmoratus*. In addition, a molecular phylogeny of crabs was established by using the 16S information. The comparison between 16S and *MLEs* based trees reveals their incongruence, and suggests either the existence of horizontal transfer events between phylogenetically distant species, or an ancestral *MLE* polymorphism followed by different evolution and stochastic loss.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Mariner-like elements (*MLEs*) belong to the *IS630/Tc1/maT/mariner* superfamily of transposons that move using a DNA-mediated "cut and paste" mechanism. *MLEs* are one of the simplest forms of autonomous DNA characterized by a single gene flanked by two short imperfect inverted terminal repeats (ITRs) (Jacobson et al., 1986; Hartl, 1989). The total length of *MLEs* is about 1300 base pairs (bp) and the unique intron-less gene, encoding a transposase of about 340–360 amino acid residues, is boarded by 5′ and 3′ untranslated regions (Plasterk et al., 1999).

The ubiquity of *MLEs* in the genome of continental organisms has been well documented (Robertson, 2002). Besides we have specifically studied *MLEs* in marine organisms and especially in crustacean from both hydrothermal and coastal origins (Casse et al., 2002, 2006; Halaimia-Toumi et al., 2004; Bui et al., 2007). To date, phylogenetic analyses of the *MLEs* sequence evolution have identified at least five *MLE* subfamilies with typically 40–56% identical amino acids: *mauritiana*, *cecropia*, *mellifera/capitata*, *irritans*, *elegans/briggsae* (Robertson, 2002). The discovery of *Tvmar1*, isolated from the protozoa *Trichomonas vaginalis*, seems to enlarge the number of *mariner* subfamily since *Tvmar* might constitute the first representative of a novel *MLE* subfamily (Silva et al., 2005).

Species phylogenies based on *MLEs* are often different from those based on the usual characters. This observation is generally interpreted in three distinct ways: horizontal transfer of TEs across boundary species, variability of evolution rates, and/or ancestral polymorphism and stochastic loss of TEs (Capy et al., 1994a,b; Lampe et al., 2003; Silva et al., 2004; Hua-Van et al., 2005). Analysis of *MLEs* isolated from some coastal and hydrothermal crustacean suggested recurrent *MLEs* horizontal transmission between hosts living in a same environment (Casse et al., 2006). Indeed *MLEs* with a high nucleotide identity have been detected in the crab *Bythograea thermydron* and in the amphipoda *Ventiella sulfuris* which shared the same hydrothermal environment (Casse et al., 2006). Beside, highly related *MLEs* have been also found in sympatric crustacean species, i.e. *Maja brachydactyla* and *Cancer pagurus* which are phylogenetically distant crab species (Casse et al., 2006). In order to further support the hypothesis that horizontal transmission of *MLEs* might occur between sympatric species, we described here *MLEs* found in the genome of a large spectrum of coastal crab species. Eighty-five (85) *MLEs* were isolated from 16 crab species that are representatives of 11 families. Each sequence was characterized as *MLEs* and placed in the phylogenetic tree of the *mariner* transposon family. Comparison with the phylogenetic data established by using the 16S information, support either the hypothesis of horizontal transfer between phylogenetically distant

* Corresponding authors. Fax: +33 2 43 83 39 17.
E-mail addresses: nathalie.casse@univ-lemans.fr (N. Casse), bchenais@univ-lemans.fr (B. Chénais).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.03.029

species, or the hypothesis of an ancestral *MLE* polymorphism followed by a distinct evolution and stochastic loss.

2. Material and methods

2.1. Sample collection and DNA extraction

Organisms studied are coastal crabs (Brachyura, Decapoda) belonging to 23 species representative of 15 families. As indicated in Table 1, samples were collected on the French (10 species), Vietnamese (11 species), Indonesian (Bali, 1 species) and New-Caledonia (1 species) coasts and conserved either in alcohol, or at −70 °C, or in liquid nitrogen until DNA extraction. The morphological identification of the species was carried out according to Apel and Spiridonov (1998), Falciai and Minervini (1992) and Stephenson and Campell (1959). Genomic DNA (gDNA) was purified by using standard procedure as previously described (Bui et al., 2007). The hydrothermal crab *Cyanagraea praedator* (Bythograeidae) was collected at the hydrothermal field 13°N on the East Pacific Rise (2620 m depth) and identified by S. Hourdez.

2.2. PCR amplifications

Amplification reactions were carried out in a 25-μl reaction volume including 100 ng of gDNA, 20 pmoles of primers, 2 mM deoxynucleotides triphosphate mixture, 1.5 mM $MgCl_2$, 2.5 μl of 10× reaction buffer (10 mM Tris, 50 mM KCl, pH 8) and 1 U of Taq DNA polymerase (Promega). The presence of *MLEs* in gDNA was investigated as previously described (Bui et al., 2007) by using, in a first attempt, the 5' ITR from the *Bytmar1* element (5′-TAC GAGGGGCGGTCAGAAAGTTATG-3′) with an annealing temperature of 55 °C and 30 amplification cycles. For species for which no amplification was detected, the presence of *MLEs* was checked using MAR primers (MAR-124F/MAR-276R and MAR-159F/MAR-249R) designed by Robertson (1993), Robertson (1997). The MAR primers allowed amplification of a well conserved region of *MLE* transposases. Primers used to amplify the 16S gene were defined by multiple alignments from crabs available in Genbank™ database (*Carcinus maenas*, AJ130811; *Chaceon affinis*, AF100914; *Panopeus herbstii*, AJ130815; *Sesarma* sp., AJ225891; *Thalassuca vocans*, Z79685; *Xantho poressa*, AJ130814). They were: *16Srep1* (5′-GT GCDAAGGTAGCATAATM-3′) and *16Srep2* (5′-CAACATCGAGTCRCA ADC-3′). Amplification was performed using the following PCR conditions: 92 °C, 30 s, 54 °C, 30 s, 72 °C, and 30 s for 36 cycles, and a final extension cycle for 10 min at 72 °C. The amplicon sizes were ranging from 361 bp (for *Perisesarma bidens*) to 378 bp (for *Myra subgranulata*) depending of the species.

2.3. Cloning and sequencing

Both *MLEs* and 16S PCR amplification products were eluted from 0.8% agarose gel using Wizard SV Gel® and PCR Clean-up System® (Promega) and cloned into the pGEM-T easy vector system I® plasmid (Promega). Then competent DH5α *Escherichia coli* were transformed and selected for the presence of recombinant plasmid based on ampicillin resistance and blue/white screening as previously described (Bui et al., 2007). Recombinant plasmids from white colonies were purified using Wizard plus miniprep kit® (Promega) before sequencing by the Cogenics-Genome Express company (http://www.gexbyweb.com/). For *MLEs* the size of the insert was previously checked by ApaI/PvuI (QBiogen) digestion and only plasmids containing an insert of approximately 1300 bp long were sequenced. For the 16S, the insert-containing colonies were confirmed by PCR using T7prom/SP6 universal primers. The selected 16S clones were sequenced using the same primers (Cogenics-Genome Express).

2.4. MLEs denomination

Each element was denominated according to Robertson and Asplund (1996) with a slight modification, using the three first letters of the genus name and the first letter of the species name followed by "mar" for mariner and the clone number, for example: *Carmmar1.1* for *Carcinus maenas* mariner clone 1.1. For each species

Table 1
Quantity of *MLEs* isolated in each Brachyura families and species studied, and accession numbers of *MLEs* and 16S sequences

Brachyura family	Species	Collected from....[a]	Number of *MLEs* obtained	Accession number for *MLE* and 16S
Ateleyclidae	*Atelecyclus undecimdentatus*	F	5	AM90691–AM906095 (AM946018)[b]
Cancridae	*Cancer pagurus*	F	6	AJ507245–AJ507250 (AM946018)
Dorippidae	*Dorippe quadridens*	V	0	n.a.[c]
Dromiidae	*Dromia personata*	F	0	n.a.
Eriphiidae	*Eriphia verrucosa*	F	5	AM906106–AM906110 (AM946021)
Goneplacidae	*Eucrate alcocki*	V	0	n.a.
Grapsidae	*Pachygapsus marmoratus*	F	9	AM231069–AM231077 (AM946022)
Homolidae	*Paromola bathyalis*	NC	5	AM906116–AM906120 (AM946029)
Leucosiidae	*Myra subgranulata*	V	5	AM906111–AM906115 (AM946027)
Parthenopidae	*Parthenope sinensis*	V	0	n.a.
Pilumnidae	*Pilumnus hirtellus*	F	5	AM906121–AM906125 (AM946023)
Pinnotheridae	*Pinnotheres pisum*	F	5	AM906126–AM906130 (AM946024)
Portunidae	*Carcinus maenas*	F	5	AM906096–AM906100 (AM410532)
	Necora puber	F	0	n.a.
	Portunus gracilimanus	V	0	n.a.
	Portunus granulatus	V	5	AM906131–AM906135 (AM410525)
	Portunus pelagicus	V	5	AM906136–AM906140 (AM410529)
	Portunus sanguinolentus	V	0	n.a.
	Portunus trituberculatus	V	5	AM906141–AM906145 (AM410530)
	Thalamita poissonii	V	5	AM906151–AM906155 (AM410536)
Sesarmidae	*Perisesarma bidens*	I	5	AM906146–AM906150 (AM946025)
Xanthidae	*Demania scaberrima*	V	5	AM906101–AM906105 (AM946026)
	Xantho poressa	F	5	AM906156–AM906160 (AM946020)
Total of *MLEs*			85	

[a] For each studied species, the collected site is indicated (i.e. F, France; V, Viet Nam; NC, New Caledonia; I, Indonesia).
[b] Accession Numbers for *MLEs*; accession numbers for 16S sequences are indicated within brackets.
[c] n.a., not accurate.

a consensus sequence (labelled with the "cons" suffix) was established with *MLEs* sharing 80–100% nucleotide homology. When within species homology was less than 50% between each *MLE*, two consensus sequences were built. When two consensuses were built for the same species there were referred as cons1 and cons2 (*Carmmarcons1* and *Carmmarcons2* for example).

2.5. Southern blot

The distribution of *MLEs* in the host genomes was checked by Southern blot as previously described (Bui et al., 2007). In order to estimate the number of *MLEs* present in the genome, for each species the consensus sequence was used to define an enzyme that cuts outside the element (see legend of Fig. 1). Several *MLEs* probes were used as indicated in the Fig. 1 legend. Labeling was performed using the random Nonaprimer Kit (QBiogen) and [α^{32}P] dATP (Perkin–Elmer Life Science). The digested gDNA was migrated in a 0.8% agarose gel, and transferred to a positively charged Nylon membrane (Hybond-N, Amersham). The membrane was hybridized overnight at 65 °C in 0.5 M Na_2HPO_4–NaH_2PO_4 pH 7.4, 7% sodium dodecyl sulfate (SDS) with the radiolabeled probe. The membrane was then washed at 65 °C for 30 min in 2× SSC–0.1% SDS (1× SSC = 0.15 M NaCl, 0.015 M trisodium citrate), for 1 h in 1× SSC–0.1% SDS, and finally for 30 min in 0.2× SSC–0.1% SDS, after which membranes were exposed at least overnight to X-ray films with an intensifying screen at −80 °C.

2.6. Sequence analyses

The isolated sequences were identified using BLAST program (Altschul et al., 1990). All sequences were submitted to EMBL database and accession numbers are indicated in Table 1. Motif signatures of *MLEs* were searched for using software from the Pôle Bioinformatique Lyonnais (http://pbil.univ-lyon1.fr/). The complete protein sequences of transposase genes from several elements of the *Tc1/maT/mariner* superfamily were downloaded from GenBank. Nucleotidic and proteic sequences were aligned using CLUSTAL W (Thompson et al., 1994) with default parameters.

The computer program Modeltest 3.04 (Posada and Crandall, 1998) was used to evaluate the fit of 56 nested models of nucleotide substitution to the data. Likelihood-ratio tests and the Akaike Information Criterion are commonly used to evaluate which model best fitted the data. We used the Akaike information criterion, as it was recently shown to have several advantages over likelihood-ratio tests (Sullivan and Joyce, 2005). The modeltest program revealed that the best model chosen for *MLEs* was the transversional substitution model with gamma correction (TVM+G). However, the TVM model is a special case of the GTR model and is not yet implemented in NJ analyses and MrBayes. Therefore, the GTR model (GTR+G) was used in the analyses. The best model for 16S sequences was HKY+I+G (Hasegawa–Kishino–Yano model with a γ-distributed rate variation across sites) and a proportion of invariables sites model was chosen as the best fitted model into the analysis of the 16S phylogenetic relationships.

Phylogenetic analyses were performed using neighbor-joining (NJ), maximum parsimony (MP), maximum likelihood (ML) with PAUP software v4.0b10 (Swofford, 2002) and Bayesian inference (BI) with MrBayes v3.0b4 (Ronquist and Huelsenbeck, 2003). The phylogenetic relationships among the crab *MLEs* were investigated from either the nucleic consensus sequences or their *in silico* translation into protein sequences, while the phylogenetic relationships among eubrachyura were investigated from the 16S nucleic sequences. NJ, ML and MCMC (Markov chain Monte Carlo) analyses were conducted with the model selected by Modeltest. Parsimony analyses were performed using the tree-bisection-reconnection (TBR) branch swapping option with 10 random addition replicates. For the BI analysis, we used three heated chains and a single cold chain in all MCMC analyses, and initiated runs with random trees. We conducted two independent MCMC runs with two million generations per run. We sampled trees (and parameters) every 100 generations. Stationarity was assessed by examining the average standard deviation of split frequencies. As the two runs converge onto the stationary distribution, we expect the average standard deviation of split frequencies to approach zero, reflecting the fact that the two tree samples become increasingly similar (Ronquist & Huelsenbeck, 2005). Moreover, the Potential Scale Reduction Factor (PSRF) should approach one as runs converge. For each run, the first 25% of sampled trees were discarded as burnin.

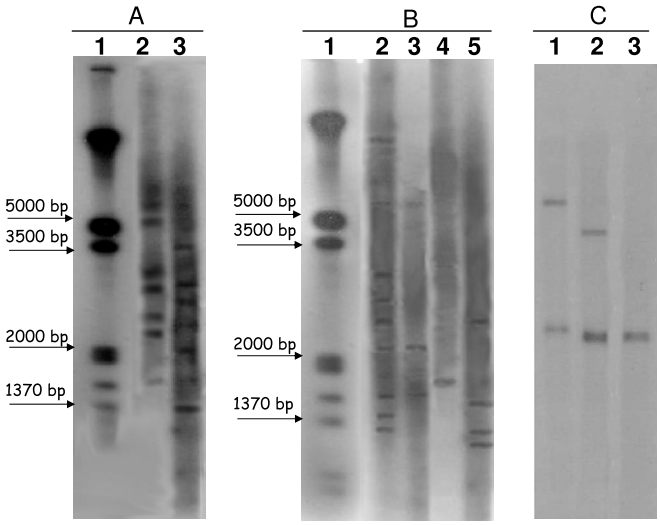

Fig. 1. *MLE* abundance in host genomes. (A) About 7 μg of gDNA from *Atelecyclus undecimdentatus* and *Portunus granulatus* (lanes 2 and 3, respectively) were digested using ApaI before Southern blot hybridization with a radiolabelled *Porgmar1.1* probe. Exposure time was 72 h at −80 °C. Lane 1: lambda HindIII/EcoRI molecular weight marker. (B) Same as A with gDNA from *Carcinus maenas*, *Portunus pelagicus*, *Portunus trituberculatus*, and *Perisesarma bidens* (lanes 2, 3, 4, 5, respectively) hybridized with a *Porpmar1.4* probe. (C) About 7 μg of gDNA from *Necora puber* were digested with either EcoRI (lane 1), Pst1 (lane 2) or both EcoRI and Pst1 (lane 3). Hybridization was done with a *Porpmar1.4* probe and the exposure time was one week at −80 °C.

3. Results

3.1. Characterisation of MLEs in coastal crabs

The presence of *MLEs* was searched by PCR using *Bytmar* ITRs derived primers and two Robertsons's primers (Robertson, 1993; Robertson, 1997) in 23 crab species recovering 14 families of brachyura coastal crabs (Table 1). A PCR fragment of about 1300 bp was found in 16 of the 23 crab species using *Bytmar* ITR (Table 1). In addition these 16 PCR products hybridized with a typical *MLE* probe, therefore confirming their *MLE* homology (data not shown). By contrast, in 7 species no amplification was observed with either *Bytmar* ITR or the two sets of MAR primers. The PCR products from the 16 *MLE*-positive species were cloned and sequenced, and were found to match with *mariner* elements by comparison with known *MLEs* sequence available in databases. Altogether, 85 complete *MLE* sequences were found in 16 crab species (Table 1). The sequence sizes were comprised between 1236 bp for *Portmar*1.2 and 1527 bp for *Carmmar*1.1 (mean size: 1313 ± 30 bp) including ITRs and UTRs (Table S1 in Supplementary data). The shortest sequence displays a deletion of about 100 bp in the 5′UTR and the longest displays a insertion of about 210 bp in

the 3′UTR. This insertion was not a repetition of part of the element and could not be identified as a known sequence in databases. The mean size of ITRs was 33 ± 3 bp, the mean sizes of UTRs were 154 ± 6 bp and 46 ± 14 bp for 5′ and 3′UTR, respectively. All the 85 sequences contained a transposase gene (mean size: 1051 ± 16 bp) but most of them display 1–11 stop codons and 2–4 frameshifts. Nevertheless 4 sequences displayed a complete open reading frame, 2 from *Pachygrapsus marmoratus* (Grapsidae), one from *Portunus granulatus* and one from *Portunus pelagicus* (Portunidae). The *Pacmmar* element (from *P. marmoratus*) has been recently studied in detail (Bui et al., 2007). It must be noted that *Pacmmar* (clone 1.3) and *Porgmar* (clone 1.5) display 99% and 100% identity at nucleic acid and amino acids levels, respectively. For each species a consensus sequence was established with *MLEs* sharing 80%–100% nucleotide homology. In seven species the homology was less than 50% between each *MLE* and then two consensus sequences were built (see Supplementary data S1).

The *in silico* translation of nucleic sequences indicated that the hallmarks of *mariner* transposases were found in most of the sequences even if slight modifications could be observed. For example the WVPHEL motif (Robertson, 1993) was found unchanged or slightly modified to WVPR(N/K)L or W(I/V)PHLL, and the YSPDLAP motif was conserved or changed to YSPELAP or YSPDLAS (see Supplementary data S2). The catalytic triad DD34D, characteristic of the *mariner* family of transposable element, was found in 65% of the sequences.

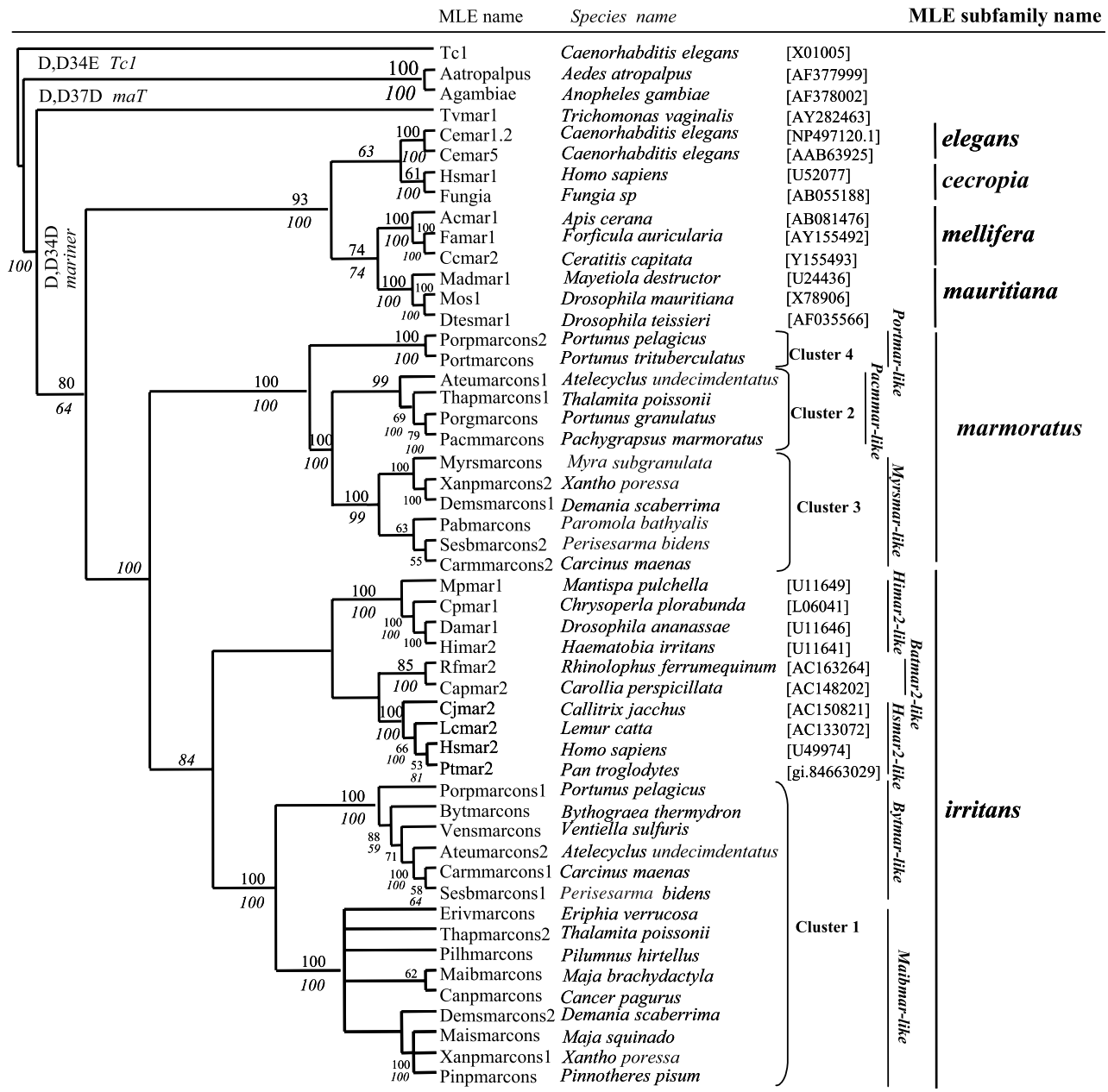

Fig. 2. Phylogenetic relationships of *MLEs* isolated herein and *mariner* sequences collected in databanks. Trees have been built using the transposase protein sequences, and only the consensus tree from the 12 most parsimonious trees was shown. *Mariner* sequences representatives of each *MLE* family have been used to build the tree established with PAUP software. Bootstrap values based on 1000 replications were indicated when above 50% using normal characters (MP node values) and italicized (NJ node values). The *MLEs* isolated from hydrothermal crustacea *Bythograea thermydron* (i.e. Bytmar), *Ventiella sulfuris* (i.e. Vensmar), and the coastal crab *Maja brachydactyla* (i.e. Maibmarcons) have been published in Casse et al. (2006). The *MLEs* isolated from the coastal crab *Maja squinado* (i.e. Maismarcons) are from personal unpublished results. All the *MLE* consensus proteic sequences used here have been supplied in Supplementary data S3.

Table 2
Pairwise divergence matrix between amino acid consensus sequences[a]

		Cluster 1												Cluster 2				Cluster 3						Cluster 4	
		Pilhmarcons	Maismarcons	Pinpmarcons	Xanpmarcons1	Canpmarcons	Maibmarcons	Demsmarcons2	Thapmarcons2	Carmmarcons1	Sesbmarcons1	Ateumarcons2	Porpmarcons1	Porgmarcons	Pacmmarcons	Thapmarcons1	Ateumarcons1	Demsmarcons1	Myrsmarcons	Carmmarcons2	Sesbmarcons2	Pabmarcons	Xanpmarcons2	Porpmarcons2	Portmarcons
Cluster 1	Erivmarcons	100	92	92	92	92	94	87	88	71	71	70	77	31	31	28	26	22	23	23	23	23	23	33	28
	Pilhmarcons		92	92	92	92	94	87	88	71	71	70	77	31	31	28	26	22	23	23	23	23	23	33	28
	Maismarcons			100	100	86	87	82	82	65	65	65	72	28	28	25	24	21	22	22	22	22	22	30	26
	Pinpmarcons				100	86	87	82	82	65	65	65	72	28	28	25	24	21	22	22	22	22	22	30	26
	Xanpmarcons1					86	87	82	82	65	65	65	72	28	28	25	24	21	22	22	22	22	22	30	26
	Canpmarcons						91	81	83	65	65	65	72	30	30	27	26	22	23	23	22	23	23	32	28
	Maibmarcons							86	83	66	66	66	73	30	30	28	25	21	22	22	22	22	22	32	28
	Demsmarcons2								77	62	62	61	68	27	27	24	22	18	19	19	19	19	19	30	26
	Thapmarcons2									62	62	62	68	29	29	26	24	21	22	22	21	22	22	30	26
	Carmmarcons1										99	99	84	28	28	26	23	22	23	22	22	22	22	29	25
	Sesbmarcons1											99	83	29	29	26	23	22	23	23	22	23	22	29	25
	Ateumarcons2												84	28	28	26	23	22	23	22	22	22	22	29	25
	Porpmarcons1													31	31	28	25	24	25	25	25	25	24	31	27
Cluster 2	Porgmarcons														100	90	77	65	66	68	69	69	65	47	43
	Pacmmarcons															90	77	65	66	68	69	69	65	47	43
	Thapmarcons1																71	64	65	67	67	67	63	43	38
	Ateumarcons1																	60	61	56	57	57	53	39	36
Cluster 3	Demsmarcons1																		97	88	88	88	82	34	32
	Myrsmarcons																			90	90	90	84	35	32
	Carmmarcons2																				99	98	92	35	32
	Sesbmarcons2																					98	93	35	32
	Pabmarcons																						93	35	32
	Xanpmarcons2																							33	30
Cluster 4	Porpmarcons2																								89

[a] Identities are given in percentage. Sequences have been aligned using CLUSTAL W software, and then alignments were transferred in GENEDOC software to obtain the identity percentage. Indicated clusters referred to the phylogenetic tree shown in Fig. 2. The *MLE* consensus sequences are supplied in the Supplementary data S3, and their names correspond to species as follows: Erivmarcons, *Eriphia verrucosa*; Pilhmarcons, *Pilumnus hirtellus*; Maismarcons, *Maja squinado*; Pinpmarcons, *Pinnotheres pisum*; Xanpmarcons, *Xantho poressa*; Canpmarcons, *Cancer pagurus*; Maibmarcons, *Maja brachydactyla*; Demsmarcons, *Demania scaberrima*; Thapmarcons, *Thalamita possoinii*; Carmmarcons, *Carcinus maenas*; Sesmarcons, *Perisesarma bidens*; Ateumarcons, *Atelecyclus undecimdentatus*; Porpmarcons, *Portunus pelagicus*; Porgmarcons, *Portunus granulatus*; Pacmmar, *Pachygrapsus marmoratus*; Myrsmarcons, *Myra subgranulata*.

3.2. Distribution of MLEs in host genomes

The distribution of these *MLEs* in their host genome was investigated by Southern blot and various results were obtained depending of the species. The presence of several bands or a smear indicate a very high number of *MLE* copy in the genome like in *Atelecyclus undecimdentatus*, *C. maenas*, *P. granulatus*, *P. pelagicus*, *Portunus trituberculatus*, and *P. bidens* (Fig. 1A and B). The large abundance of *MLEs* in *P. marmoratus* genome was previously reported (Bui et al., 2007). By contrast only one or two bands were observed in the genome of *C. pagurus* and *Necora puber* suggesting the presence of a limited number of *MLEs* (Fig. 1C). For *N. puber* no amplification has been obtained with the primer sets while a weak hybridization after one week exposition at −80 °C has been observed in Southern blot using the sequence of *Porpmar* as a probe. These observations suggest that *MLEs* detected in the genome of *N. puber* is probably quite different from the others crabs elements.

3.3. Phylogenetic position of crab MLEs in the mariner family

The topologies of the trees obtained with the NJ and MP methods from either nucleic or protein sequences were similar, and only differed for the most apical nodes (Fig. 2). The 23 consensus sequences were dispatched in four distinct clusters supported by high bootstrap values (98–100%). The percentage of identity between each cluster varies from 18 to 69% (Table 2). Inside each group these percentages vary between 62 and 99% in cluster 1, 71 and 90% in cluster 2, 82 and 98 in cluster 3 and reach 89% between the 2 sequences included in cluster 4 (Table 2).

Referring to the *mariner* family of transposable elements, at least five subfamilies have been described to date, i.e. *mauritiana*, *mellifera*/*capitata*, *irritans*, *cecropia*, and *elegans*/*briggsae* (Robertson, 2002). In order to place the above characterized crab *MLEs* in the *mariner* phylogenetic tree, several elements representative of the five subfamilies were aligned together with crab *MLEs* and with some *Tc1* and *maT* elements as external groups. The MP analysis generated 12 equally parsimonious trees and the resulting consensus tree display a similar typology than the tree obtained by NJ analysis (Fig. 2). The Fig. 2 obviously shows that the *MLEs* isolated from coastal crabs set in two different places in the *mariner* tree. First, the element of the cluster 1 are positioned in the *irritans* subfamily but separated in two clades. Some of these elements joined the previously described *Bytmar*-like lineage, which include *MLEs* from hydrothermal invertebrates (Halaimia-Toumi et al.,

2004; Casse et al., 2006). The others formed a new lineage supported by a bootstrap value of 100%, and which may be called *Maibmar*-like. Second, the crab *MLEs* from clusters 2, 3, and 4 did not joined any known subfamily and formed a new clade supported by a 100% bootstrap value (Fig. 2). The cluster 1 (lineages *Bytmar*-like and *Maibmar*-like) located in the *irritans* subfamily, displays 20–30% similarity at the amino acids level with the three other clusters. Clusters 2 and 3, corresponding to the lineages *Pacmmar*-like and *Myrsmar*-like, respectively, show more than 50% similarity between each other. The cluster 4 (lineage *Portmar*-like) is clearly distinct with 35% similarity with clusters 2 and 3. This suggests the existence of a new subfamily of *mariner* elements, which may be called *marmoratus*, where each cluster 2, 3, and 4 formed a new lineage, i.e. *Pacmmar*-like, *Myrsmar*-like, and *Portmar*-like, respectively (Fig. 2).

3.4. Comparison of 16S and MLE-based phylogenies of crabs

The topologies of phylogenetic trees obtained with the *MLEs* information from one hand and the 16S ribosomal RNA sequences from the other hand have been compared (Fig. 3A and B). The analysis of 16S information using NJ, ML, MP, and BI methods allowed the construction of a partial phylogenetic tree of eubrachyura. The results shown in Fig. 3A support the existence of the Cancroidae and Grapsoidae superfamilies. In addition, the Bythograeoidae superfamily (hydrothermal crabs) appeared to be the sister clade of the Xanthoidea. The Pinnotheroidea superfamily could be the sister clade of the Grapsoidea, but the bootstrap support of this node is low (Fig. 3A).

The *MLEs* based tree (Fig. 3B) indicated that several distant families of crabs could share elements that were very closed. For example, *MLEs* found in *P. granulatus* (Portunidae) are identical with *MLEs* of *P. marmoratus* at the amino-acid level (Grapsidae). A high homology was also found between *MLEs* from the Portunidae (*Thalamita poissonii*, *cons2*) and Cancridae (*C. pagurus*) families (99%), and between *MLEs* from the Xanthidae (*Demania scaberrima*) and Leucosiidae (*M. subgranulata*) families (97%). Reciprocally, various *MLEs* could be present in the same family of crab as seen for example in the portunids family in which *MLEs* from *P. granulatus* shared only 31%–47% of identity with *MLEs* from *P. pelagicus* and *P. trituberculatus*, respectively. In addition, one individual may display more than one kind of *MLEs* in its genome as shown in the

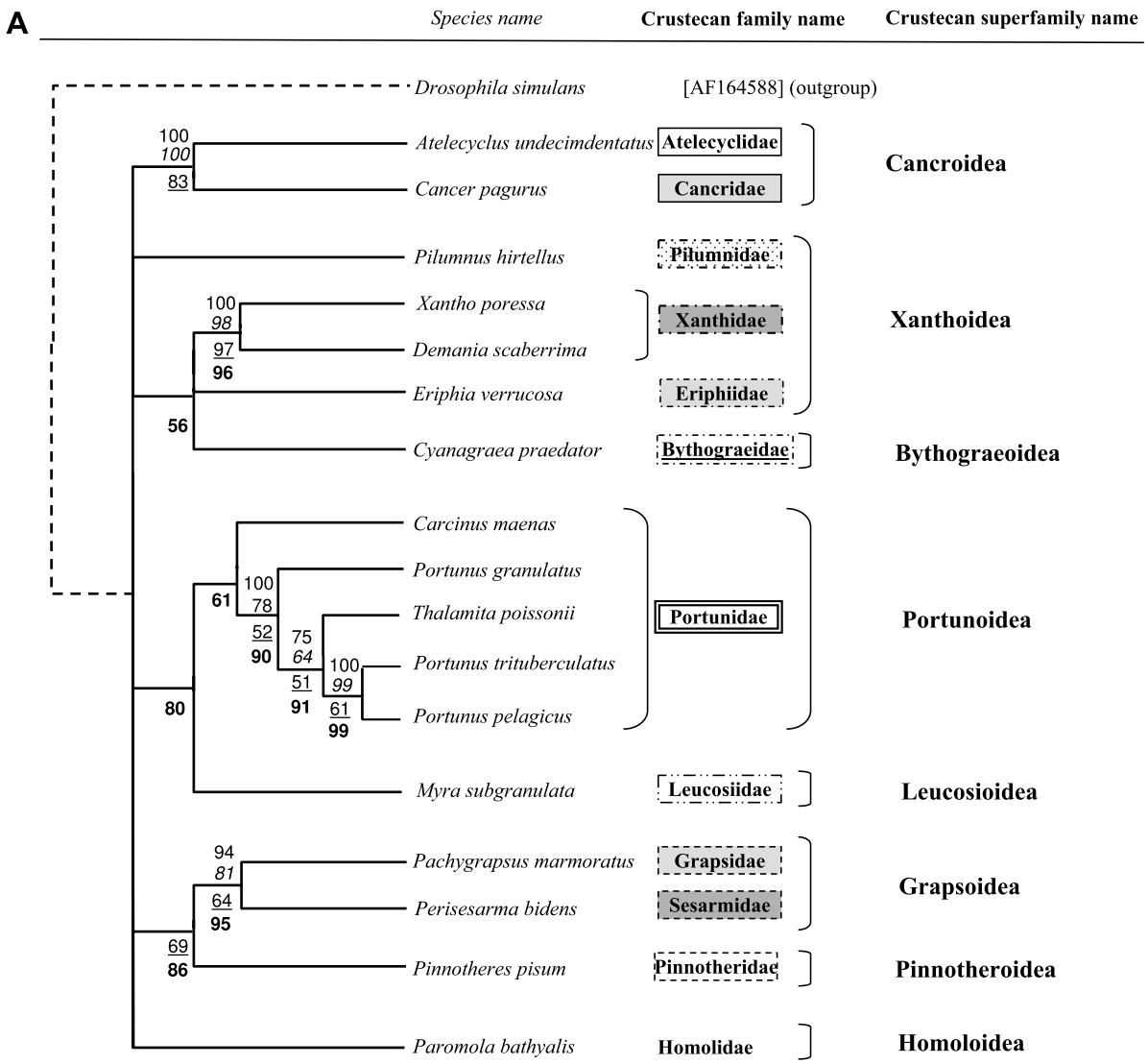

Fig. 3. Comparison of phylogenetic relationships of crabs using 16S sequences (A) and *MLEs* nucleic sequences (B). The trees were established using NJ, MP, and ML with PAUP software, and Bayesian inference (BI) methods. Bootstrap values above 50% are shown in normal characters (MP), italicized (NJ), underlined (ML) and bold (BI). All the *MLE* consensus nucleic sequences used to build the tree are supplied in Supplementary data S4. The consensus *MLE* sequence from the hydrothermal Bythgraeidae crab *Cyanagraea praedator* (i.e. Cyapmarcons) was built from unpublished sequences (Accession Nos. AJ507216–AJ507218); the 16S sequence of this crab was registered as AM946028.

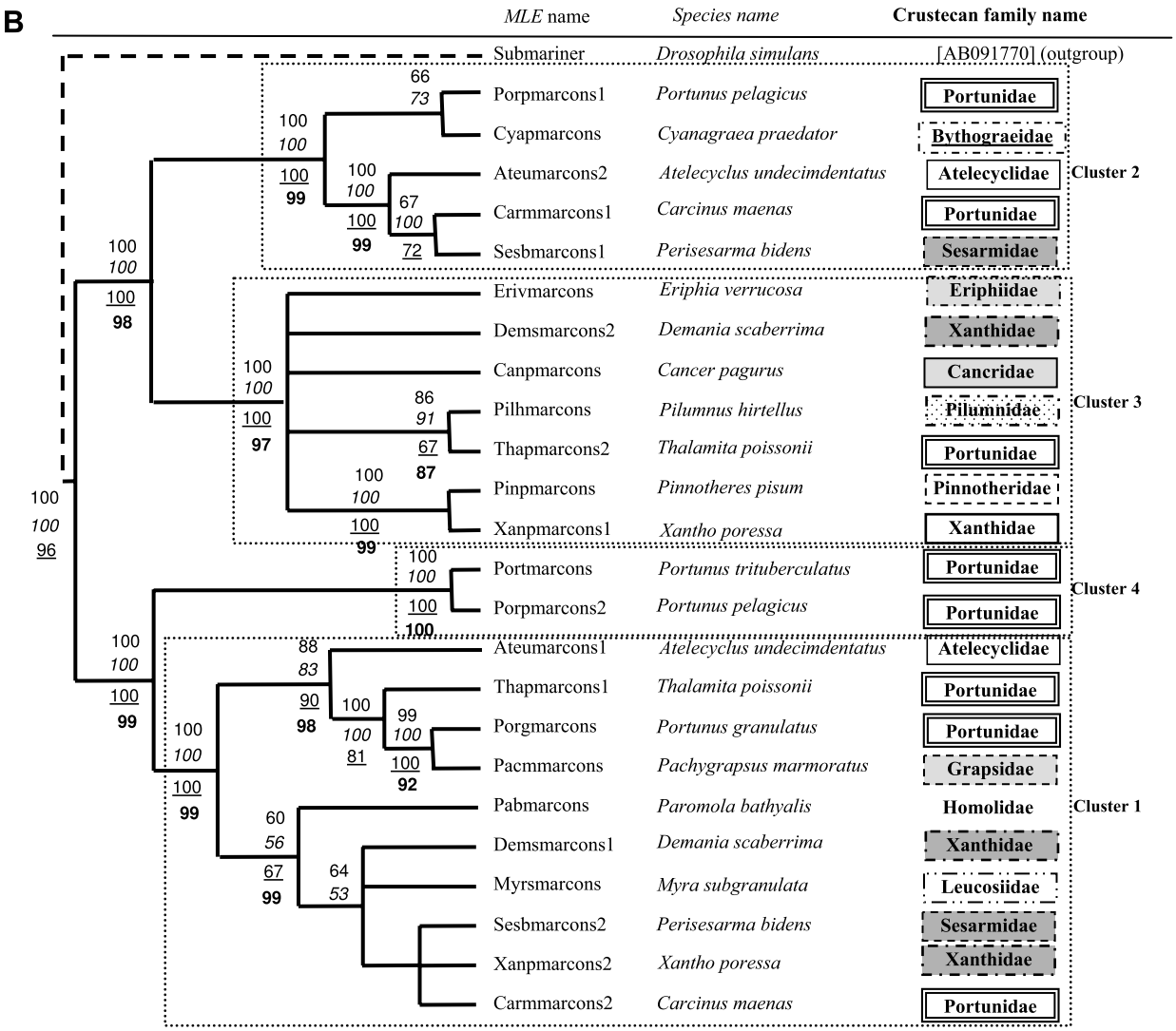

Fig. 3 (*continued*)

cases of *A. undecimdentatus*, *C. maenas*, *D. scaberrima*, *P. pelagicus*, *P. bidens*, *T. poissonii*, and *X. poressa* (Fig. 3B) that contain *MLE* representatives of *marmoratus* and *irritans* subfamilies.

Altogether these results obviously indicate the incongruence of *MLEs*-based and 16S-based phylogenies of crabs, and suggest either the hypothesis of horizontal transfer between phylogenetically distant species of crab, or the existence of an ancestral polymorphism followed by distinct evolution and stochastic loss of *MLEs*.

4. Discussion

Beyond to enlarge the list of *MLEs*, particularly those from marine invertebrates, the present study analysed the *MLEs* distribution in a wide range of coastal crab species. This study showed that most of the species studied (16 out of 23) displayed *MLEs* in their genome. Although both PCR amplification and Southern hybridization were used, it should be noted that the absence of *MLEs* in seven species may results for a default of detection of quite divergent elements in some genomes. The detected *MLEs* displayed the classical features of *MLEs* including the size, ITRs sequence and transposase hallmarks. Four sequences with uninterrupted ORF were found, one in *P. granulatus*, one in *P. pelagicus* and two in *P. marmoratus*. The *Pacmmar* elements from *P. marmoratus* were recently described in details (Bui et al., 2007). The *Porgmar* and *Pacmmar* sequences are 100% identical at the amino acids level. The *Porpmar* sequence possesses more than 80% identity in amino acids with the *Bytmar1* discovered in the genome of the hydrothermal crab *B. thermydron* (Halaimia-Toumi et al., 2004). The translated *Porpmar* sequence does not generated an N-flag in the DNA binding region of the transposase as observed for *Bytmar1* sequences. In the literature, only eleven *MLEs* with uninterrupted ORF have been reported (Jacobson et al., 1986; Garcia-Fernandez et al., 1995; Gomulski et al., 1997, 2001; Leroy et al., 2003; Barry et al., 2004; Halaimia-Toumi et al., 2004; Silva et al., 2005; Casse et al., 2006; Ren et al., 2006; Bui et al., 2007) among which only *Mos1* and *Famar1* were demonstrated to be naturally active elements (Medhora et al., 1991; Lampe et al., 1999; Barry et al., 2004). By contrast, *Himar1*, a consensus established by the nucleotide majority rule (Lampe et al., 1996) and *Hsmar-Ra*, a sequence obtained from the reconstructed ancestral form (Miskey et al., 2007), displays a complete ORF encoding an active transposase. The functional activity of the *Pacmmar* transposases is under investigation in our laboratory. With less than 35% amino acids identity, *Pacmmar* and *Porpmar* transposases have quite distinct sequences, and then it might be interesting to investigate also the putative activity of the *Porpmar* element isolated from *P. pelagicus*.

Almost half of the new elements described here (above referred as cluster 1) fit with the *irritans* subfamily of *MLEs* as previously

reported for *MLEs* found in marine organisms (Casse et al., 2002, 2006; Halaimia-Toumi et al., 2004). However, whereas some of them were grouped together with the *Bytmar*-like lineage according to previous report (Casse et al., 2006), the others appeared to create a new lineage (*Maibmar*-like) within the *irritans* subfamily of *MLEs*. Moreover, the second half of *MLEs* characterized in this study (above referred as clusters 2, 3, and 4) are grouped in a new branch which may be distinct from the *irritans* subfamily and constituted a novel *mariner* subfamily called *marmoratus*. Despite the *Pacmmar* element has been previously described as a new lineage in the *irritans* subfamily (Bui et al., 2007), in the present study numerous *MLEs* of coastal crabs integrated the *Pacmmar* group which then rise as a subfamily, i.e. *marmoratus* subfamily.

Although partial, the 16S-based phylogeny of Brachyura reported here is in agreement with the previous analyses based on morphologic and molecular characters. The phylogenetic tree confirmed that Grapsidae/Sesarmidae and Atelecyclidae/Cancridae constitute superfamilies (i.e. Grapsoidea and Cancroidea, respectively) as reported by Martin and Davis (2001). Compared to the 16S-based phylogeny, the phylogenetic tree constructed with the *MLEs* information was obviously incongruent and two additional arguments were found suggesting horizontal transfer events, i.e. presence of identical *MLEs* in distant species (e.g. *P. marmoratus* (Grapsidae) and *P. granulatus* (Portunidae), and sporadic distribution of *MLEs* in closely related species. Concerning *mariner* elements nine cases of horizontal transfer have been described in the literature (reviewed in Silva et al., 2004) supporting their widespread distribution. Nevertheless, the demonstration of horizontal transfer event is quite difficult and at least three main alternative hypotheses may explain incongruence of the 16S and TE-based phylogenies. First, phylogenetic reconstruction artefact may introduces mistakes or incertitude in the phylogeny, and this is particularly important when studied sequences have distinct speed of evolution, which is the case between *MLEs* and 16S. Second, ancestral duplication and/or polymorphism may create paralogous copies of *MLEs* which may have a distinct evolution in different genomes. Third, the host genome is able to repress the TE activity (Prud'homme et al., 1995; Hartl et al., 1997), and at the ultimate step TE would be eliminated from the population by stochastic loss or vertical extinction. Then *MLE* copy may be deleted in a species genome but not in another even if they are phylogenetically related, and this may generates a sporadic *MLEs* distribution. Therefore, duplication/deletion events associated with distinct speed of evolution may also explain incongruence of the phylogenies based on *MLEs* or 16S genes.

In conclusion, our results showed the abundance of *MLEs* in crab genomes and suggested that horizontal transfer events and/or ancestral polymorphism and stochastic loss may be responsible for both the observed intraspecific diversities and interspecific homology of *MLEs*. In order to reinforce one of these hypotheses, further analyses would be necessary such as the determination of dn/ds ratios of host genes and *MLEs* as described by Silva and Kidwell (2000) for *Drosophila*. Unfortunately crustacean genomes are poorly documented and a reference gene could not be collected from databases for all the species we studied here. Then its remains impossible to compare the divergence of host genes and *MLE* sequences in crustacean species. Horizontal transfer or ancestral polymorphism, one can hope that the increased number of sequencing genomes may help to answer to this question.

Acknowledgments

This work was supported by a PhD grant from Agence Universitaire de la Francophonie to Q.-T.B. The authors are grateful to Monika Gosh for english revision of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.03.029.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.-W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403.

Apel, M., Spiridonov, V.A., 1998. Taxonomy and zoogreography of the portunid crabs (Crustacea: Decapoda: Brachyura: Portunidae) of the Arabian Gulf and adjacent waters. Fauna Arabia 17, 159–331.

Barry, E.G., Witherspoon, D.J., Lampe, D.J., 2004. A bacterial genetic screen identifies functional coding sequences of the insect *mariner* transposable element *Famar1* amplified from the genome of the Earwig, *Forficula auricularia*. Genetics 166, 823–833.

Bui, Q.T., Delaurière, L., Casse, N., Nicolas, V., Laulier, M., Chénais, B., 2007. Molecular characterization and phylogenetic position of a new *mariner*-like element in the coastal crab, *Pachygrapsus marmoratus*. Gene 396, 248–256.

Capy, P., Anxolabéhère, D., Langin, T., 1994a. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends Genet. 10, 7–12.

Capy, P., Langin, T., Bigot, Y., Brunet, F., Daboussi, M.J., Périquet, G., David, J.R., Hartl, D.L., 1994b. Horizontal transmission versus ancient origin: *mariner* in the witness box. Genetica 93, 161–170.

Casse, N., Pradier, E., Demattei, M.V., Bigot, Y., Laulier, M., 2002. *Mariner* transposons are genetic parasites occurring in the genome of hydrothermal invertebrates. Cah. Biol. Mar. 43, 325–328.

Casse, N., Bui, Q.T., Nicolas, V., Renault, S., Bigot, Y., Laulier, M., 2006. Species sympatry and horizontal transfers of *Mariner* transposons in marine crustacean genomes. Mol. Phylogenet. Evol. 40, 609–619.

Falciai, L., Minervini, R., 1992. In: Delachaux, and Niestlé (Eds.), Guides des homards, crabes, langoustes, crevettes et autres crustacés décapodes d'Europe. Lausanne-Paris, pp. 287.

Garcia-Fernandez, J., Bayascas-Ramirez, J.R., Marfany, G., Munoz-Marmol, A.M., Casali, A., Baguna, J., Salo, E., 1995. High copy number of highly similar *mariner*-Like Transposons in planarian (Platyhelminthe): evidence for trans-phyla horizontal transfer. Mol. Biol. Evol. 12, 421–431.

Gomulski, L.M., Torti, C., Malacrida, A.R., Gasperi, G., 1997. *Ccmar1*, a full-length *mariner* element from the Mediterranean fruit fly, *Ceratitis capitata*. Insect Mol. Biol. 6, 241–253.

Gomulski, L.M., Torti, C., Bonizzoni, M., Moralli, D., Raimondi, E., Capy, P., Gasperi, G., Malacrida, A.R., 2001. A new basal subfamily of *mariner* element in *Ceratitis rosa* and other tephritid flies. J. Mol. Evol. 53, 597–606.

Halaimia-Toumi, N., Casse, N., Demattei, M.V., Renault, S., Pradier, E., Bigot, Y., Laulier, M., 2004. The GC-rich transposon *Bytmar1* from the deep-sea hydrothermal crab, *Bythograea thermydron*, may encode three transposase isoforms from a single ORF. J. Mol. Evol. 59, 747–760.

Hartl, D.L., 1989. Transposable element *mariner* in *Drosophila* species. In: Berg, D.E., Howe, M.M. (Eds.), Mobile DNA. American Society for Microbiology Press, Washington, DC, pp. 531–536.

Hartl, D.L., Lohe, A.R., Lozovskaya, E.R., 1997. Modern thoughts on an ancient mariner: function, evolution, regulation. Ann. Rev. Genet. 31, 337–358.

Hua-Van, A., Le Rouzic, A., Maisonhaute, C., Capy, P., 2005. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. Cytogenet. Genome Res. 110, 426–440.

Jacobson, J.W., Medhora, M.M., Hartl, D.L., 1986. Molecular structure of a somatically unstable transposable element in *Drosophila*. Proc. Natl. Acad. Sci. USA 83, 8684–8688.

Lampe, D.J., Churchill, M.E.A., Robertson, H.M., 1996. A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. EMBO J. 15, 5470–5479.

Lampe, D.J., Akerley, B.J., Runbin, E.J., Mekalanos, J.J., Robertson, H.M., 1999. Hyperactive transposase mutants of the *Himar1 mariner* transposon. Proc. Natl. Acad. Sci. USA 96, 11428–11443.

Lampe, D.J., Witherspoon, D.J., Soto-Adames, F.N., Robertson, H.M., 2003. Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. Mol. Biol. Evol. 20, 554–562.

Leroy, H., Castagnone-Sereno, P., Renault, S., Augé-Gouillou, C., Bigot, Y., Abad, P., 2003. Characterization of *Mcmar1*, a *mariner* -like element with large ITR from the phytoparasitic nematode *Meloidogyne chitwoodi*. Gene 304, 35–41.

Martin, J.W., Davis, G.E., 2001. An Updated Classification of the recent Crustacea, Natural History Museum of Los Angeles Country. Sciences series 39, 1–124.

Medhora, M., Maruyama, K., Hartl, D.L., 1991. Molecular and functional analysis of the *mariner* mutator element *Mos1* in *Drosophila*. Genetics 128, 311–318.

Miskey, C., Papp, B., Matés, L., Sinzelle, L., Keller, H., Izvak, Z., Ivics, Z., 2007. The ancient *Mariner* sails again: transposition of the human *Hsmar1* element by a reconstructed transposase and activities of the SETMAR Protein on transposon ends. Mol. Cell. Biol. 27, 4586–4600.

Plasterk, R.H., Izsvak, Z., Ivics, Z., 1999. Resident aliens the *Tc1/mariner* superfamily of transposable elements. Trends Genet. 15, 326–332.

Posada, D., Crandall, K., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., Bucheton, A., 1995. Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. Genetics 139, 697–711.

Ren, X., Park, Y., Miller, T.A., 2006. Intact *mariner* -like element in tobacco budworm *Heliothis virescens* (Lepidoptera, Noctuidae). Insect Mol. Biol. 15, 743–748.

Robertson, H.M., 1993. The *mariner* transposable element is widespread in insects. Nature 362, 241–245.

Robertson, H.M., Asplund, M.L., 1996. *Bmmar1*: a basal lineage of the mariner family of transposable elements in the silkworm moth, *Bombyx mori*. Insect Biochem. Mol. Biol. 26, 945–954.

Robertson, H.M., 1997. Multiple *mariner* transposons in flatworms and hydras are related to those of insects. J. Hered. 88, 195–201.

Robertson, H.M., 2002. Evolution of DNA transposons in eukaryotes. In: Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A.M. (Eds.), Mobile DNA II. American Society for Microbiology Press, Washington DC, pp. 1093–1110.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylo-genetic inference under mixed models. Bioinformatics 19, 1572–1574.

Silva, J.C., Kidwell, M.G., 2000. Horizontal transfer and selection in the evolution of P elements. Mol. Biol. Evol. 17 (10), 1542–1557.

Silva, J.C., Loreto, E.L., Clark, J.B., 2004. Factors that affect the horizontal transfer of transposable elements. Curr. Issues Mol. Biol. 6, 57–72.

Silva, J.C., Bastida, F., Bidwell, S.L., Johnson, P.J., Carlton, J.M., 2005. A potentially functional *mariner* transposable element in the protist *Trichomonas vaginalis*. Mol. Biol. Evol. 22, 126–134.

Stephenson, W., Campell, B., 1959. The Australian portunids (Crustacea: Portunidae). Aust. J. Mar. Freschwater Res. 10, 84–124.

Sullivan, J., Joyce, P., 2005. Model selection in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 36, 445–466.

Swofford, D., 2002. PAUP (Phylogenetic Analysis Using Parsimony). Sinauer Associates, Sunderland Massachusetts, USA.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.