

## Sequences of mitochondrial DNA suggest that *Echinocardium cordatum* is a complex of several sympatric or hybridizing species: A pilot study

A. Chenuil & J.-P Féral

*Observatoire Océanologique, Banyuls-sur-mer, France*

**ABSTRACT:** Sequence data from the 16S ribosomal gene of the mitochondrial genome were obtained for 10 individuals of the irregular sea urchin *Echinocardium cordatum* (Pennant 1777) [Loveniidae, Spatangoida], with one individual of *Echinocardium flavescens* (O.F. Müller, 1776) as an outgroup, plus one individual of *Abatus cordatus* (Verrill, 1876) [Schizasteridae, Spatangoida] as a more distant outgroup. Eighteen additional individuals were analysed for their *Taa I* restriction digestion profile at the same locus. This preliminary data set gave unexpected results. Eight different *E. cordatum* haplotypes were found and their phylogenetic relationships were established. They form two well supported monophyletic groups, clades A and B. Haplotypes of Glade A were found only in the samples from the Atlantic ocean, whereas haplotypes of Glade B were found in both Mediterranean and Atlantic locations. A and B haplotypes are found sympatrically in an Atlantic sample. Clade B can be divided in Glade B1 (two haplotypes) and Glade B2 (four haplotypes). Using fossil evidence within the genus *Echinocardium*, we inferred that more than six million years separate clades A and B, and more than two million years separate clades B 1 and B2. These large divergence times, the absence of haplotypes displaying nucleotide sequences intermediate between these groups of haplotypes, and the reduced divergence within each Glade suggest that they belong to different taxa, which were probably geographically separated in the past. Sequence data or genotypes from nuclear markers are necessary to establish if these putative pairs of taxa are now hybridizing or sympatric species.

### 1 INTRODUCTION

*Echinocardium cordatum* is an irregular sea urchin, which lives buried in the sediment (mostly fine sand to muddy sand and mud). Its reproduction results in a planctotrophic larva, potentially allowing high dispersal, therefore the species is expected to display a weak geographic structure of gene frequencies. Its distribution is antitropical. Previous genetic surveys gave interesting, though preliminary, results. With allozymes, (Féral *et al.*, 1995) observed a strong genetic differentiation between Atlantic and Mediterranean sites, no differentiation between Atlantic sites; in the same paper, phylogenetic trees inferred from 28S ribosomal DNA sequences produced a very robust cluster containing the individuals from two Atlantic sites (Channel and French Brittany) and another one, equally robust, with individuals from the Catalonian coast (Mediterranean) and individuals from the Galician coast (Atlantic). In addition to these contradicting results, Féral *et al.* (1998) observed, very close to Mediterranean populations with conventional allozymic alleles, one population with completely diagnostic alleles at one locus.

A phylogeographic study therefore appeared necessary to establish specific limits within the *E. cordatum* cluster. Since its mutation rate is high, and because it does not recombine, and also because its coalescence time is reduced relative to nuclear DNA, mitochondrial DNA is ideal to characterize geographic variation within species, and to distinguish evolutionary lineages which were separate in the past but are now sympatric and may hybridise. It is also suitable to distinguish species that diverged very recently. We therefore started sequencing the 16S mitochondrial DNA of several individuals sampled in the Atlantic Ocean and Mediterranean Sea.

### 2 MATERIALS AND METHODS 2.1

#### *Biological materials*

Sampling sites were located at the Channel/North Sea limit (North of France), along the Galician coast (North-West of Spain), in the Catalonian coast (Gulf of Lion), in the French Riviera (Ligurian Sea), and in the Aegean Sea near Athens (Greece) (Figure 1).

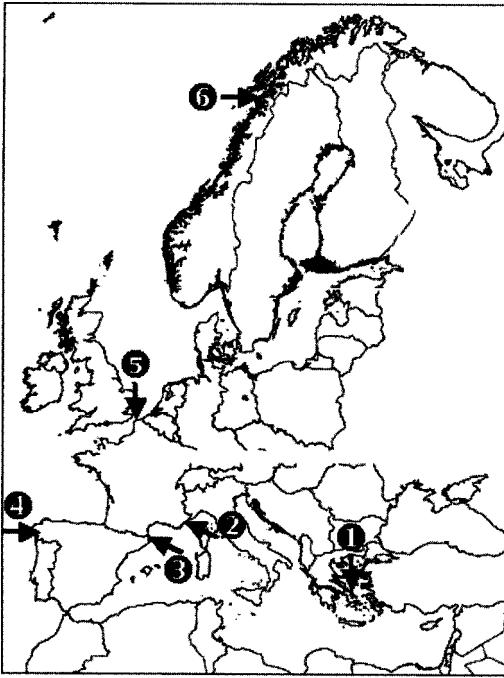


Figure 1. Sampling sites of *Echinocardium cordatum* and *E. flavescens* - 1/Aegean Sea (near Athens, GR), 2/Ligurian Sea (Villefranche-sur-mer, FR), 3/Gulf of Lion (Banyuls-sur-mer, FR), Coast of Galicia (Camarinas, SP), 5/North Sea (Wimereux, FR), 6/Norwegian Sea (Bode, NO).

One individual of *Echinocardium flavescens* from Norway (Bods/Valosen) was taken as a close out-group, and one individual of *Abatus cordatus* (also a Spatangoid irregular sea urchin, sampled from Kerguelen Islands) as a more distant outgroup.

### 2.2 Materials and molecular techniques

Some samples had been conserved in ethanol for up to five years before DNA was extracted, for others, DNA was extracted from fresh tissue, or tissue conserved 2 months in a solution composed of 20% DMSO, 0.25 M EDTA, NaCl saturated and equilibrated at pH 8. Several protocols were used for DNA extraction: CTAB extraction with phenol/chloroform (Winnepenninckx *et al.*, 1993), Nucleon Phytopure DNA extraction (APB society), or a simpler technique using the Chelex 100 chelating resin (Sigma society). For Chelex 100 extractions, tissue was put in 500  $\mu$ l of a 10% suspension of Chelex 100 in water with proteinase and kept at 60°C overnight, then boiled for 10 minutes, frozen at -20°C. Care was taken to avoid incorporation of chelating resin beads which could interfere with subsequent reactions. It

seems that a cycle of freezing/thawing allowed better amplification of newly extracted DNA, whatever the extraction technique. In all cases, less than 1 mm<sup>3</sup> of tissue was put into the extraction buffer with proteinase K (1 to 4  $\mu$ l of proteinase K at 20mg/ml) for overnight digestion at 60°C. Even when complete dissolution failed, amplification of the 16S mtDNA fragment succeeded. For CTAB and Nucleon extractions, DNA pellets were suspended in 100  $\mu$ l of sterile nanopure water. When 2.5 p.l of this DNA solution was not detectable in agarose minigels stained with ethidium bromide or SYBR Gold (Molecular Probes society), or was extremely degraded, it was used without further dilution in the polymerase chain reaction (PCR). Otherwise, DNA samples were diluted 10 times in water before PCR. For Chelex 100 extracts, 1  $\mu$ l of supernatant was taken without dilution. PCR were performed in 20  $\mu$ l using 1  $\mu$ l of the DNA solution with 0.15  $\mu$ l of Qbiataq 5 u/ $\mu$ l (Qbiogen society), 2  $\mu$ l of the buffer supplied by manufacturer giving final concentrations of 3.5 mM MgCl<sub>2</sub> and 0.15 mg/ml BSA, 3.3  $\mu$ l of 2 mM of each dNTP (more than conventionally, as recommended by Qbiogen), and 20pmol of each primer, 16SF2 (5'GTT TCG CCT GTT TAC CAA AAA CAT) and 16SR (5'CGA ACA GAC CAA CCC TTA AAA GC) (Palumbi, 1996). One cycle of 5 minutes at 95°C was followed by 40 cycles of 40 seconds at 94°C, 30 seconds at 65°C and 2 minutes at 72°C, on a PCT200 thermocycler. All templates gave unique and intense bands of 592 by approximately containing 400 ng to 1.6  $\mu$ g DNA per 20  $\mu$ l. After first sequences were obtained, in order to rapidly and economically distinguish one haplotype from the others, we performed a restriction digestion by 0.5 units of *Taal* (MBI Fermentas) of 5  $\mu$ l of crude PCR product in a total of 10 p.l in 1X of the supplied buffer (Y+). After the addition of 2-31d of bromophenol blue, the product was visualized in a 2% or 3% agarose minigel containing ethidium bromide run 30 to 40 minutes at 100V Controls with no DNA template were used for each PCR reaction. We never observed an amplification product for those negative controls although at present several hundreds of PCR have been performed for this locus, using several dozens of negative controls. After using 3 p.l of the PCR for a control minigel, the remaining PCR product was sequenced on automated sequencers with the primer 16SR and, for five samples also by the 16SF2 primer, by a private company.

### 2.3 Data analysis methods

Before analysing haplotypes within *Echinocardium cordatum*, we checked that the Spatangoida used in this study (*Abatus plus Echinocardium* individuals) clustered into a monophyletic group, and that all *Echinocardium* individuals also formed a monophyletic

unit, using regular sea urchins as outgroups (Genbank references: AJ 130798 for *Hemicentrotus pulcherrimus*, NC\_001770 for *Arbacia lixula*, NC\_001572 for *Paracentrotus lividus*, NC\_001453 for *Strongylocentrotus purpuratus*). DNA sequences were aligned using Bioedit (Hall, 1999), and phylogenetic reconstructions used the PHYLIP software (Felsenstein, 1993). Neighbour Joining (NJ) with Kimura's distance and a transition/transversion ratio of 2, parsimony and maximum likelihood methods (ML) were used to build phylogenetic trees. One thousand bootstrapped data sets were produced to estimate branch robustness for the first two methods. For ML trees, we give the estimates of the probability that the branch length equals zero, for information, but, as was emphasized by Felsenstein in the PHYLIP documentation, we are aware that this is not a reliable and conservative test.

### 3 RESULTS

#### 3.1 Sequence data

About 550 bases per PCR product were read. Sequencing in both directions gave complementary sequences. For two individuals respectively the 40 and 70 first nucleotides were missing. Eight different haplotypes were found among the 10 sequenced individuals of *Echinocardium cordatum* (Table 1, Figure 2).

#### 3.2 Sequence alignment

Alignment was straightforward within spatangoids. When regular sea urchins were included, some nucleotide positions had to be removed, and 425 sites were kept once positions corresponding to the primers were removed (data set "a"). Without regular sea urchins, but including partially sequenced individuals, 462 sites

remained (data set "b"); without partially sequenced individuals, 509 sites were kept (data set "c").

#### 3.3 Phylogenetic relationships among haplotypes (Figure 2, Table 2)

All distance methods gave the same topology.

- a) Using a regular sea urchin sequence from Genbank (*Hemicentrotus pulcherrimus*) as the most external outgroup (data set "a"), we obtained very clearcut phylogenetic relationships between different recognized species: *Abatus cordatus* branched off first, followed by *Echinocardium flavescens* (monophyly of the genus *Echinocardium*), and then by *E. cordatum* sequences. These nested monophyletic groups were supported by bootstrap values of 100% for parsimony and NJ trees, and the corresponding branch lengths of the ML tree were estimated significantly different from zero. Within *E. cordatum*, well-supported clades were revealed: Glade A (2 haplotypes), and Glade B (4 haplotypes). These two clades appear robust with all used methods (bootstrap values were of 100% for clade A and 95% for Glade B for both NJ and parsimony trees, and for ML trees branch lengths were estimated significantly different from zero). Similar conclusions were obtained with other species of regular sea urchins as outgroups.
- b) With the data set containing all sequenced haplotypes (data set "b": Note that the two haplotypes of Glade A have the same sequence, the sites for which they differed being removed from this data set.), relationships within Glade B could be established more precisely. Two haplotypes formed a well-supported Glade, named B1 (further details are given in Table 1). Another Glade, B2 (four more haplotypes of Glade B), was also found by all three methods, but was not strongly supported.

Table 1. Number of individuals analysed by restriction digestion (RFLP) or by sequencing of the 16S mt-DNA amplified fragment, for each sampling location (see figure 2 for clade definition). For sites, numbers are those of Figure 1.

	North sea	Atlantic	Gulf of Lion	French Riviera	Aegean sea
Sites	©	O	©	O	O
Sequence					
A	1	1			
B1		1		1	2
B2			4		
RFLP					
B2 pattern			8		7
Non-B2 pattern					1
Total	1	2	12	4	9

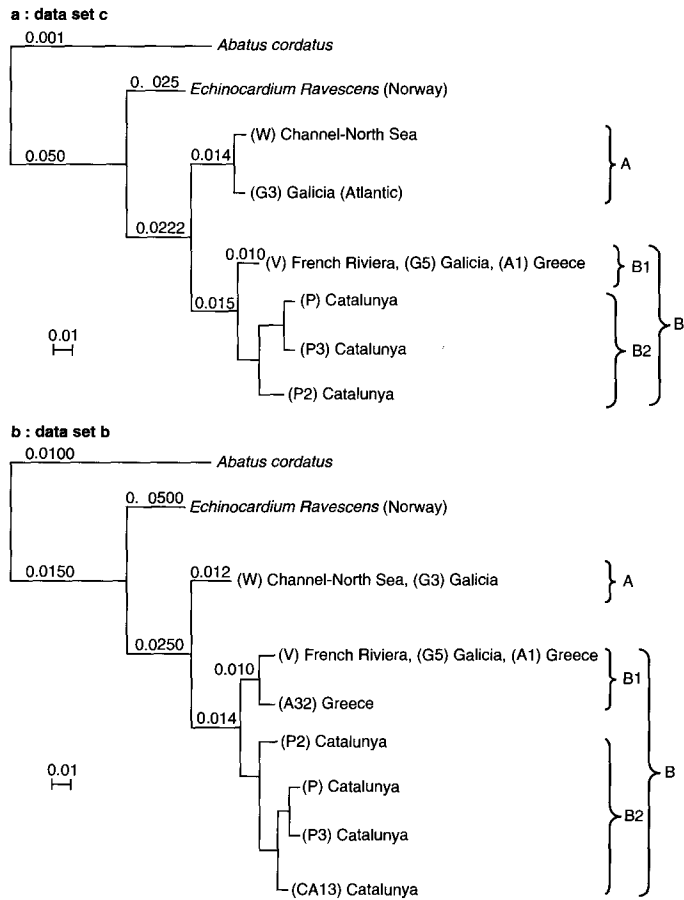


Figure 2. ML trees for data sets c (Fig. 2a) and data set b (Fig. 2b). Branch lengths are indicated for well-supported clades (bootstrap values are given in Table 2). The names of the individuals displaying each haplotype are given in parentheses followed by their sampling site.

Table 2. Results of phylogenetic analyses. For each data set, the % of bootstrap values supporting a given Clade is given, in line one for parsimony, and in line two for neighbour joining trees. In line three, we give, for the ML trees, the estimated levels of significance of the test of the hypothesis that branch length is null.

Data set	Nb of nucleotids	Clade <i>Echinocardium</i>	Clade <i>E. cordatum</i>	Clade A	Clade B	Clade B1	Clade B2
a	425	100	100	100	95	SS	NF
		100	100	100	95	SS	81
		**	**	**	**	**	NS
b	462	NR	100	SS	99	97	51
		NR	100	SS	98	99	60
		NR	**	**	**	**	NS
c	509	NR	100	96	99	SS	58
		NR	100	99	99	SS	82
		NR	**	**	**	**	NS

(\*\*;  $P < 0.01$ ). NS: Not significant. SS: A single sequence represented this clade for this data set. NF: Clade not found in the consensus tree.

c) Using the data set of larger sequence length (data set "c"), Glade B2 was supported more strongly than with shorter data sets, and clades that were already strongly supported were confirmed. It may be noted that there is only one most parsimonious tree, which has the same topology as the consensus trees from bootstrapped parsimony and neighbour joining trees and as the ML tree.

### 3.4 Divergence times between clearly identified clades of haplotypes

We know, by combining fossil data and phylogenetic data, that *Echinocardium flavescens* and *E. cordatum* diverged more than 12 million years (myrs) ago. Fossils morphologically similar to *E. mediterraneum* and *E. cordatum* were observed in a geological stratum of this age (Philippe, 1998), and geological and molecular analyses (Laurin *et al.*, 1994; David and Laurin, 1996) established that *E. flavescens* branched off before the split occurred between *E. mediterraneum* and *E. cordatum*. In our 16S sequences, *Es flavescens* and *E. cordatum* display from 33 to 35 differences among 509 sites. 16S mtDNA therefore evolves at less than 0.5% differences per myrs in *Echinocardium*. Haplotypes from Glade A and Glade B differ by 17 to 18 differences, and those from Glade B 1 and Glade B2, by 6 to 8 differences, leading to estimates of divergence times of 6.18 myrs between Glade A and B, and 2.47 myrs between clades B1 and B2. By contrast, there are only 2 differences between the two haplotypes of Glade A, only one between the two B 1 haplotypes and one between the three pairs of B2 haplotypes. The use of Kimura's distance - which takes into account possible multiple changes or reversions-, with a transition/transversion ratio of 2, slightly increased these estimates to 6.7 and 2.8 myrs.

### 3.5 Geographic distribution of haplotypes

Using restriction digestion of the PCR product with *Taal*, we determined whether or not an individual haplotype belonged to the B 1 Glade: the sequence of *Echinocardium flavescens* and clades A and B2 of *E. cordatum* correspond to one profile, whereas sequences from Glade B 1 produce a different profile. Eighteen additional individuals were therefore analysed by this technique in addition to the ten sequenced individuals. Haplotypes from Glade A are found in the Atlantic (*s. l.*) samples, but not in the Mediterranean (*s. l.*) samples analysed. Haplotypes from Glade B 1 are found in both seas, and haplotypes from Glade B2 are only found in samples from the Mediterranean sea (Table 1).

Haplotypes from clades A and B 1, and from Glade B2 and from another Glade (restriction digestion pattern "non-B2") are found in sympatry, respectively in the Galician coast (Atlantic), and in Greece (Figure 2).

## 4 DISCUSSION

### 4.1 Evolutionary origin of the different clades of haplotypes

Within the sample of individuals identified as *Echinocardium cordatum*, different clades were defined, which probably diverged more than 2.5 and 6.2 myrs ago. Theoretically, even such remote divergence times are not sufficient to conclude that different species are involved. Since mtDNA does not recombine, the life-time of a given neutral mutant only depends on the demographic history of the species. If the effective size of the species is large enough, haplotypes may persist a long time before being eliminated by genetic drift. An additional, noteworthy characteristic of the data set is the absence of intermediate haplotypes between those of clades A and B, and to a lesser extent, those of clades B 1 and B2. Any haplotype from Glade A differs by at least 17 nucleotides from any haplotype from Glade B, any pair of haplotypes from Glade B 1 and Glade B2 differs by at least 6 nucleotides, and within each Glade, pairs of haplotypes differ at most by two nucleotides. Two groups of hypotheses can account for this pattern. Under the first hypothesis, *E. cordatum* samples may belong to one single species, with a very large effective size (necessary to explain the very ancient most recent common ancestor), and which underwent severe bottleneck events (to explain the absence of intermediate haplotypes, which were lost by random genetic drift). The bottleneck(s) hypothesis is theoretically possible (one just has to imagine an *ad hoc* scenario) but such long coalescence times (6 myrs) as well as such large gaps between conspecific haplotypes (17 substitutions) seem unlikely within a species. Several studies using 16S or more rapidly evolving mitochondrial sequences were performed in echinoderms, and average within-species diversity never exceeded 0.84%, indeed were generally much lower, and haplotypes whose divergence approached 1 myr always belonged to different species, which were identified either by their morphology or by their allopatric distribution (Lessios *et al.*, 1999; Lessios *et al.*, 1998, Lessios *et al.*, 2001b; Lessios *et al.*, 2001a; Hrnicevich and Foltz 1996; Hrnicevich *et al.*, 2000).

Under the second hypothesis, a split occurred within the range of the ancestral *E. cordatum* species, one of the "populations" evolved "A" haplotypes, the other evolved "B" haplotypes (the same reasoning may apply for B 1 and B2 haplotypes). Under this vicariant-type hypothesis, within each of these newly isolated populations, "intermediate" haplotypes appeared, which at first were very close to the common ancestral haplotype, with time, more and more divergent haplotypes appeared (by mutation) and some more closely related and ancestral haplotypes disappeared (by genetic drift) till now (up to 17-18 differences). Within

each population, random genetic drift resulted in the coalescence of the presently observed haplotypes to an ostensible [?] common ancestral haplotype which seems relatively recent, compared to the 6 myrs of the common ancestor to "A" and "B" haplotypes, because within any one of clades A, B 1 and B2, there are only 0-2 differences.

#### 4.2 Geographic distribution of the haplotypes

A given haplotype can be found several thousand kilometres apart (the haplotype of the B 1 Glade on the coasts of Galicia, French Riviera, and Greece). This is not surprising because this species has a very high dispersal potential.

In two areas we found haplotypes from different clades in sympatry (A and B in Galicia, and Glade B2 and another Glade in Greece). The different taxa, which probably compose *Echinocardium cordatum*, are therefore sympatric. Today, these taxa may be hybridising. This hypothesis is supported by morphological data: (i) a morphological study allowed to distinguish a typical "*E. cordatum*" morphotype and a "*E. fenauxi* morphotype". 28S rDNA sequences were obtained on the same individuals, and phylogenetic trees from DNA sequences grouped individuals according to locality rather than morphology (Laurin *et al.*, 1994); (ii) morphological measures of individuals from different locations formed relatively well separated areas in a factorial plane, but some individuals had intermediate phenotypes, particularly those from Galicia (David *et al.*, 1999), the sampling area where clades A and B are found in sympatry.

Finally, we did not find haplotypes from Glade A or B1 on the Catalanian coast (12 individuals from three locations were analysed). If confirmed, the absence of other haplotypes than B2 from the Catalanian coast would suggest either a past endemism and a present isolation of this area from influx of *E. cordatum* larvae, or a competitive exclusion of other haplotypes by B2 haplotypes. This last hypothesis is contradicted by the fact that in Greece B2 haplotypes can coexist with other ones.

## 5 CONCLUSION AND FUTURE RESEARCH

Although morphological variation has already been reported within *Echinocardium cordatum*, it was still considered as a single species, probably because the disruption between morphological and molecular variation could have been a result of an environmental control of shape and also because intermediate morphotypes were observed. Our preliminary genetic study strongly suggests the presence of several taxa within *E. cordatum*. If we had used nuclear DNA sequences instead of mitochondrial DNA, we would

probably not have identified the B 1 and B2 clades because nuclear DNA has larger coalescence times (Palumbi *et al.*, 2001). Clades B1 and B2 would be more strongly separated if we used a more rapidly evolving sequence, such as COI (preliminary data confirm a more rapid evolutionary rate for COI). We plan to test the alternative hypothesis that disruption between morphology and genetics is due to hybridisation, by analysing the same individuals with 16S mtDNA plus unlinked genetic markers (we started sequencing the D8 region of the 28S rDNA and we are trying to find microsatellite codominant markers in this species). These individuals are also being characterised morphologically by our biometrician colleagues (B. David and coll.).

Much more numerous individuals (from well chosen areas) will be sequenced for their 16S mtDNA in order to compare the coalescence times of haplotypes from the different clades and to compare the demographic history of those clades. We will reconstruct the origin of these clades by also considering paleogeographical data.

However, the best test of the taxonomic status of the *Echinocardium cordatum* clades consists in performing controlled crosses between individuals from different mitochondrial clades. If, as preliminary data suggest, these entities are hybridising, the only evidence supporting their specific status would be the observation of a hybrid dysgenesis (reduced fertility or viability of hybrids).

## ACKNOWLEDGEMENTS

Thanks are due to all those who helped us in the field or collected the urchins for us (Norway: N. Hagen, B. David, B. Laurin. Galicia: B. David. Villefranche-sur-mer: B. Hesse. Wimereux: D. Davoult). We also thank S.v. Boletzky who improved the English text. This work was supported by a grant of IFB (*Institut Français de la Biodiversité*).

## REFERENCES

- David B & Laurin B (1996) Morphometrics and cladistics: measuring phylogeny in the sea urchin *Echinocardium*. *Evolution* **50**, 348-359.
- David B, Laurin B, Choné T & Magniez F (1998) Morphological disparity in the genus *Echinocardium* (Echinoidea: Spatangoida). In MD Candia Carnevali & F Bonasoro (eds), *Echinoderm Research 1998*. Pp. 253-259, Balkema, Rotterdam.
- Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
- Féral J-P, Poulin E, Derelle E, Gallardo S, Chambon C (1995) Genetic differentiation of *Echinocardium cordatum*

- as revealed by allozymes and RNA sequencing. In R Emson, AB Smith & AC Campbell (eds), *Echinoderm Research 1995*. Pp. 41-42, Balkema, Rotterdam.
- Féral J-P, Poulin E, Oubelkheir K (1996) Geographic and genetic differentiation of *Echinocardium cordatum* (Pennant) - The current state of a complex question: The status of *Echinocardium fenauxi* (Péquignat). In R Mooi & M Telford (eds). *Echinoderms: San Francisco*. Pp. 647-649, Balkema, Rotterdam.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**, 95-98.
- Hrinkevich AW, Axayacatl R-O, Foltz DW (2000) Phylogenetic analysis of molecular lineages in a species-rich subgenus of sea stars (*Leptasterias* subgenus *Hexasterias*). *Amer Zool.* **40**, 365-374.
- Hrinkevich AW, Foltz DW (1996) Mitochondrial DNA sequence variation in a sea star (*Leptasterias* spp.) species complex. *Molec. Phylogen. Evol.* **6**, 408-415.
- Laurin B, Féral J-P, David B, Derelle E (1994) Polytypism in the spatangoid sea urchin *Echinocardium*: A morphological vs molecular approach. In B David, A Guille, JP Féral & M Roux (eds). *Echinoderms through time*. Pp. 739-745, Balkema, Rotterdam.
- Lessios HL, Garrido MJ, Kessing BD (2001b) Demographic history of *Diadema antillarum*, a keystone herbivore on caribbean reefs. *Proc. R. Soc. Lond. B* **268**, 1-7.
- Lessios HA, Kessing BD, Pearse JS (2001a) Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution* **55**, 955-975.
- Lessios HA, Kessing BD, Robertson DR (1998) Massive gene flow across the world's most potent marine biogeographic barrier. *Proc. R. Soc. London - ser. B* **265**, 583-588.
- Lessios HA, Kessing BD, Robertson DR, Pauley G (1999) Phylogeography of the pantropical sea urchin *Euclidaris* in relation to land barriers and ocean currents. *Evolution* **53**, 806-817.
- Palumbi AR, Cipriano F, Hare MH (2001) Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution* **55**, 859-868.
- Palumbi SR (1996) *Nucleic Acids II: The polymerase reaction*. In DM Hillis, C Moritz, BK Mable (eds), *Molecular systematics*. Sinauer, Sunderland, Massachusetts.
- Philippe M (1998) *Les échinides miocènes du bassin du Rhône: révision systématique*. Nouvelles archives du Muséum d'Histoire Naturelle de Lyon. fascicule 36: 441 pp.
- Winnepenninckx B, Backeljau T, De Wachter R (1993) Extraction of high molecular weight DNA for molluscs. *Trends in Genetics* **9**, 407.