J Mol Evol (2002) 54:246–257
DOI: 10.1007/s00239-001-0006-0

JOURNAL OF MOLECULAR EVOLUTION

Exploring the Phylogenetic Utility of ITS Sequences for Animals: A Test Case for Abalone (*Haliotis*)

Annette W. Coleman,[1] Victor D. Vacquier[2]

[1] Division of Biology and Medicine, Brown University, Providence, RI 02912, USA
[2] Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093-0202, USA

Received: 20 April 2001 / Accepted: 9 July 2001

Abstract. To evaluate the general utility of sequences of the nuclear rDNA internal transcribed spacer (ITS) regions for phylogenetic analyses of animal species groups and their broader relationships, sequences were obtained for 19 species of the genus *Haliotis* plus a keyhole limpet and a more distantly related gastropod, the Chilean abalone. Three subclades of *Haliotis* species appear consistently, each encompassing little variation. They are (A) the North Pacific species, (B) the European species, and (C) the Australia species. The one Caribbean species examined clearly groups with the North Pacific clade, not the European clade. *H. midae* (South Africa) and *H. diversicolor supertexta* (Taiwan) both diverge basal to the European and Australian species groups in the phylogenetic trees. Sequence comparisons showed that one species of *Haliotis, H. iris* from New Zealand, is quite distant from the remaining *Haliotis* species, almost as much as the more obvious outgroup, the keyhole limpet, an observation common to other DNA sequence analyses of these taxa. Using the rate of nucleotide change calculated from the sister Caribbean-Pacific pair, the length of the *H. iris* long branch is compatible with the suggestion that its ancestry became isolated on New Zealand at Gondwandan breakup. Use of ITS permits a totally independent estimate of the phylogenetic relationships, yet branching order was very similar to that established using other DNA regions studied previously, including those under strong positive selection. Knowledge of the RNA transcript secondary structure is particularly useful in the optimal alignment of more distantly related taxa. The RNA transcript secondary structure of *Haliotis* ITS2 shows conservation of features found also in ITS2 of angiosperms and algal taxa. Since ITS, particularly ITS2, is not saturated with nucleotide changes even at the family level, it should be useful for phylogenetic reconstruction of animal groups, not just at the species and genus levels but perhaps also for families and above.

Key words: Abalone — Biological species — *Haliotis* — Internal transcribed spacer — ITS — Phylogeny — Secondary structure

Introduction

DNA-based phylogenetic reconstructions of animal genera and species, with conspicuous exceptions (e.g., Schlotterer et al. 1994), have relied heavily on sequences of genes in nuclear rDNA and mitochondrial DNA. Nuclear rDNA genes display relatively few variant nucleotide positions at the lowest taxonomic levels, while mtDNA can discriminate among populations. By contrast, those intergenic spacer regions of nuclear ribosomal genes transcribed into RNA (ITS1 and ITS2) have proven very useful in evaluating plant, fungal and algal lineages, from subspecies to family and even ordinal levels (Baldwin 1992). The ITS is relatively conserved in its

Correspondence to: A.W. Coleman; *email:* annette_coleman@brown.edu

evolution, presumably because its transcript secondary structure provides at least some of the signals that guide processing into small, large, and 5.8S ribosomal RNA (Hadjiolova et al. 1994; van Nues et al. 1995). The ability to predict transcript secondary structure has enhanced the value of ITS regions for phylogenetic studies since this insight is important to guiding sequence alignment at higher taxonomic levels (Mai and Coleman 1997).

Among metazoans, the genus *Haliotis* was chosen for this study for two reasons: first, to assess the usefulness of ITS sequences for reconstructing the phylogenetic history of an animal group and, second, to compare the ITS results with known hybridization data for the same species. There are about 55–60 recognized species of *Haliotis* in the world (Lindberg 1992; Geiger 1998, 1999). Geiger (1999) has summarized the wealth of available information on morphology, distribution, phylogeny studies, and the fossil record. Also, the genus has a considerable literature concerning gamete affinities and cross-fertilization among species, an aspect of species biology particularly interesting to compare to a phylogenetic reconstruction based on ITS sequences. Furthermore, *Haliotis* is the only animal genus for which DNA sequences are available from multiple species for two gene products (one on the sperm and the second on the egg) that interact at a major step in gamete recognition at fertilization (Swanson and Vacquier 1998). Thus, it would be possible to assess the degree of congruence of phylogenetic estimates based on ITS with those based on gene regions coding for proteins involved in gamete interaction, where positive selection for coevolution has been reported.

Species of the genus *Haliotis* are united by many morphological characters, probably the most distinctive synapomorphy being the arc of perforations in the shell. However, after an exhaustive reinvestigation of morphological characters in *Haliotis* species, Geiger (1999) concluded that the radula, epipodium, and hypobrachial gland structures are all useful to identify species but not valuable for phylogenetic reconstruction of the evolutionary history in the genus at our current state of knowledge. He further notes that shell characters such as shape, size, and color are relatively plastic, and not phylogenetically useful. Hence, molecular evidence assumes an important role in phylogenetic reconstruction of species in this genus.

The ITS analysis has revealed three strongly supported subclades of *Haliotis,* each geographically discrete. The phylogenetic reconstruction is in almost complete agreement with previous molecular analyses as well as interfertility studies. The recognizable conservation of ITS RNA transcript secondary structural features, even in the relatively distantly related gastropod compared here, *Concholepas,* suggests that the shared secondary structure features might be very useful to help in ITS alignment above the genus level when further mollusc sequences are available.

Materials and Methods

The protocol to obtain Internal Transcribed Spacer (ITS) sequences was similar to that utilized previously (Mai and Coleman 1997; Coleman et al. 2001). DNA or tissue (Table 1) was from the same animals used by Lee and Vacquier (1995); in addition, DNA extracts from tissue of a specimen each of *H. pourtalesii* (specimen 152707 of the Los Angeles County Natural History Museum), *H. roberti,* and the keyhole limpet (*Megathura crenulata*), and from a juvenile of the Chilean gastropod *Concholepas concholepas* were included for comparisons. DNA was extracted by grinding tissue in InstaGene Matrix (BioRad Laboratories, Hercules, CA) and used as template for polymerase chain reactions (PCR) using the following primers:

- **Primers for entire ITS-1-5.8S-ITS-2:**
  "G-FOR" (5′-GGGATCCGTTTCCGTAGGTGAACCTGC), a sequence in the small subunit rDNA similar to the ITS1 primer of White *et al.* (1990), and
  "G-REV" (5′-GGGATCCATATGCTTAAGTTCAGCGGGT-3′) a sequence in the large subunit rDNA used by Mai and Coleman (1997).

  Three primers specific for *Haliotis* were:
  Hal-FOR (5′TGAACCTGCGGAAGGATCATTAACG-3′) overlapping G-FOR but extending internally, and
  Hal-REV (5′AGTCTCGTCTGATCTGAGGTC-3′) adjacent but entirely internal to G-REV primer.
  ABA-FOR (5′CTGCGGAAGGATCATTAACGATATC-3′) is a forward primer extending five nucleotide positions yet more internal to Hal-FOR.
- **Primers located in 5.8S gene:**
  The major primers were the reverse complementary pair ITS2 and ITS3 of White *et al.* (1990).
  A-FOR (5′-TCGATGAAGAGCGCAGCC-3′) was an additional primer for the 5.8S region that includes one more nucleotide 3′ to the White *et al.* (1990) ITS2 primer.

PCR mixtures utilized 1.25 mM $MgCl_2$ and a PCR protocol as in Mai and Coleman (1997). Taq polymerase was added only after reaction mixtures had reached the initial 95°C heating step. Bands of appropriate size were gel purified and either used as template for further PCR reactions or ligated into pT7 Blue vector (Novagen, Madison, WI), used to transform *E. coli,* and plasmid with insert was isolated using the Wizard Plus Miniprep System (Promega, Madison, WI). Individual minipreps were assessed before sequencing both for size of insert and presence/absence of certain diagnostic restriction endonuclease cut sites found in the 5.8S region of *Haliotis, Xba*1 near the 5′ end, and *Sma*1 near the 3′ end.

Sequences were obtained for both strands of the ITS and 5.8S rDNA regions using primers in the vector and the ABI Prism 377 DNA Sequencer and Big Dye Chemistry protocol. Sequences were aligned by eye using MacVector and AssemblyLIGN software (Kodak, International Biotechnologies, Inc. New Haven, CT) and guided by ITS1 and ITS2 transcript secondary structure derived by methods described in Mai and Coleman (1997). Transcript folding utilizing the website www//mfold2.wustl.edu/~mfold/rna/form1.cgi implementing **mfold,** version 3.0 (Zuker et al. 1999), using the default settings.

Phylogenetic analyses of the aligned sequences utilized PAUP* version 4.0b4a (Sinauer Assoc. Inc., Sunderland, MA) for cladistic analyses (maximum parsimony; heuristic and branch and bound options) and distance analyses (Kimura two-parameter matrix, neighbor-joining algorithm option with minimum evolution criterion). ITS1 and ITS2 were assessed both separately and together, to see if they were self-consistent, with all nucleotide positions weighted equally and gaps coded as missing data. Additional parsimony assessments utilized the variant of each gap equalling the fifth character. Where intraspecies variant sites were found, the appropriate IUPAC symbol was used. The repeatability of results was examined for both parsimony and distance trees by bootstrap repetitions. The *p* values represent pairwise mean

Table 1. Organisms sequenced

Organism	Trivial Name	Locale	GenBank No.
Class Gastropoda			
Order Vetigastropoda, Family Haliotidae			
Haliotis sorenseni	white	La Jolla, CA	AF296850
H. cracherodii	black	La Jolla, CA	AF296857
H. kamtschatkana	pinto	Friday Harbor, WA	AF296853
H. fulgens	green	La Jolla, CA	AF296859
H. rufescens #1	red	San Diego, CA	AF296855
H. rufescens #2	red	Mendocino, CA	AF296851
H. corrugata	pink	La Jolla, CA	AF296856
H. walallensis	flat	La Jolla, CA	AF296854
H. roberti		Cocos I, Costa Rica	AF306942
H. discus hannai		Japan	AF296858
H. pourtalesii		Louisiana coast	AF296871
H. tuberculata #1		Roscoff, France	AF296860
H. tuberculata #2		Naples, Italy	AF296861
H. tuberculata lamellosa		Greece	AF296862
H. roei		Australia	AF296866
H. scalaris		Australia	AF296864
H. cyclobates		Australia	AF296865
H. rubra		Australia	AF296867
H. midae		South Africa	AF296863
H. diversicolor supertexta		Taiwan	AF296868
H. iris		New Zealand	AF296869
Order Vetigastropoda, Family Fissurellidae			
Megathura crenulata	keyhole limpet	La Jolla, CA	AF296849
Order Caenogastropoda, Family Muricidae			
Concholepas concholepas	abalone, el loco	Puerto Montt, Chile	AF296870

character distances (adjusted for missing data) as calculated by PAUP*. Sequences have been deposited in GenBank (Table 1) and the alignment is available from TreeBASE as SN789-2181 (http://www.herbaria.harvard.edu/treebase/).

Results

The standard universal ITS primers do, in fact, amplify *Haliotis* ITS, but they also appear to be preferential for algal, fungal, and protozoan templates over metazoan ones, and among metazoans, preferential for flatworm templates versus molluscan. Five of the DNA extracts from alcohol-fixed testes yielded ITS sequences that most closely resemble Trematoda, among the GenBank sequences currently available. The only template DNAs successful on the first attempt in producing a single, correct, band from PCR were those where phenol-chlorofom DNA purification methods were used on samples of dissected internal tissue, or masses of sperm, apparently ensuring the predominence of *Haliotis* over foreign DNA, at appropriate high dilution. The Hal-FOR and Hal-REV primers then gave single PCR bands of the correct DNA. The ABA-FOR primer is specific for *Haliotis* species and did not work on keyhole limpet, nor even on *H. iris*, both of which differ in sequence from the remaining *Haliotis* species at these positions.

Because no ITS or 5.8S sequences from *Haliotis* were available for comparison, the *Haliotis* sequences were identified by their similarity to the 5.8S sequence of the snail *Arion rufus* (EMBO # X00131). All *Haliotis* 5.8S sequences were identical except for *H. diversicolor* (one substitution) and *H. iris* (three substitutions). The 5.8S of keyhole limpet differs from the majority *Haliotis* 5.8S by only one substitution. The 5.8S contained no useful information for phylogenetic reconstruction and was omitted from analyses.

Both the ITS1 and ITS2 of all these organisms proved similar in lengths and base composition to those typical of the vast majority of other eukaryote groups (Baldwin 1992; Torres et al. 1990). *Haliotis* species' ITS1 was 257–318 nucleotides in length and 51–57% GC; ITS2 was 289–303 nucleotides with 51–58% GC content. The keyhole limpet ITS1 was 280 nt and 52% GC, and ITS2 was 235 nt with 60% GC. The *Concholepas* ITS2 was 388 nt in length, an increase primarily due to two insertions with high GC content, producing an overall 62% GC content. Alignment of the *Haliotis* species was straightforward since there is great similarity among the species' sequences and little suggestion of saturation in the data. Only two nucleotide positions in the entire alignment have all four nucleotides represented, and these two positions are both in the unusual ITS1 insert (Table 2). About 5% of nucleotide positions in the alignment of all the *Haliotis* species have three nucleotides present.

Complete sequences were obtained for all but *H. ru-*

Table 2. ITS1 insert sequences

all N. Pacific spp.	.amwc>GGGGtGcAAGCgcgcGCTTCtCCCC<.....agtcg.
Med-N. Atlantic spp.	....
H. midae (So. Af.)	.agtw>GGGGttGAAGCgcgcGCTTCtCCCC<c.....gatcg.
H. scalaris (Aust)	.agta>GGGGttGAAGTgcgcGCTTCtCCCC<gaacgc....gatcg.
H. cyclobates (Aust.)	.agta>GGGGttGAAGTgcgcGCTTCtCCCC<gaacgctcgatcg.
H. roei (Aust.)	.agta>GGGGttGAAGTgcgcGCTTCtCCCC<gaacgctcgmtcg.
H. rubra (Aust.)	.agta>GGGattGAAGTgcgcGCTTCtcCCC<gaacgctcgatcg.
H. divers. (Taiwan)	.ttcg>GGGgttGAAGTgcgcGCTTCtaCCC<a.....
H. iris (New Z.)	.....>GTTAttTGGC<.....attg.

M = A or C, W = T or A, period = space inserted for alignment; carets mark 5′ and 3′ boundaries of stem-loop; upper case nucleotides form the pairings in the RNA transcript. "All N. Pacific" = subclade A. Med-N. Atlantic = subclade B.

bra (missing ITS2) and *Concholepas* (missing ITS1). Clones of ITS were sequenced from at least two separate PCR reactions from DNA extracts of a single animal of *H. cracherodii, H. kamtschatkana, H. rufescens, H. tuberculata, H. roei, H. midae, H. iris,* and the keyhole limpet. Variation in ITS1 of repeats from a single animal was 0–5 nucleotide substitutions and 0–3 indels; variation in ITS2 was 0–2 nucleotide substitutions and 0–2 indels. Variation seen between two individuals of the same species collected at different places (Table 1) was no greater. The two *H. tuburculata* sequences (from Naples and Roscoff) differed at only one site. The Mendocino sequence of *H. rufescens* revealed no variants not already found in the two San Diego sequences of this species.

Secondary Structure of the RNA Transcript of ITS1 and ITS2

Secondary structures of the RNA primary transcript were derived for all the sequences, and were essentially identical for the keyhole limpet and the genus *Haliotis*, for both ITS1 and ITS2. The proposed structure was supported by numerous single nucleotide changes that preserved pairing in the paired regions of helix. The secondary structures were primarily useful in helping align the keyhole limpet and *Concholepas* sequences with those of *Haliotis* species. They provide structure to help confirm the correctness of the alignment, since DNA regions without protein products lack the innate punctuation provided by codon triplets.

The ITS1 transcripts have a secondary structure of six stem-loops (termed 1, 2, 3, 4, 5, and 6), easily aligned among *Haliotis* species, except for numbers 3 and 6 of *H. iris* (data not shown, available upon request). In addition, there is a major indel, a 9–38 bp insertion/deletion forming an additional stem-loop between loops 2 and 3. This indel is the only region in the entire ITS that presented any question of the most appropriate alignment. The indel is entirely absent in the Mediterranean-Atlantic group, but present in all the other species (Table 2). The keyhole limpet has at this position four nearly identical repeats of 18 nucleotides in length.

The secondary structure of the ITS2 transcript, shown in Fig. 1, was shared by all the organisms, including *Concholepas.* It has four stem-loops (termed I, II, III, and IV), of which loop IV is the most variable. Both *H. iris* and the keyhole limpet, compared with the other species, show nucleotide changes in the mid region of stem III, affecting both sides of the helix but maintaining pairing. However, they share with the other *Haliotis* species identical pairings for the basal three pairings of loop I and the basal five pairings of II. The most highly conserved subsequence in the alignment, identical among all the sequences including keyhole limpet, is an 18 nucleotide stretch on the 5′ side of loop III near the tip (Fig. 1); the only variant is a single transition found in *H. iris* that preserves the pairing potential. The ITS2 of *Concholepas,* though its alignment with the others was essentially unusable for phylogenetic purposes, nevertheless had recognizable similarity of secondary structure and sequence in ITS2, despite two indels that add a 35 nucleotide extension to hairpin loop II and a 53 nucleotide hairpin loop emerging from hairpin loop III.

ITS2 transcript secondary structure has been studied extensively among other eukaryote groups, and proposed secondary structures are supported not just by single nucleotide changes in stems that preserve the capacity to pair, but also by Compensatory Base Changes (CBCs) found by comparison among groups of related species. CBCs are sites of compensatory nucleotide change involving both of the members of a pair in the double helix of an RNA transcript secondary structure (Gutell et al. 1994); they are considered initial proof of the proposed secondary structure, as applied so successfully to transcripts of ribosomal genes. The insert in Fig. 1 illustrates where the relatively conserved pairing positions are found in other organisms studied. Excepting *H. iris,* the *Haliotis* species have no CBCs in any of the paired positions of the ITS2 transcript helices I, II, and III, although there are a number of one-sided changes that compensate to maintain the helix pairing with the opposite strand. *H. iris,* by comparison with the other *Haliotis* species, has a multiplicity of compensatory base changes; for example it has two CBCs near the terminus of loop 1, one at the terminus of loop II and an extended

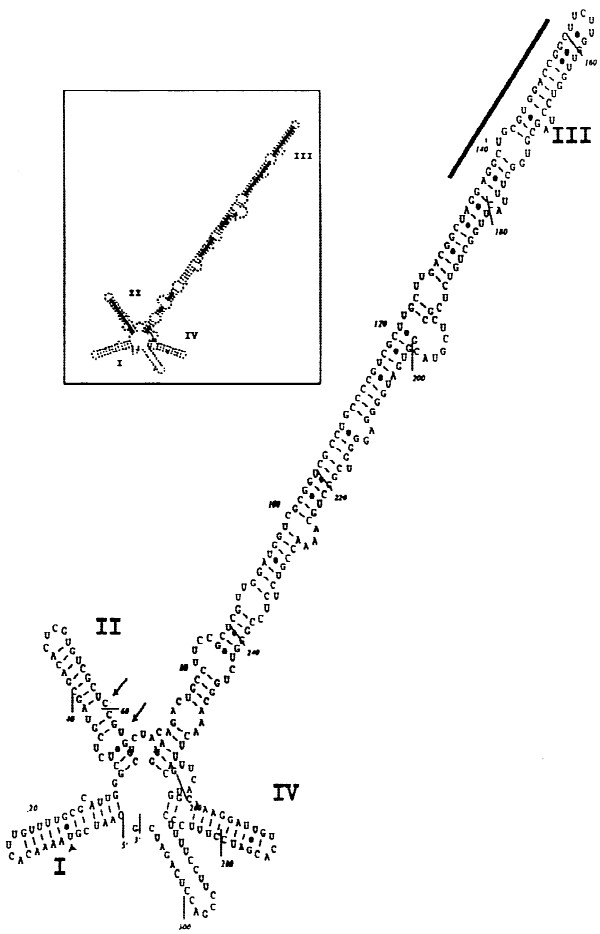

Fig. 1. Diagram of the secondary structure of the RNA transcript of ITS2 of *H. rufescens* (dG = 135.9). The four hairpin loops are labelled I, II, III, and IV. Pyrimidine-pyrimidine pairings in II (arrows) and the region of highest primary sequence conservation on the 5′ side of III (solid line) are designated. Arrowhead designates a nucleotide position mentioned in the text that alternates between C and U, always pairing with the G opposite. Inset: cartoon illustrating with heavy lines the regions of highly conserved pairings, compared with other phyla (Coleman 2000).

(30 ntp) region of difference in the middle of the stem of III. These strongly support the transcript folding for at least loops I, II, and III.

Phylogenetic Reconstructions by ITS Comparisons

The ITS1 plus ITS2 alignment totals 716 nucleotide positions, of which 431 are variable, but only 113 of these are parsimony informative. Searches using either the heuristic or the branch and bound option of parsimony, ignoring indels, all produce the same 435 most parsimonious trees of 610 steps for the alignment. The 50% majority rule cladogram shown in Fig. 2 is derived by parsimony analysis, and illustrates the characteristics common to the various analyses. The data here include all *Haliotis* species except *H. rubra* (omitted because the ITS2 is lacking), plus the keyhole limpet, and all sites in ITS1 and ITS2. Three major subclades consistently showing high bootstrap support are labelled A, B, and C, and represent, respectively, North Pacific (plus Caribbean), European (Mediterranean plus Eastern Atlantic), and Australian species. These same subclades are supported by all analyses, whether ITS1, ITS2, or both are included, whether polymorphic nucleotide positions are excluded or not, and whether the parsimony, or distance with minimum evolution, criteria are employed. Coding gaps as a fifth nucleotide also produces all the clades seen with greater than 80% bootstrap support in Figs. 2 and 3.

Of the additional three *Haliotis* species not included in subclades A, B, and C in Figs. 2 and 3, two, *H. midiae* and *H. diversicolor,* are both positioned just ancestral to subclades B and C in the various trees, but to various degrees, depending upon the method and whether ITS1, ITS2, or the combined data are used. Figures 2 and 3 illustrate two of their positionings, and a common third arrangement is with *H. diversicolor* branching basal to *H. midiae,* followed by subclade C.

With respect to the final species, *H. iris* from New Zealand, there is only one placement. In all parsimony and distance derived trees, all the *Haliotis* species except *Haliotis iris* are united into one clade with 100% bootstrap support. Figure 3, a distance tree (Kimura two-parameter matrix, neighbor-joining algorythm) utilizing all the *Haliotis* species and both ITS1 and ITS2, illustrates the most visually striking aspect of the data, the basal trichotomy of long branches separating keyhole limpet from *H. iris* from all the remaining *Haliotis* species. These branch lengths are 4–10 times longer than those in any more terminal parts of the trees. The pairwise comparison values illustrate the basis of the phenomenon [for *H. iris* versus other *H. spp,* $p = .48–.49$ for ITS1, $p = .45–48$ for ITS2; for keyhole limpet versus non-*iris H. spp,* $p = .50–.54$ for ITS1, $p = .38–.40$ for ITS2; for *H. iris* versus keyhole limpet, $p = .59$ for ITS1, $p = .46$ for ITS2].

The tree in Fig. 3 also emphasizes visually how little variation there is within subclades A, B, and C. For the North Pacific species, the range of $p = 0–.049$ in ITS1, $p = 0–.044$ in ITS2; for the European species no difference in ITS1 and $p = .003$ in ITS2; and for the Australian species $p = .014–.017$ in ITS1, $p = .014–.020$ in ITS2. Although variation within subclades A, B, and C is limited, it still contributes to hierarchical structure, best understood by relation to the transcript secondary structure. For example, in ITS2 of clade A, there are 25 variable sites, of which eight are parsimony informative. All but one of these eight sites are found in single stranded regions of the secondary structure. This might be expected since positions remaining unpaired in the transcript secondary structure are known in other organisms to evolve at a significantly higher rate than those paired in transcript helices. The eighth parsimony informative

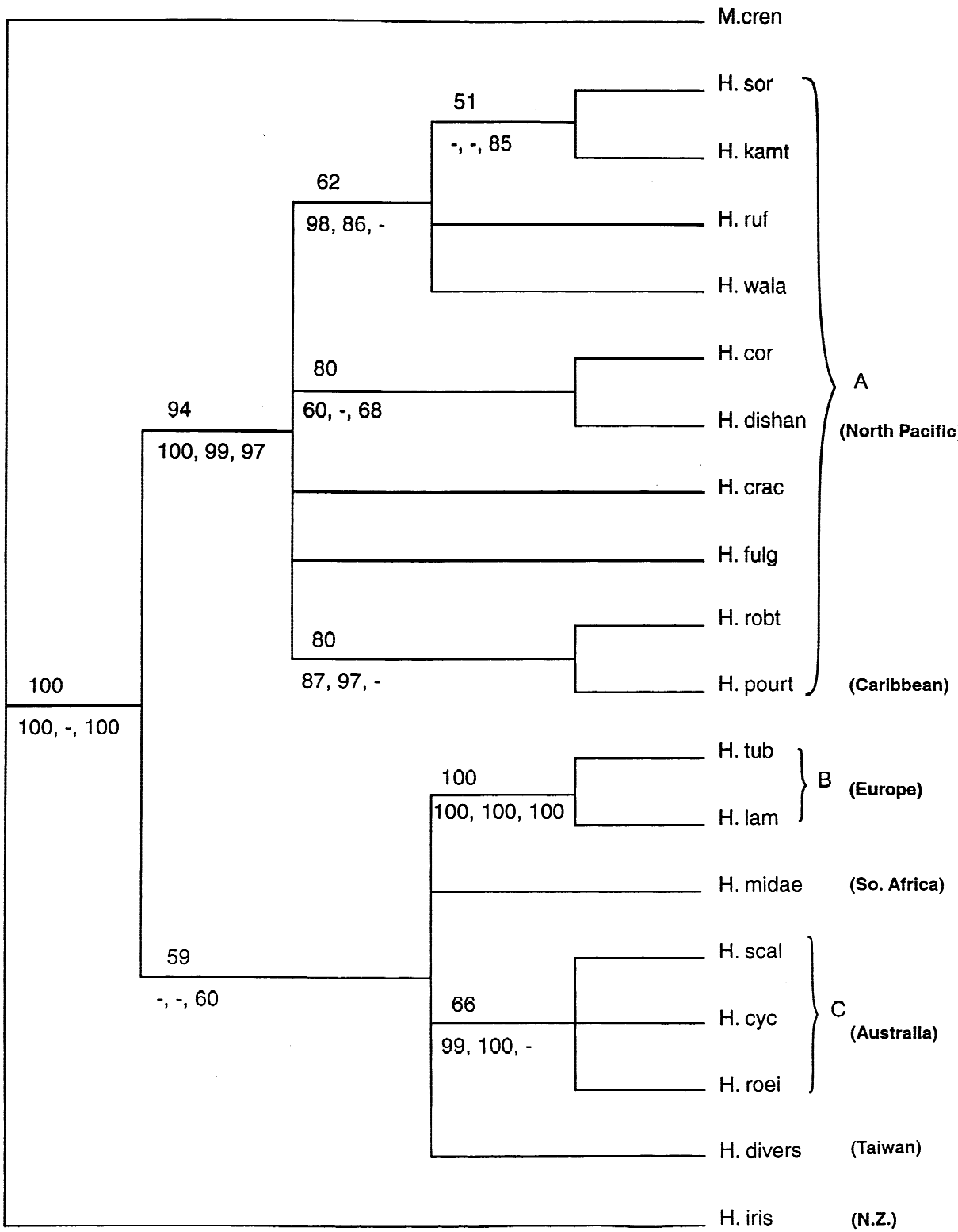

Fig. 2. Fifty percent majority rule consensus tree, obtained after 500 bootstrap repetitions for ITS1 + ITS2 parsimony analysis (heuristic, TBR default) with gaps coded as missing. Above the lines, percent frequencies greater than 50 are shown. Below the lines are shown 50% consensus bootstrap values for the same data including gaps as a fifth base for, respectively, ITS1 + 2, ITS1 only, or ITS2 only. Brackets indicate the three subclades (A, B, and C). The HI = 0.11, and CI I = 0.92. *H. rubra* was omitted because ITS2 was not available.

site is located in a helix found in the stem-loop of ITS2 loop I (arrowhead, Fig 1). Half of the North Pacific species sequences have a U in this position, pairing with a G in the transcript. This position has a C (pairing with the same G) in *H. corrugata*, *H. discus hannai*, *H. pourtalesii*, *H. roberti*, and *H. fulgens*, and all the non-North Pacific species of *Haliotis* as well. If truly a synapomorphy, this would suggest that *H. cracherodii* should join

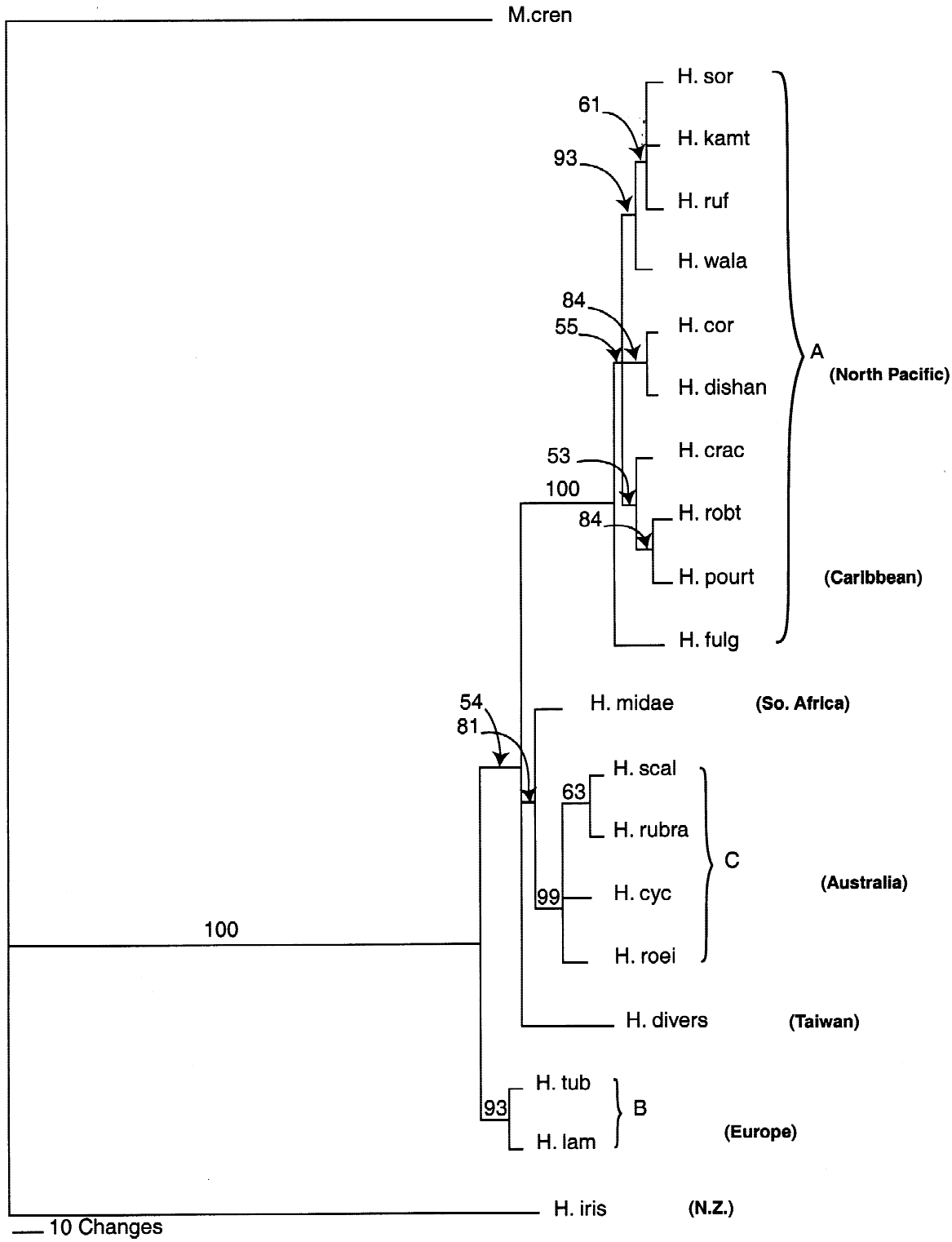

Fig. 3. Fifty percent majority rule tree obtained from 1000 bootstrap replicates by the distance, Kimura two-parameter, neighbor-joining analysis, using minimum evolution objective, for ITS1 and ITS2. For this analysis, all positions found variable within a species were omitted. The HI = 0.09 and CI = 0.91. Major subclades are designated as in Fig. 2. *H. rubra* is included here to illustrate its position; the lack of ITS2 has little effect, presumably because there are only four variable positions among all the Australian ITS2 sequences.

the clade of four other *Haliotis* species (*H. sorenseni, H. rufescens, H. walallensis,* and *H. kamtschatkana*) in subclade A. In Fig. 3, its position is different, but with only low bootstrap support.

Two additional aspects of the insertion/deletions deserve special attention. The first is the presence or absence, and sequence, of the indel of ITS1 shown in Table 2. This indel is extremely phylogenetically informative,

and could be utilized alone to create the trees. The second aspect, the position and length of minor indels, all in single stranded regions, is less informative. The characteristics of these indels, absence or presence as well as length, support the branching patterns in Figs. 2 and 3, as shown by coding gaps as a fifth state for parsimony analysis. Most of these indels actually reflect increase or decrease in homopolymer stretches of sequence, presumably reflecting the propensity for slippage during DNA replication over evolutionary time. Three indels in ITS1 involve homopolymers (0–5 T's between stem-loops 3 and 4, or 0–7 A's between stem-loops 4 and 5, or 0–3 T's distal to stem-loop 6). In ITS2, no similar polynucleotide tracts appear, but the presence or absence, and sequence, of insertions/deletions in terminal loops of I (2 nt), III (5 nt), and a bulge on the 3′ side of III (4–6 nt) fully support the reported clades.

Discussion

ITS As a Phylogenetic Tool

The extent of concerted evolution. Ribosomal repeats exist in multiple tandem copies in the genome. It is obviously important to have some estimate of how effectively concerted evolution (Hillis et al. 1991) has homogenized the many copies in an individual. Past experience suggests that, other than in known recent hybrids (Campbell et al. 1997; Coleman, in preparation), homogenization, as detected among ITS copies from the same individual, is nearly perfect. At most, there are 1–5 nucleotide substitutions, located in regions showing single-stranded structure in the RNA transcript secondary structure. The same generalization seems to apply to these *Haliotis* ITS sequences. No more than five variant positions were found in any individual, and the diversity of repeats observed between two individuals of one species was no greater.

A further question is whether any of the seventeen nucleotide polymorphisms found within a species, or by repeat sequencing of an organism (eleven in ITS1, six in ITS2), might be attributed to Taq polymerase error. Such error would be expected to occur randomly throughout the sequence. However, all but two of the positions displaying polymorphisms, either intra-individual or intra-species, are located in regions that remain single stranded in the RNA transcript secondary structure (Fig. 1). Since single stranded positions comprise only about 37% of ITS1 and 30% of ITS2, the polymorphisms are not randomly distributed. Positions remaining single stranded in the RNA transcript are not under selection for pairing, and at least some may be free to vary between two tolerated nucleotides, explaining the polymorphisms. Overall, with the maximum variation found within a species being five positions for the entire ITS1+2 (totaling 716 positions), it appears that concerted evolution is sufficient to keep all but a very few ITS positions homogeneous within a species of *Haliotis.*

ITS2 transcript secondary structure among eukaryotes. Secondary structure of the ITS2 RNA transcript has been analyzed most extensively among sets of related species in the green lineage of evolution (Coleman 2000); it appears that a common transcript folding pattern is exhibited from *Chlamydomonas* to the flowering plants. The *Haliotis* ITS2 transcript secondary structure shows a remarkable similarity (compare Fig. 1 here to Fig. 3 of Mai and Coleman 1997) to that found in the other species groups where ITS2 secondary structure has been derived by extensive species comparisons (Mai and Coleman 1997; Coleman 2000). Not only does the *Haliotis* ITS2 transcript form four stem-loops, the fourth of which is by far the most variable, but two other hallmark characteristics are present. Typically, the second hairpin loop has a pyrimidine-pyrimidine bulge in the fourth, fifth, sixth, or seventh pairing position. *Haliotis* ITS2 has a pyrimidine-pyrimidine bulge at pairing position 4 and at pairing position 7 from the base of stem II (Fig. 1, arrows) and so also does the keyhole limpet structure. Also similar is the location on the 5′ side of hairpin loop III of the longest absolutely conserved region of primary sequence among even taxonomically distant relatives. At exactly this position, all the *Haliotis* species and the keyhole limpet have their longest shared nucleotide sequence, 18 nucleotides, identical with the exception of *H. iris* (one nucleotide substitution, a transition that preserves pairing). Even *Concholepas* retains a ten nucleotide sequence of identity in this precise position of the secondary structure.

Unfortunately for our purposes, Haliotidae is a monogeneric family, and no ITS sequences of other molluscs of the same order are yet available for comparison at higher taxonomic levels. However, the remarkable similarity of keyhole limpet and *Haliotis* ITS, as well as the observation of recognizable sequence and secondary structure similarity with *Concholepas,* suggest that knowledge of comparative secondary structure of ITS2 regions could greatly aid in guiding alignment of sequences from different families of molluscs to obtain the maximal amount of phylogenetic information.

Comparisons with other Haliotis *phylogenetic studies.* Reconstruction of the phylogenetic history of the *Haliotis* species has been attempted previously, using both gene loci and other aspects of DNA. For example, a tandem repeat of undefined function is present in the nuclear DNA in certain *Haliotis* species (*H. rufescens, sorenseni, kamtschatkana, walallensis,* and *corrugata*), but is not present in DNA of *H. cracherodii* or *H. fulgens* (Muchmore et al. 1998). The placement of *H. cracherodii* and *H. fulgens* in Figs. 2 and 3 does not conflict with a possible single acquisition of this DNA character.

Table 3. Known hybrid animals.

	wht	red	pk	grn	pnto	flt	blk	disc	mad	giga	rub	laev
wht	x											
red	ZM	x										
pk	ZmZ	ZM	x									
grn		ZmZ	ZM	x								
pnto	ZmZ	Zm	ZmZ	x								
flt	ZmZ	ZmZ	Zm		ZmZ	x						
blk					ZmZ		x					
disc					ZmZ			x				
mad								zm	x			
giga								zm	zm	x		
rub											x	
laev											ZmZ	x

Where no data are presented below the diagonal, no attempts at crosses have been reported.
x = parentals, ZM = mature F_1, Zm = small numbers of mature F_1, ZmZ = rare mature F_1, zm = F_1 reported, maturity unknown, giga = *H. gigantea* (Japan), laev = *H. laevigata* (S. Australia), mad = *H. madaka* (Japan).
From experimental crosses: disc, pnto ((Lyon and Vacquier 1999); disc, mad, giga (Sakai, personal communication); rub, laev (Brown 1995).
From analysis of shells: wht, red, pk, grn, pnto, flt (Owen et al. 1970).

With respect to particular genes and their associated regions, phylogenetic reconstructions are available for the cDNA sequences of the sperm protein lysin, the flanking 3′ untranslated sequence of lysin (Swanson and Vacquier 1998; Lee et al. 1995), the 18 kD protein related to lysin (Vacquier et al. 1997), and the *mt*COI gene (Metz et al. 1998). In the North Pacific species ITS subclade (A), the foursome of California species grouped together in Figs. 2 and 3 also group together in trees for these four additional loci. Trees of all four loci also locate *H. fulgens* basal to the remaining California species, but with low bootstrap support except in the case of lysin. This same result is seen in Fig. 3, but lacking strong bootstrap support. The pairing of *H. corrugata* with *H. discus hannai,* exclusive of other species, is supported by moderate bootstrap levels in Figs. 2 and 3. This pairing has less than 50% bootstrap support in trees from lysin and its flanking 3′ region, but appears in the mtCOI tree (somewhere between 50–70% bootstrap support). No comparison is available for the more strongly supported pairing in Figs. 2 and 3, *H. roberti* with *H. pourtalesii,* since no previous study has utilized a Caribbean species (*H. pourtalesii*). The prior studies support no other exclusive pairings among subclade A species.

For lysin and its flanking 3′ sequence, where a more worldwide selection of species was included, exactly the same three strongly supported subclades, A, B, and C, are found in all the trees, and subclade A is separated from subclades B and C by a branch with 100% bootstrap support, as in the ITS results. The location of *H. diversicolor* is basal to *H. midae,* and both are basal to subclades B and C, although this positioning has relatively low bootstrap support for both lysin and its 3′ sequence. As in the ITS analysis, *H. iris* is on a long branch, associated with the subclade A species in the lysin analyses and with the other species in the lysin 3′ untranslated sequence study. Trees in both studies are unrooted.

Overall, there are no conflicts between the ITS analyses and the consensus of previous phylogenetic reconstructions using other DNA loci. These congruencies among all the DNA sequencing analyses are perhaps the more remarkable since ITS has been assumed to evolve neutrally with respect to speciation mechanisms while lysin is one of the best examples of a gene evolving rapidly under strong positive selection for speciation (cf. Ting et al. 2000). Lysin appears to evolve 4–50 times faster than the fastest evolving proteins common to mice and humans (Kresge et al. 2001). In sum, not only is the ITS phylogenetic reconstruction as resolved in its robust support of clades as those derived from other DNA studies, but also it is strongly supported by the unusual ITS1 indel.

ITS Similarity Versus Interbreeding Results in Haliotis

From the ITS analyses, the North Pacific species appear to be closely related. If so, it is of interest to examine whether they can potentially interbreeed successfully. Four papers (Lyon and Vacquier 1999; Leighton and Lewis 1982; Hoshikawa et al. 1998; Vacquier and Lee 1993) present results from various experimental pairings of eight North Pacific species. These include data on frequency of fertilization in interspecies pairings and titer of lysin needed for vitelline envelope dissolution. Most reciprocal pairings, but not all, produce similar results in terms of percent fertilization. In every case attempted, at least some fertilization was reported, using standard concentrations of sperm; however the significance of such data may be limited by the fact that, in *Haliotis,* it is known that high densities of sperm will effect fertilization in crosses (Leighton and Lewis 1982).

Table 3 compiles the available data concerning postzygotic development of hybrids among ten species of the North Pacific clade and two species of the Australian clade, including all cases where development has occurred and, where known, if hybrids mature. Obviously, the frequency of hybrid adults found in nature must reflect a combination of opportunity and genetic potential. Although many potential crosses have yet to be attempted experimentally, one might infer from the table that the North Pacific species are capable of some minimal level of gene exchange, at least under experimental conditions, since hybrid animals develop to maturity. Unfortunately, almost no backcrosses have yet been reported. Interestingly, the five crosses producing the lowest F_1 survival all involved *H. fulgens* as one parent (Leighton and Lewis 1982). This parallels the basal branching of *H. fulgens* with respect to the other representatives of subclade A. The Australian species intercross successfully but may or may not be compatible at all with North Pacific species, since there are no recorded attempts at crossing. No data are available concerning crosses between taxa from Europe (subclade B). Within subclades A and C, then, postzygotic barriers to interbreeding are not yet fully developed.

Transcript secondary structure characteristics permit a second type of comparison with other eukaryotes. In all the other eukaryote groups where a broad array of species has been compared for both ITS2 sequence secondary structure and tested for any vestige of interspecies sexual compatibility, an interesting correlation has been found. When sufficient evolutionary distance has accumulated to produce even one CBC in the relatively conserved pairing positions of the ITS2 transcript secondary structure, taxa differing by the CBC are observed experimentally to be totally incapable of intercrossing (reviewed in Coleman 2000). This general observation is presumably true also for *Haliotis*, since the only species displaying an ITS2 CBC with respect to the remainder of the genus, *H. iris*, is almost certainly completely sexually isolated from all the others.

An understanding of how recently diverged the California species are is important to ongoing studies of their genes active in gamete recognition. For example, in order to focus experimental studies on secondary and tertiary structure of lysin, it is valuable to be able to distinguish biologically important regions of sequence difference between species from any variation that presumably has little effect, as suggested by its showing significant variation even within a species. Lee *et al.* (1995) reported finding no more than three nucleotide differences in the lysin gene between individuals of a species, but two subregions that varied between closely related species. Subsequently, Lyon and Vacquier (1999) exchanged one of these, the 3′ terminal sequence of sperm lysin, between the genes of two closely related species and quantitated the abilities of these chimeric lysins to dissolve homologous and heterologous egg vitelline envelopes. Their results confirmed the importance of the 3′ terminal species-specific sequences of lysin in vitelline envelope dissolution. The other region of the gene sequence predicted to be important to gamete interaction in pairings between the California species has not yet been tested experimentally. In addition, there is a major disjunction in similarity between the lysin sequences of the North Pacific species compared to those of the remaining *Haliotis* species in Fig. 2 of Lee and Vacquier (1995), which might represent a degree of evolutionary change resulting in no gamete interaction at all.

Evolutionary History and Biogeography

The branching pattern of the Taiwan-Australia-South Africa assortment of species remains unresolved. From this limited data set one might conclude that the common ancestor of the *Haliotis* species had a major ITS1 insert (Table 2), and that all the North Pacific species diverged fairly recently from one ancestral type, hence their uniformity for this insert. The Mediterranean-Atlantic species is likely derived from a different ancestor, one that had lost the insert. If this interpretation is correct, then the ancestral relationship of Australian-Taiwan-South Africa species, all containing one or another variant forms of the insert, will only be sorted out when some of the numerous smaller abalone species of the Indo-Malayan Archipelago, the most species-diverse area for abalone (Geiger 1999), have been sequenced and compared. The bias of interest has been toward the larger, North Pacific species, of great economic importance.

The earliest recognizable *Haliotis* fossils are upper Cretaceous (about 70 mya) from the Caribbean/California region, although there are fossil shells known from 200 mya with a perforation similar to those in abalone, suggesting an earlier beginning for the family (Lindberg 1991). The fossil record also suggests that the Pacific clade was present and fairly widespread as early as 5 mya (Lindberg 1992). The most interesting phylogenetic result of the ITS analyses is the robust inclusion of *H. pourtalesii* (from the Caribbean) in the otherwise North Pacific clade, not with the Mediterranean-Atlantic clade. The *H. pourtalesii/H. roberti* pair differ at a total of only ten nucleotide positions in ITS1 and ITS2. This observation implies that relatively little sequence change has occurred in ITS, at least since the rise of the Meso-American land bridge 3–3.5 million years ago. If this interpretation is correct, it provides a landmark date for their vicariance.

The present ITS analysis, and prior DNA studies as well, suggest that the New Zealand species, *H. iris*, is evolutionarily distant from all the other species of the genus. This could merely be a case of rapid evolution producing a long branch, or it might possibly reflect an ancient isolation of the lineage leading to *H. iris*. If the

latter is the case, the length of the *H. iris* branch must be sufficient at least to satisfy the known timing (about 80 mya) of the isolation of New Zealand from other land masses. A very rough test of this latter explanation can be made by calculating the rate of ITS change since the Meso-American vicariance event, and then, assuming a constant rate of ITS change, using this rate to assess the length of the *H. iris* branch.

If one brackets the timing of the separation of *H. roberti* from *H. pourtalesii* in the range of 3–4 mya, knowing their ITS1+ITS2 difference ($p = 0.0173$), one obtains a rate of nucleotide substitution of 0.0029–0.0022 per million years (0.0173 divided by 2 to estimate the distance from the branch point, and the result divided by either 3 my or 4 my, respectively). These rates multiplied by 80 my give 0.232–0.176 minimal expected branch length from the node separating isolated New Zealand species from other species. The observed difference between *H. iris* and the remaining *Haliotis* species is 0.478 (range of $p = 0.471$–0.486), which divided by two (to obtain the branch length from the node) is 0.239. This is longer than the minimum required to fit the hypothesis (0.232–0.176). Thus, it appears that the remoteness of *H. iris* may reflect the isolation of its ancestry since the separation of New Zealand from any other land mass, although other explanations are always possible. If the rate of nucleotide substitution is extended into the past, it yields a possible ancestor common to all the living species of *Haliotis* at ca. 83–111 mya. Although such calculations are subject to assumptions at each step, the result is remarkably compatible with the fossil record and history of continental drift. Perhaps additional DNA information from southern hemisphere abalone will illuminate the position of *H. iris* from New Zealand, the species that now appears from DNA studies to be so distantly related.

Acknowledgements. Our grateful thanks to Dr. Eduardo Bustos, IFOP, Puerto Montt, Chile, for the gift of juvenile *Concholepas concholepas* from the abalone hatchery; to Dr. D.L. Geiger, Natural History Museum of Los Angeles County, for tissue from *H. pourtalesii;* and to Dr. H. Chaney, Santa Barbara Museum of Natural History, for a sample of *H. roberti* from Cocos Island, its only known habitat. This manuscript benefited from numerous beneficial suggestions from the editor and the referees.

References

Baldwin BG (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. Mol Phylogenet Evol 1:3–16

Brown D (1995) Genetic evidence for hybridisation between *Haliotis rubra* and *H. laevigata.* Mar Biol 123:80–93

Campbell CS, Wojciechowski UF, Baldwin BG, Lawrence AA, Donoghue MJ (1997) Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic comples (Rosaceae). Mol Biol Evol 14: 81–90

Coleman AW (2000) The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. Protist 151:1–9

Coleman AW, Jaenicke L, Starr RC (2001) Genetics and sexual behavior of the pheromone producer, *Chlamydomonas allensworthii* (Chlorophyceae). J Phycol 37:1–5

Geiger DL (1998) Recent genera and species of the family Haliotidae (Gastropoda: Vetigastropoda). The Nautilus 111:85–116

Geiger DL (1999) PhD dissertation, U of Southern Calif, Los Angeles, CA, pp. 423

Gutell RR, Larsen H, Woese CR (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol Rev 58:10–26

Hadjiolova KV, Normann A, Cavaille J, Soupene E, Mazan S, Hadjiolov AA, Bachellerie J-P (1994) Processing of truncated mouse or human rRNA transcribed from ribosomal minigenes transfected into mouse cells. Mol Cell Biol 14:4044–4056

Hillis DM, Moritz C, Porter CA, Baker RJ (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science 251:308–310

Hoshikawa H, Sakai Y, Kijima A (1998) Growth characteristics of the hybrid between pinto abalone, *Haliotis kamtschatkana* Jonas, and ezo abalone, *H. discus hannai* Ino, under high and low temperature. J Shellfish Res 17:673–677

Kresge N, Vacquier VD, Stout CD (2001) Abalone lysin: the dissolving and evolving sperm protein. BioEssays 23:95–103

Lee Y-H, Vacquier VD (1995) Evolution and systematics in Haliotidae (Mollusca: Gastropoda): inferences from DNA sequences of sperm lysin. Marine Biology 124:267–278

Lee Y-H, Ota T, Vacquier VD (1995) Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol Biol Evol 12:231–238

Leighton DL, Lewis CA (1982) Experimental hybridization in abalones. Int J Invert Reproduction 5:273–282

Lindberg DR (1992) Evolution, distribution and systematics of Haliotidae. In: Shepherd SA, Tegner MJ, del Proo G (eds) Abalone of the world: biology, fisheries and culture. Blackwell Scientific, London, pp. 3–19

Lyon JD, Vacquier VD (1999) Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. Dev Biol 214:151–159

Mai JC, Coleman AW (1997) The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. J Mol Evol 44:258–271

Metz EC, Robles-Sikisaka R, Vacquier VD (1998) Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochonchial DNA. Proc Natl Acad Sci USA 95:10676–10681

Muchmore M-E, Moy GW, Swanson WJ, Vacquier VD (1998) Direct sequencing of genomic DNA for characterization of a satellite DNA in five species of Eastern Pacific abalone. Mol Marine Biol Biotech 7: 1–6

Owen RS, McLean JH, Meyer RJ (1970) Hybridization in the eastern Pacific abalones (*Haliotis*). Bull Los Angeles City Mus Nat Hist Sci 9:1–37

Schlotterer C, Hauser M-T, von Haeseler A, Tautz D (1994) Comparative evolutionary analysis of rDNA ITS regions in *Drosophila.* Mol Biol Evol 11:513–522

Swanson WJ, Vacquier VD (1998) Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science 281: 710–712

Ting C-T, Tsaur S-C, Wu C-I (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus.* Proc Natl Acad Sci USA 97: 5313–5316

Torres RA, Ganal M, Hemleben V (1990) GC balance in the internal transcribed spacers ITS1 and ITS2 of nuclear ribosomal RNA genes. J Mol Evol 30:170–181

Vacquier VD, Lee Y-H (1993) Abalone sperm lysin: unusual mode of evolution of a gamete recognition protein. Zygote 1:181–196

Vacquier VD, Swanson WJ, Lee Y-H (1997) Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? J Mol Evol 44:S15–S22

van Nues RW, Rientjes JMJ, Morre SA, Mollee E, Planta RJ, Venema J, Raue HA (1995) Evolutionarily conserved structural elements are critical for processing of Internal Transcribed Spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. J Mol Biol 250: 24–36

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols. Academic Press, New York, pp. 315–322

Zuker M, Mathews DH, Turner DH (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski J, Clark BFC (eds) RNA biochemistry and biotechnology, NATO ASE Series. Klewer Academic Publishers, Hingham, MA, pp. 11–43