Molecular Phylogenetics and Evolution 48 (2008) 718–727

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Utility of arginine kinase for resolution of phylogenetic relationships among brachyuran genera and families

Brian C. Mahon [a,b], Joseph E. Neigel [a,*]

[a] Department of Biology, University of Louisiana at Lafayette, P.O. Box 42451, Lafayette, LA 70504, USA
[b] Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA

ARTICLE INFO

Article history:
Received 14 December 2007
Revised 2 May 2008
Accepted 3 May 2008
Available online 11 May 2008

Keywords:
Decapoda
Brachyura
Arginine kinase
Phylogeny
Cytochrome oxidase I
Phylogenetic informativeness

ABSTRACT

The molecular phylogenetics of decapod crustaceans has been based on sequence data from a limited number of genes. These have included rapidly evolving mitochondrial genes, which are most appropriate for studies of closely related species, and slowly evolving nuclear ribosomal RNA genes, which have been most useful for resolution of deep branches within the Decapoda. Here we examine the utility of the nuclear gene that encodes arginine kinase for phylogenetic reconstruction at intermediate levels (relationships among genera and families) within the decapod infraorder Brachyura (the true crabs). Analyses based on arginine kinase sequences were compared and combined with those for the mitochondrial cytochrome oxidase I gene. All of the genera in our taxon sample were resolved with high support with arginine kinase data alone. However, some of these genera were grouped into clades that are in conflict with recognized brachyuran families. A phylogeny based on cytochrome oxidase I was consistent with the arginine kinase phylogeny, but with weaker support. A recently proposed measure of phylogenetic informativeness indicated that arginine kinase was generally more informative than cytochrome oxidase I for relationships above the level of genus. Combined analysis of data from both genes provided strong support for clades that are in conflict with current assignments of genera to the families Epialtidae, Mithracidae, Pisidae, and Portunidae.

1. Introduction

The crustacean order Decapoda is an ecologically and economically important group that includes crabs, shrimps, lobsters and crayfish. Phylogenetic relationships within the Decapoda have been the subjects of numerous molecular systematic investigations. Some of the earliest were based on the nuclear large subunit ribosomal RNA gene (18S), and focused on relationships among suborders and families (Kim and Abele, 1990; Perez-Losada et al., 2002; Spears et al., 1992). After Cunningham and co-workers (1992) used a phylogeny based on the mitochondrial large subunit ribosomal RNA gene (16S) to argue that lithodid crabs were part of a paguroid clade, 16S became widely used for crustacean molecular systematics. Unlike 18S, 16S evolves rapidly enough to resolve phylogenetic relationships among closely related species (Schubart et al., 2000). For some studies of decapod systematics, data from multiple ribosomal RNA genes have been combined (e.g. Ahyong and O'Meally, 2004; Tudge and Cunningham, 2002). A few of the protein-coding genes of the mitochondrial genome have also been used in decapod systematics. The gene that encodes cytochrome oxidase subunit I (COXI) is used both in decapod phylogenetics and also as a "DNA barcode" for species identification (Hebert et al., 2003). Cytochrome oxidase II has been used for the molecular systematics of crabs in the family Aeglidae (Perez-Losada et al., 2004).

Relatively few phylogenetic analyses of decapods have used the sequences of protein-coding genes from the nuclear genome. A phylogenetic analysis based on glucose-6-phosphate isomerase (GPI) and elongation factor-1α (EF-1α) along with the mitochondrial protein-coding gene cytochrome oxidase subunit I (COXI) provided evidence of three major clades within the snapping shrimp genus *Alpheus* (Williams et al., 2001). Porter et al. (2005) combined data for three ribosomal genes (16S, 18S and the nuclear 28S gene) with sequences of the nuclear histone H3 gene to examine relationships among decapod infraorders. Regier and Shultz (2001) demonstrated the usefulness of the nuclear protein-coding gene elongation factor-2 (EF-2) to resolve higher-level relationships within the Arthropoda, such as the controversial relationships among the subphyla Crustacea, Hexapoda and Myriapoda.

For accurate phylogenetic inference, gene sequences must be correctly aligned and provide a phylogenetic signal that is not degraded by saturation at variable sites. Of the genes that have been most commonly used for crustacean systematics, ribosomal RNA sequences can be difficult to align without ambiguity and mitochondrial protein-coding genes often prone to mutational saturation. Neigel and Mahon (2007) argued that there has been

* Corresponding author. Fax: +1 337 482 5660.
E-mail address: jneigel@louisiana.edu (J.E. Neigel).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.05.005

too much reliance on these few genes in crustacean systematics and that additional nuclear protein-coding loci are needed to generate accurate phylogenies. In their review, they noted that of the 14 loci that have been used for crustacean systematics, five are mitochondrial and four are within the nuclear ribosomal repeat unit. Of the remaining five, four encode proteins with highly conserved amino acid sequences that are most useful for resolution of higher-level relationships within the Arthropoda. Only one gene, glucose 6-phosphate isomerase, qualifies as a nuclear gene that encodes a protein that is not highly conserved at the amino acid level (Williams et al., 2001). In contrast, over 30 nuclear protein-coding loci have been used in insect systematics (Caterino et al., 2000; Hardy, 2007). Clearly, there is a need for more protein-coding nuclear genes in crustacean systematics, especially for questions about phylogenetic relationships at levels that fall between the optimal ranges of highly conserved nuclear genes and rapidly evolving mitochondrial genes.

The average rate of sequence divergence can provide a rough indication of how suitable a gene might be for the resolution of deep versus shallow phylogenetic relationships. However, because the actual distribution of rates among sites is critical, the average rate is not a reliable measure of how useful a gene sequence will be for a particular phylogenetic problem. Townsend (2007) recently proposed such a measure, phylogenetic informativeness (PI), based on the number of sites expected to have changed along the short internal branches of a polytomy but not along the long branches that lead from the polytomy to the terminal taxa. PI is calculated from estimated rates of change at each site rather than an average rate for the entire sequence. Thus a sequence with a few rapidly evolving sites among many invariant sites will have a different PI than a sequence with the same overall rate of divergence but a more even distribution across sites. Profiles of relative phylogenetic informativeness (RPI) can be used to show the proportional contributions of different genes to the resolution of polytomies across a range of depths within a phylogeny.

Here we evaluate a region of the nuclear gene that encodes arginine kinase (AK) for phylogenetic analysis of the Brachyura. AK belongs to the phosphagen kinase family of enzymes, which catalyze the reversible transfer of phosphoryl groups between ATP and phosphagen energy stores (Ellington, 2001). This activity functions to buffer ATP levels under conditions of high demand (Meyer et al., 1984; Walliman et al., 1992), regulate levels of inorganic phosphate (Meyer et al., 1986), and transport energy intracellularly (Bessman and Capenter, 1985). AK has been identified in most animal groups, except vertebrates, although AK in echinoderms appears to have evolved independently of AK in other animals (Suzuki et al., 1999). We conducted our phylogenetic analysis with AK sequences in parallel with an analysis based on the mitochondrial gene COXI. This allowed us to compare our results for AK with those for one of the most commonly used genes in crustacean systematics.

2. Materials and methods

2.1. Specimen collection

Live specimens were collected by snorkeling or SCUBA, or purchased from commercial vendors (Table 1). Tissue was removed immediately after animals were sacrificed so that DNA could be extracted from fresh tissue that was not frozen or preserved in alcohol. Most specimens were photographed, preserved as vouchers and deposited in our collection.

2.2. Taxon sampling

Taxa were selected to represent groupings at several levels: species within genera, genera within families, and families within superfamilies (Table 1). We used the brachyuran section of Martin and Davis (2001) for an overall taxonomic framework and McLaughlin and colleagues (2005) for individual assignments of species to families. Based on these classifications, our taxon sample for AK sequences includes five species within the genus *Cancer*, seven species representing four families within the superfamily Majoidea, six species representing two families within the superfamily Xanthoidea, and nine species representing three families within the superfamily Grapsoidea. We also included representatives of the families Portunidae, Ocypodidae and Pinnotheridae. Sequences from two non-brachyuran decapod species were used as outgroups: the anomuran *Oedignathus inermis* and the thalassinidean *Lepidopthalmus louisianensis*.

Our sample of COXI sequences included most, but not all of the species included in our taxon sample for AK (Table 1). Except where sequences where obtained from GenBank, AK and COXI sequences were obtained from the same specimen. For phylogenies that were based all or in part on COXI sequences, we used *O. inermis* as the sole outgroup. We did not use *L. louisianensis* in these cases because its COXI sequence exhibited unusual (for decapods) biases in nucleotide composition; which have been shown to produce errors in phylogenetic inference (Hassanin et al., 2005).

2.3. Primer design

PCR primers for AK were based on alignment of cDNA sequences from several species (GenBank Accession Nos.: AF167313, AF233355, AF233356, AF233357, and AF288785). We used the computer program NAR (Rychlik and Rhoads, 1989) to avoid primer sequences with significant secondary structure or potential for dimer formation. Two pairs of primers were designed to amplify overlapping regions of the second exon of the AK gene (Fig. 1). The first pair were designated AKF2 (forward) and AKR2 (reverse) and the second pair AKF3 (forward) and AKR3 (reverse). AKR3 is a modification of the ENDR1 primer (Kotlyar et al., 2000) designed for amplification of AK cDNA from *Carcinus maenas* and *Callinectes sapidus*. Degenerate primers for amplification of a portion of the COXI gene (primers COIF2 and COIR2, product size 658 bp or COIF2B and COIR2, 622 bp) were based on primers commonly used for insects (Simon et al., 1994), which we modified for brachyurans (Table 2).

2.4. DNA extraction, amplification and sequencing

Genomic DNA from fresh muscle tissue or testes was extracted and purified following PureGene (Gentra Systems) extraction protocols for animal tissue. Extracted DNA was quantified on a TD-360 Fluorometer (Turner Designs) and visualized on a 0.7% agarose gel stained with ethidium bromide to verify DNA quality. PCR reactions were conducted in a Stratagene RoboCycler thermal cycler with the following amplification profile: 10 min at 95 °C, followed by 35 cycles of 1 min at 95 °C, 30 s at the optimal annealing temperature, and 30 s at 72 °C and followed by an additional 7 min at 72 °C. Table 2 lists optimal annealing temperatures for each pair of primers.

Each 25 μl PCR reaction contained 200 μM of each dNTP; 300 nM each PCR primer; 1× PCR buffer (10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin, Perkin Elmer); 1.25 U AmpliTaq Gold polymerase (Perkin Elmer); and template DNA (10–60 ng). PCR products were visualized on a 1.5% agarose gel to verify a single product of the correct size was amplified. Two microliters of ExoSAP-it (USB) were added to the remaining PCR product (~20 μl) and incubated at 37 °C for 15 min to remove residual primers and dNTPs, followed by 15 min at 85 °C to heat deactivate the enzymes. Amplicons were sequenced with Big Dye™

Table 1
Taxonomic sampling, collection locations or supplier and GenBank Accession Nos. (GB #)

Superfamily	Family	Genus and species	Collection location	AK	COXI
Infraorder Anomura					
Paguroidea	Lithodidae	*Oedignathus inermis*	Bodega Bay, CA	EU329144	EU329164
Infraorder Thalassinidea					
Callianassoidea	Callianassidae	*Lepidopthalmus louisianensis*	Bay St. Louis, MS	EU329129	None
Infraorder Brachyura: Subsection Heterotremata					
Cancroidea	Cancridae	*Cancer antennarius*	Bodega Bay, CA	EU329139	EU329149
		Cancer irroratus	Gulf of Maine, Inc., ME	EU329138	EU329150
		Cancer jordani	Bodega Bay, CA	EU329137	EU329151
		Cancer magister	Albertsons, LA (from OR)	EU329136	EU329152
		Cancer productus	Bodega Bay, CA	EU329145	EU329153
Majoidea	Epialtidae	*Pugettia producta*	Bodega Bay, CA	EU329117	EU329168
	Mithracidae	*Microphrys bicornutus*	Bocas del Toro, Panama	EU329125	EU329161
		Mithraculus forceps	Bocas del Toro, Panama	EU329123	EU329162
		Mithraculus sculptus	Bocas del Toro, Panama	EU329124	EU329163
	Pisidae	*Loxorhynchus crispatus*	Bodega Bay, CA	EU329128	EU329158
		Scyra acutifrons	Bodega Bay, CA	EU329116	EU329169
	Tychidae	*Pitho lherminieri*	Bocas del Toro, Panama	EU329119	EU329167
Portunoidea	Portunidae	*Arenaeus cribrarius*	Bocas del Toro, Panama	EU329142	EU329146
		Callinectes sapidus	T-Bob's Seafood, LA	EU329140	EU329148
		Carcinus maenas 1	Gulf of Maine, Inc , ME	EU329135	EU329154
		Carcinus maenas 2	GenBank sequence	AF167313	None
		Portunus sp.	Bocas del Toro, Panama	EU329118	None
Xanthoidea	Menippidae	*Eriphia gonagra*	Bocas del Toro, Panama	EU329134	EU329155
		Menippe adina	Terrebonne Bay, LA	EU329127	EU329159
		Menippe mercenaria	Beaufort, NC	EU329126	EU329160
	Panopeidae	*Eurypanopeus abbreviatus*	S. Padre Island, TX	EU329133	None
		Eurytium limosum	Tampa Bay, FL	EU329132	None
		Panopeus obesus	Tampa Bay, FL	EU329120	None
Infraorder Brachyura: Subsection Thoracotremata					
Grapsoidea	Grapsidae	*Pachygrapsus crassipes*	Bodega Bay, CA	EU329121	EU329165
		Pachygrapsus gracilis	Bocas del Toro, Panama	EU329122	EU329166
		Pachygrapsus marmoratus	GenBank sequence	AF288785	None
	Sesarmidae	*Aratus pisonii*	Tampa Bay, FL	EU329143	None
		Armases cinereum	Tampa Bay, FL	EU329141	EU329147
		Sesarma reticulatum	Cypermort Point, LA	EU329115	EU329170
	Varunidae	*Chasmagnathus granulata*	GenBank sequence	AF233357	None
		Eriocheir sinensis	GenBank sequence	AF233356	AY274302
		Hemigrapsus oregonensis	Bodega Bay, CA	EU329130	EU329157
Ocypodoidea	Ocypodidae	*Uca longisignalis*	Cypermort Point, LA	EU329114	EU329171
		Uca minax	Cypermort Point, LA	EU329113	EU329172
		Uca pugilator	Tampa Bay, FL	EU329112	EU329173
Pinnotheroidea	Pinnotheridae	*Fabia subquadrata*	Bodega Bay, CA	EU329131	EU329156

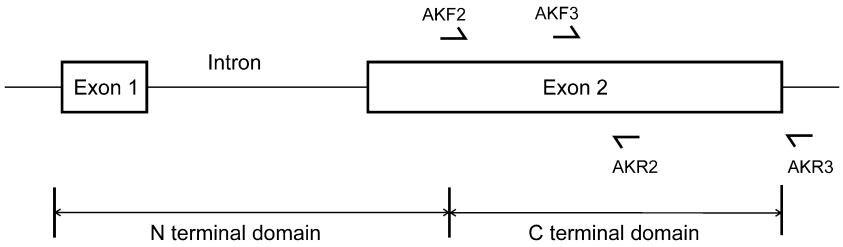

Fig. 1. Map of the arginine kinase gene showing the location of PCR primers.

(Applied Biosystems) terminator cycle sequencing, purified through Sephadex spin columns, dried down, resuspended in high-deionized formamide, and analyzed with an ABI Prism™ Genetic Analyzer with Sequencing Analysis software (PE Applied Biosystems).

In addition to the sequences that we determined, we obtained five from GenBank to use in our phylogenetic analyses. These included one of the two AK sequences for *Carcinus maenas* (AF167313), AK sequences for *Chasmagnathus granulata* (AF233357), *Eriocheir sinensis* (AF233356), and *Pachygrapsus marmoratus* (AF288785) and a COXI sequence for *Eriocheir sinensis* (AY274302).

2.5. Data analysis

Sequences were aligned with ClustalW v1.8.3 (Chenna et al., 2003) set to 15 for the gap open penalty, and 6.66 for the gap extension penalty. The Perl script MrAIC 1.3 (Nylander, 2004) in combination with PHYML v2.4.4 (Guindon and Gascuel, 2003) was used to choose the best models of sequence evolution for each data set by the Akaike Information Criterion (AIC). PAUP* 4.0b10 (Swofford, 2002) with the heuristic search option was used for maximum parsimony (MP) analysis, PHYML v2.4.4 for maximum likelihood (ML) analysis, and MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001) for Bayesian analysis (BA). Nonparametric boot-

Table 2
Primer pairs used for this study

Name	Insect equivalent	Direction	Position	Primer sequence 5′–3′	T_a
Arginine kinase					
AKF2		Forward	391–411	AAY GTG GAC CCC GAT GGC AAA	61
AKR2		Reverse	832–852	CAG GCG GTC RTG GTG RGA GAA	
AKF3		Forward	760–780	ATG GGY GGT GAY YTG GGC CAG	61
AKR3		Reverse	1148–1167	CGG CTC GCC CTC ACC CAT A	
Cytochrome oxidase 1 (COXI)					
COIF2	C1-J-1718	Forward	1496–1521	GGA GGA TTT GGA AAT TGA TTA GTW CC	53
COIF2B	C1-J-1751	Forward	1532–1554	GGA GCH CCT GAT ATR GCT TTY CC	
COIR2	C1-J-2329	Reverse	2129–2154	ACD GTR AAY ATR TGR TGN GCT CAD AC	

T_a indicates the optimized annealing temperature for each primer pair. Position indicates the primer position in respect to the GenBank reference sequence for the species *Carcinus maenas* (AK) and *Callinectes sapidus* (COXI) (GenBank Accession Nos. AF167313 and AY363392, respectively). "Insect equivalent" is the name of the insect primer for COXI as reported in Simon et al. (1994).

strap values are reported as percentages of the total number of replicates. ML bootstrap support (MLBS) and MP bootstrap support (MPBS) were assessed for 2000 replicates of each analysis. BA was conducted with two independent runs, each with one cold chain and three heated chains with the temperature parameter set to 0.1. Each run was sampled every 300 generations and continued at least until the standard deviation of the splits was below 0.006, which required between three million and 10 million generations. Data from the first 25% of each run was discarded as a burn-in period to allow the chains to reach stationary distributions. Bayesian posterior probabilities (BPP) are reported as percentages (100 represents a probability of 1). For ML analysis of AK and COXI combined, sequences were concatenated. For the BA analysis of the combined sequences, each sequence was assigned to a separate data partition with independent parameter estimation. Newick format tree files were opened in Tree Explorer portion of MEGA 3.1 (Kumar et al., 2004) and Corel Draw 12 (Corel Corporation) was used for editing. Support values are shown only for branches in which either the MPBS or MLBS value was above 50.

Phylogenetic informativeness (PI) and relative phylogenetic informativeness (RPI) were calculated as described in Townsend (2007), with some modifications. Rates of evolution were estimated for individual sites with the program DNARates (Olsen, G.J., Pracht, S., and Overbeek, R., unpublished) based on the phylogenetic tree inferred by ML analysis of combined AK and COXI sequences as described above (hereon referred to as the combined ML tree) . However, rather than convert sequence divergence estimates to estimates of absolute time as in Townsend (2007), we calculated PI and RPI at each node in the combined ML tree by the following procedure. We used the fixed tree option of PHYML to estimate branch lengths separately for AK and COXI on the combined ML tree. We then used a Perl script (available upon request) and the Bioperl libraries (Stajich et al., 2002) to calculate PI for AK and COXI and RPI for AK from the rate estimates generated by DNARates, the combined ML tree, and the estimated branch lengths for AK and COXI. For each internal node in the combined ML tree, the script used equation 10 in Townsend (2007) to calculate PI values for AK and COXI from node height (distance to leaf node) and the site-specific rate estimates. The RPI of AK at each internal node was calculated as the percent of the total PI contributed by AK.

3. Results

3.1. PCR amplification and sequencing

During the initial development and testing of PCR primers for AK we found it helpful to use undegraded samples of genomic DNA (as judged by gel electrophoresis). In many instances we could not amplify AK sequences of moderate length (1–2 kb) from DNA extracted from either alcohol-preserved or frozen tissues, although we were able to amplify COXI sequences from these samples. In contrast, we have achieved a high success rate with DNA isolated from fresh field-collected specimens. Thus our preferred procedure has become to perform DNA extractions on fresh material at field sites.

We initially attempted to amplify the entire coding region of the AK gene with overlapping pairs of primers. However, we discovered a previously unreported intron that varied in size among taxa (data not presented here). We therefore decided to limit our analysis to the larger, second exon. With two pairs of primers, we amplified overlapping regions that encode the majority of the carboxy-terminal domain of the AK protein (Fig. 1).

Partial sequences of AK were determined for 32 brachyuran species and two outgroup species, and for a subset of 28 of those species for COXI (Table 1). All sequences lacked stop codons and were aligned without indels. Thus all sequences appeared to be functional protein-coding genes with no indications of being pseudogenes.

3.2. Phylogenies based on AK

Based on the Akaike Information Criterion (AIC) (Akaike, 1974) MrAIC selected the General Time Reversible model with invariant sites and a gamma distribution of rates across sites (GTRIG) as the most appropriate model of sequence evolution. With this model, maximum likelihood (ML) and Bayesian (BA) methods yielded phylogenies with similar topologies, although BA posterior probabilities (BPP) were generally higher than ML bootstrap values (MLBS) (Fig. 2). For five recognized genera (*Cancer, Menippe, Mithraculus, Pachygrapsus* and *Uca*), our taxon sample included multiple species from each genus. In the AK phylogeny, these genera were monophyletic with MLBS between 94 and 100, and BPP of 100. For seven recognized families our taxon sample included multiple genera within each family. Three of these families (Panopeidae, Sesarmidae, and Varunidae) were monophyletic, with MLBS between 93 and 100, and BPP of 100. The Portunidae were polyphyletic; three of the four species were sister to a group that included both *Carcinus maenas* (Portunidae) and the genus *Cancer* (Cancridae). However, the grouping of *Carcinus* with *Cancer* was not strongly supported (MLBS 53, BPP 71). In our taxon sample the superfamily Xanthoidea was represented by the Menippidae and Panopeidae and weakly supported by BA (BPP of 66). The Panopeidae were resolved with strong support (MLBS 100, BPP 100), although the Menippidae were not because the placement of *Eriphia gonagra* was unresolved. There was strong support for the superfamily Majoidea (MLBS 99, BPP 97). However, two families within the Majoidea were polyphyletic: the Mithracidae because

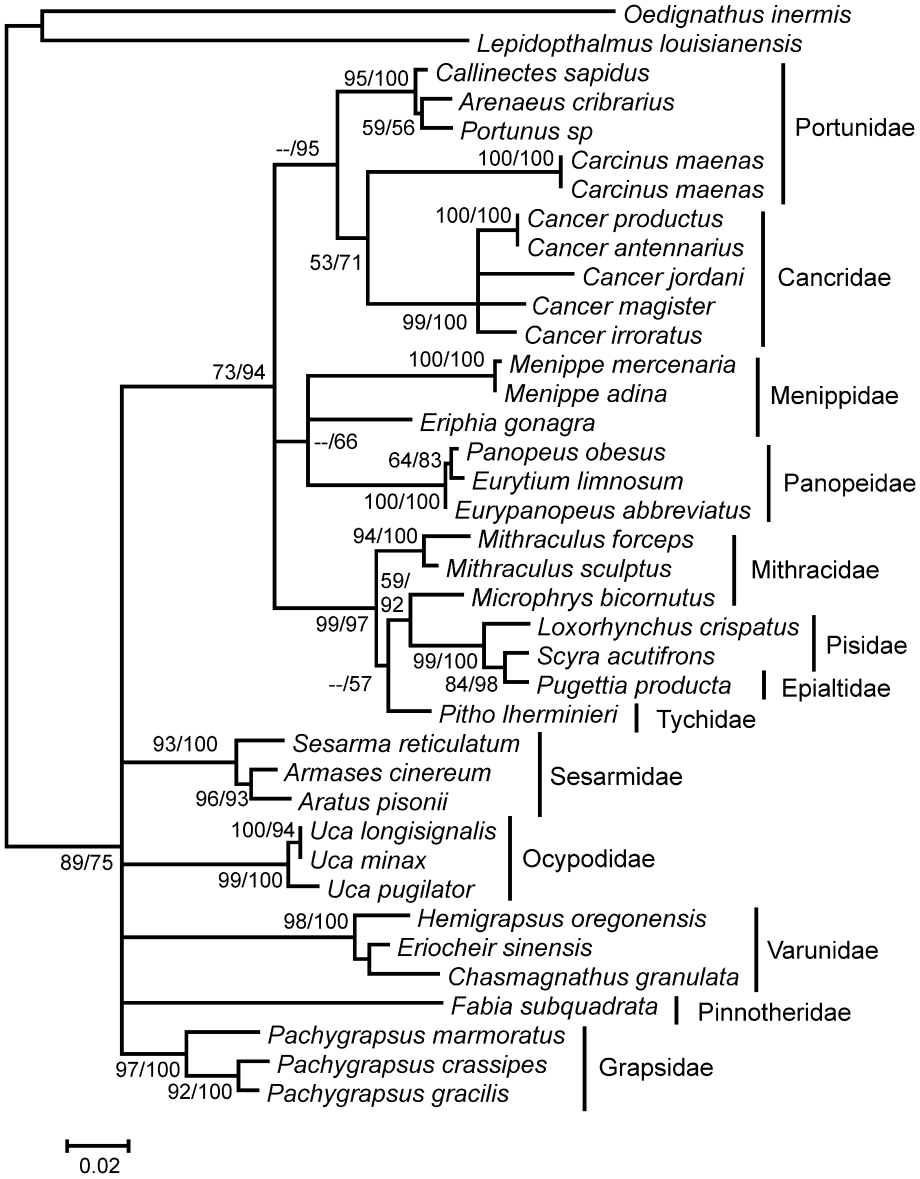

Fig. 2. Phylogeny inferred by maximum likelihood and Bayesian analysis for brachyuran arginine kinase sequences. Numbers at each branch indicate maximum likelihood bootstrap support/Bayesian posterior probabilities.

of the placement of *Microphyrs bicornutus* with the Pisidae, and the Pisidae because of the sister relationship of *Scyra acutifrons* with the epialtid *Pugettia producta*. The superfamily Grapsoidea (represented by Grapsidae, Sesarmidae and Varunidae) was unresolved. The brachyuran subsection Heterotremata was resolved with a MLBS of 73 and a BPP of 94; however the subsection Thoracotremata was not resolved.

3.3. Phylogenies based on COXI

MrAIC selected the GTRIG model as the best model of sequence evolution for the COXI sequences. ML and BA methods reconstructed congruent phylogenies, but differed, sometimes considerably, in support values for particular branches (Fig. 3). For example, BA resolved a majoid clade with a BPP of 94, while the MLBS for this clade was only 16. In comparison with the phylogeny based on AK, there was generally lower support for individual branches and fewer branches were resolved by the criterion of support values above 50. This lack of resolution and lower support with COI was not due solely to the reduced number of taxa used for the COI phylogeny. Similar differences were seen when AK and COXI trees were constructed for the same set of taxa (results not shown).

Of the three groups found in the AK phylogeny that were in conflict with current taxonomic assignments to families, two of these were also found with support values above 50 in the COXI phylogeny: (1) the group *Carcinus maenas* + *Cancer* spp., which conflicts with assignments of these species to Portunidae and Cancridae respectively; and (2) the sister relationship of *S. acutifrons* (Pisidae) with *P. producta* (Epialtidae) rather than its confamilial *L. crispatus* (Pisidae).

3.4. Phylogenies based on combined data

ML and BA analyses of combined AK and COXI sequences were performed for the set of taxa that were common to both data sets (in most cases, the same specimens were used). MrAIC indicated the GTRIG model was the most suitable for the concatenated sequences. The ML and BA trees for the combined data were very

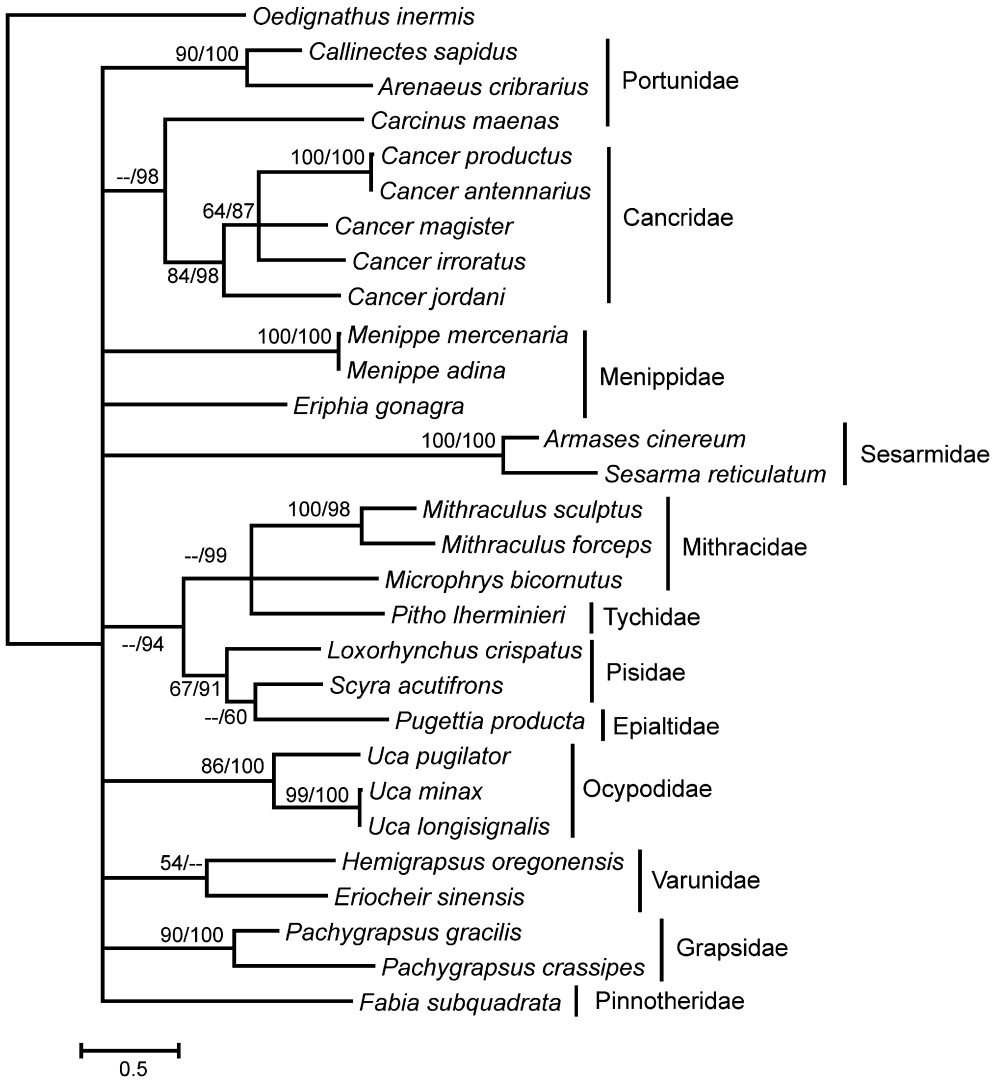

Fig. 3. Phylogeny inferred by maximum likelihood and Bayesian analysis for brachyuran cytochrome oxidase I sequences. Numbers at each branch indicate maximum likelihood bootstrap support/Bayesian posterior probabilities.

similar, differing mainly in support values for particular branches (Fig. 4). These trees were also generally concordant with those based on AK and COXI sequences separately, but with a few differences. Support values were generally higher for the combined data, which resulted in resolution of additional relationships. A sister relationship between *C. magister* and *C. productus* + *C. antennarius* was resolved with MLBS of 61 but a BPP below 50. The family Menippidae was resolved, although with low support (MLBS 61, BPP 63). The Sesarmidae appear as the sister taxon to the Ocypodidae, rather than to either of the two grapsoid families (BPP 62). In conflict with the phylogeny inferred from AK sequences alone, a sister relationship between *M. bicornutus* and *P. lherminieri* was resolved by ML (MLBS 58) but not BA.

3.5. Phylogenetic informativeness

The estimated rate of sequence divergence for COXI (averaged across sites) was six times higher than for AK. This suggests that in comparison with COXI, the phylogenetic relationships for which AK is most informative should be deeper. Profiles of phylogenetic informativeness (PI) against ML estimates of sequence divergence for AK and COI revealed that these sequences also differ in the sensitivity of PI to sequence divergence (Fig. 5). COXI sequences are most informative across a much narrower range of sequence divergence than are AK sequences. Thus while the PI of COXI is 9.4 times higher than the PI of AK at a sequence divergence of 0.02, for sequence divergence above 0.34, the PI of AK is consistently at least 1.5 times higher than the PI of COXI.

Direct comparison of the PI of AK and COXI data at individual nodes was made by calculation of the relative phylogenetic informativeness (RPI) of AK for each node of the combined ML tree (Fig. 6). We found RPI values to be especially sensitive to estimates of branch lengths for COXI because of the strong dependency of the PI of COXI on sequence divergence (Fig. 5). At many nodes, node heights for COXI were well above the range for which COXI sequences are most informative, while node heights for AK were lower and within the optimal range for AK sequences. In this context, we note that branch lengths estimated by PHYML for COXI on the combined ML tree were longer than those estimated without the fixed tree option of PHYML (Fig. 3).

For some nodes corresponding to genera, the RPI of AK was less than 50 and as low as zero, which suggests COXI should be more informative for resolution of relationships within those genera, although AK could be more informative for the resolution of the genera themselves. For deeper relationships, RPI was generally

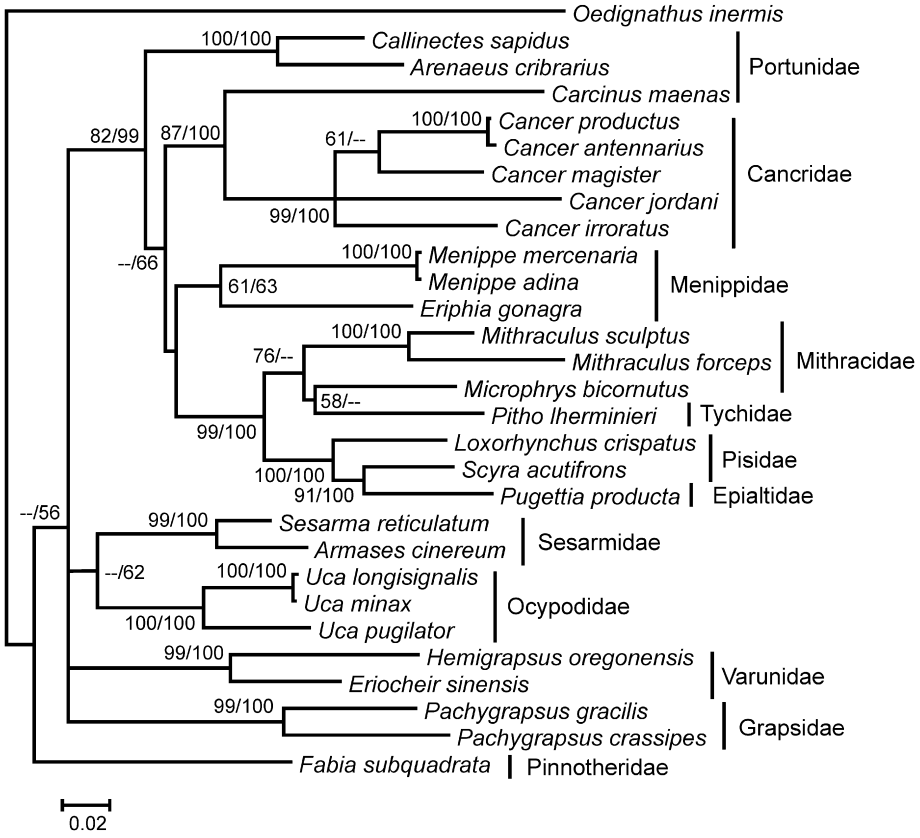

Fig. 4. Phylogeny inferred by maximum likelihood and Bayesian analysis for combined arginine kinase and cytochrome oxidase I sequences. Numbers at each branch indicate maximum likelihood bootstrap support/Bayesian posterior probabilities.

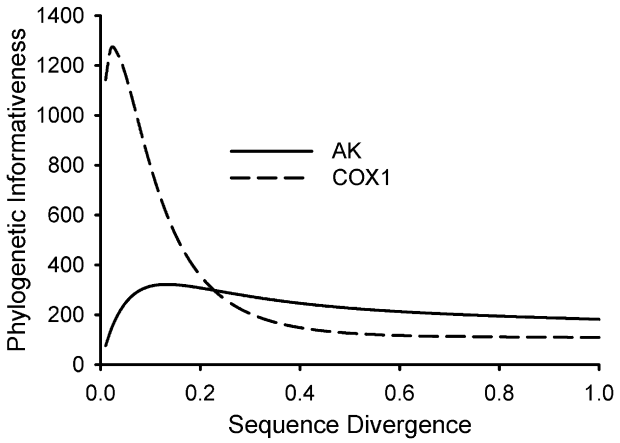

Fig. 5. Profiles of phylogenetic informativeness for brachyuran arginine kinase and cytochrome oxidase I sequences.

above 50, indicating AK should be more informative. Comparisons of these RPI values with the results of our ML and BA phylogenetic analyses are complicated by two factors. First, we might expect that a measure of phylogenetic informativeness would be less relevant to ML or BA analysis than to maximum parsimony (MP) analysis, because MP depends entirely on phylogenetically informative characters, while ML and BA do not. Second, the number of AK sequences we analyzed was greater than the number of COXI sequences, so that differences in support values could have been caused by differences in taxon sampling. We therefore performed MP analyses of AK and COXI sequences for an identical set of 29 taxa to provide the most relevant context in which to examine phylogenetic informativeness. We note that the consistency index for AK sequences (0.467) was nearly twice that for COXI sequences (0.242). Overall, the phylogenies inferred by MP were not as well resolved as those inferred by either ML or BA, especially for deeper relationships (Fig. 7 compared to Figs. 2 and 3). RPI values (Fig. 6) indicated that AK sequences should be more informative than COXI sequences for most of the polytomies in the tree (internal nodes that connect more than two leaf nodes). This was generally true; MP analysis of AK sequences resolved 16 of a possible 28 internal nodes, while MP analysis of COXI sequences resolved only 11 of 28 possible internal nodes.

4. Discussion

The main intent of this investigation was to assess the utility of the arginine kinase (AK) gene for resolution of brachyuran relationships. Although it is generally not possible to perform the ideal assessment, which would be a comparison of phylogenies based on the sequence with the true phylogeny, there are two criteria that are widely employed. First, it is expected that if a phylogeny inferred from sequence data is accurate, it will be at least in rough agreement with other well-supported phylogenies based on other characters. Second, it is desirable that the data provide strong statistical support for phylogenies, which implies robustness if not necessarily accuracy. For our assessments of the phylogenetic utility of AK sequences by the first criterion, we compared phylogenies based on AK sequences with phylogenies based on cytochrome oxidase I (COXI) sequences and with relationships implied by current taxonomic assignments. For the second criterion we used a nonparametric bootstrap of a maximum likelihood (ML) analysis and posterior probabilities from a Bayesian analysis (BA). In addition to these two standard criteria, we explored the use of a recently

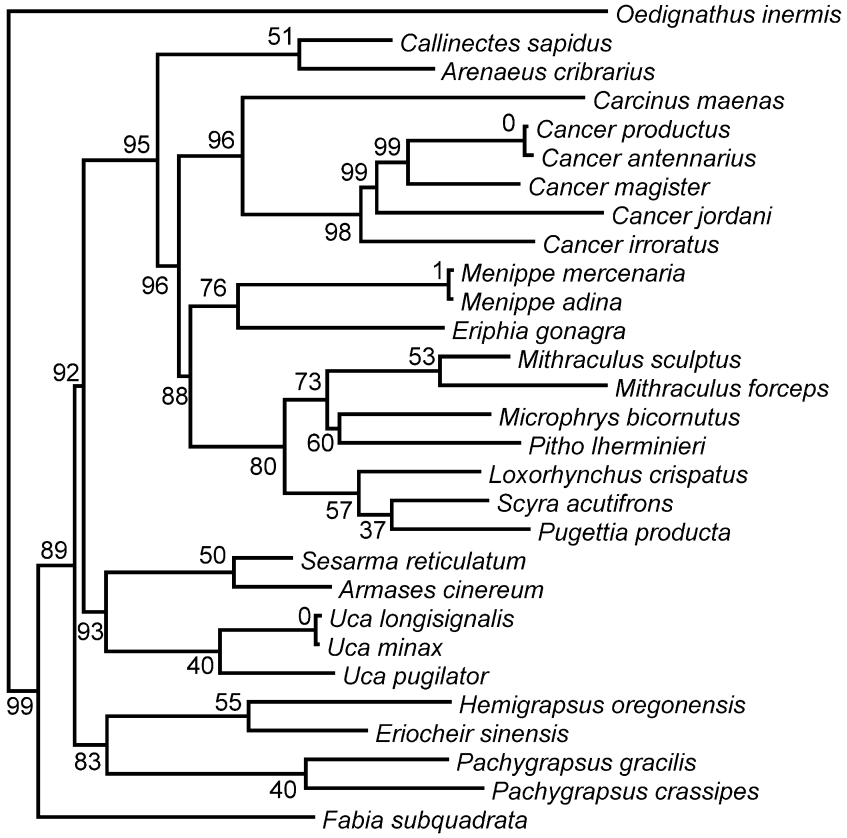

Fig. 6. Relative phylogenetic informativeness of arginine kinase sequence data for each node of a phylogeny based on combined arginine kinase and cytochrome oxidase I sequences.

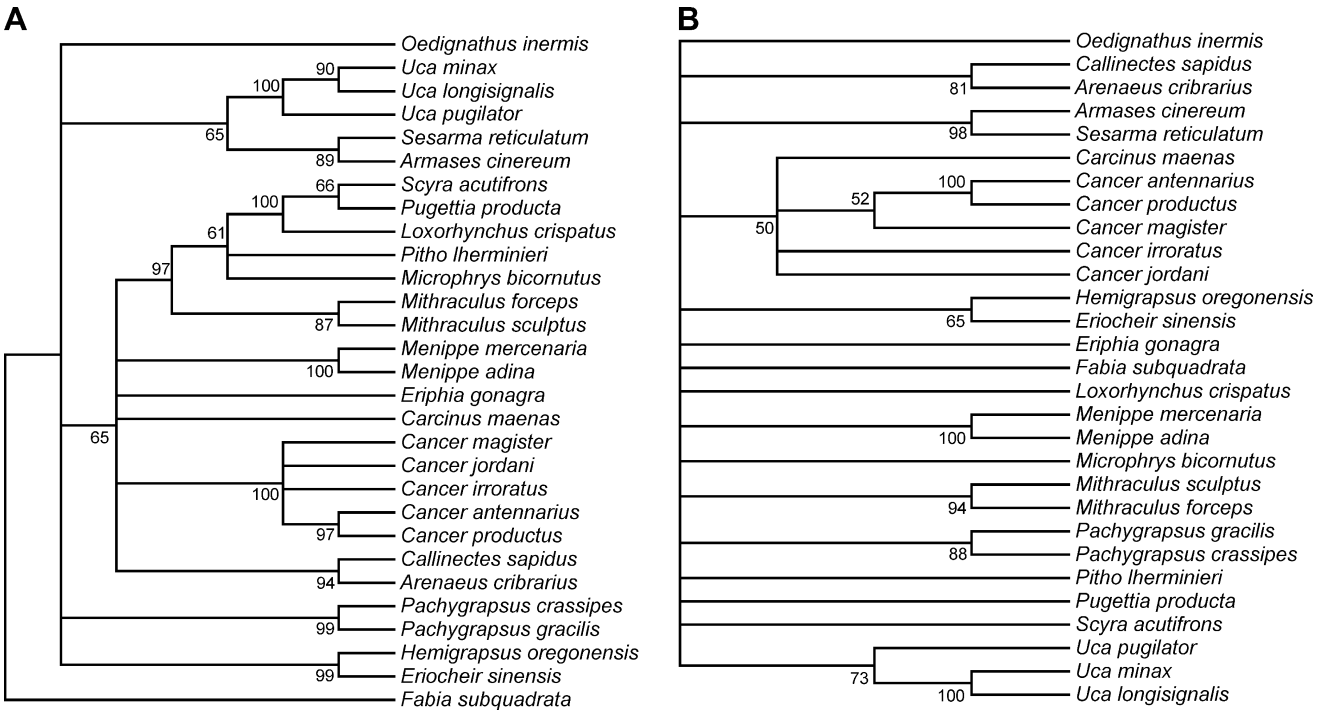

Fig. 7. Phylogenies inferred by maximum parsimony for (A) arginine kinase and (B) cytochrome oxidase I sequences for the same set of taxa. Numbers at each branch indicate bootstrap support.

developed measure of phylogenetic informativeness (Townsend, 2007). This measure is based on the expected number of phylogenetically informative sites in a sequence for nodes at specific depths in a tree.

Phylogenies based on either ML or BA of AK sequences were generally consistent with recently proposed brachyuran taxonomic assignments, but with some exceptions. Concordance was perfect for genera; all five that were represented by multiple species in our taxon sample were placed into monophyletic groups with strong statistical support. However, the AK data did not support the monophyly of four of the seven recognized families represented by multiple genera in our sample. In three of these cases, there was support for polyphyletic relationships rather than simply lack of resolution. One of these cases involved *Carcinus maenas*, which is assigned to Portunidae, but appears as sister to the genus *Cancer* (Cancridae) in our AK phylogenies. The group *C. maenas* + *Cancer* spp. was also strongly supported by our Bayesian analysis of COXI sequences (BPP 98). Although we are not suggesting reassignment of *C. maenas* on the basis of this evidence, we note that it is atypical among portunids in lacking modification of the last pereiopods into flattened swimming legs. A recent cladistic analysis of brachyuran foregut characters (Brosing et al., 2007) supported a close sister relationship between the Portunidae and Cancridae, however it did place *C. maenas* within the Portunidae. An analysis of mitochondrial 16S sequences also indicated a close relationship between these families, and weak support for a polyphyletic Portunidae (Schubart et al, 2000). We suggest that further investigation of these relationships with additional characters and taxa is warranted.

The two other cases of conflict between relationships inferred from AK sequences and current taxonomic assignments involved the families Pisidae, Mitracidae and Epialtidae within the superfamily Majoidea. However these conflicts were not surprising, because evidence for the paraphyly or polyphyly of these nominal families has also been found in larval characters (Marques and Pohle, 2003; Marques et al., 2003; Pohle and Marques, 2000), a combination of 16S, 28S and COXI sequences (Hultgren and Stachowicz, 2008), and the COXI sequences presented here (Fig. 3). There is thus sufficient reason to doubt that these majoid families represent natural monophyletic groups. It appears that AK sequences in combination with other data could be useful in the resolution of these majoid relationships.

The nodes in phylogenies inferred from AK sequences were generally better supported than those based on COXI sequences, and more nodes were resolved by the criterion of support values above 50. This is noteworthy because COXI is among the most commonly used genes for brachyuran systematics; our results suggest that AK would be a better choice for single-gene phylogenies. However, neither COXI nor AK sequences were sufficient to resolve every node. Better results were obtained with the two genes in combination, but a few nodes remained unresolved. These results highlight the need for multiple genes for accurate and robust inference of brachyuran phylogenies.

Townsend's (2007) measure of phylogenetic informativeness is intended to measure the contribution of different sequences to the resolution of nodes (soft polytomies) at specific depths in a phylogenetic tree. Plots of phylogenetic informativeness for AK and COXI were strikingly different. The plot for AK was nearly level over a broad range of sequence divergence, while the plot for COXI displayed a very sharp peak at low sequence divergence. This reflects differences in the rate and pattern of base substitution for the two sequences. Overall, AK evolves at a moderate rate with a relatively broad range of rates across sites. COXI exhibits a rapid rate of synonymous substitutions, but much lower rates of nonsynonymous substitutions. These contrasting patterns suggest various strategies for combining data from different sequences to achieve specific ends. For resolution of polytomies within a narrow range of phylogenetic depths (such as might result from a rapid radiation), sequences could be selected that are narrowly informative for nodes at those depths. In contrast, for resolution of polytomies across a wide range of phylogenetic depths sequences like AK, with broad ranges of substitution rates, could be used or sequences that evolve at different rates could be combined to produce an overall broad distribution.

Relative phylogenetic informativeness (RPI) of AK was a rough, but not exact predictor of which sequence (AK or COXI) was best at resolving individual polytomies (nodes with more than two descendant taxa) by MP analysis. In general, most of the polytomies had RPI values above 50 (Fig. 6). As might be predicted from these higher RPI values, more polytomies were resolved with AK than with COXI (Fig. 7). However, RPI did not predict differences in support for the two polytomies resolved by MP analysis of COXI sequences. The RPI for the node that connects *Cancer antennarius*, *C. productus*, and *C. magister* was 99, indicating that AK should be much more informative than COXI. However, this node was resolved by MP analysis of COXI (with weak support), but not AK. The node that connects the three species of *Uca* had an RPI of 40, although BS support was higher for AK (100) than for COXI (73). We suggest that these discrepancies could result from statistical errors in the estimation of phylogenetic informativeness, the inherent stochasticity in the occurrence of parsimony-informative sites, or non-phylogenetic signals that could influence bootstrap values. It would be useful to examine the relationship between bootstrap values and PI in additional cases to assess the latter as a predictive tool.

It is our hope that AK will be added to the set of genes routinely used for crustacean systematics. This would help to reduce the current bias towards mitochondrial and ribosomal RNA genes, which are not always well suited for phylogenetic analysis (Neigel and Mahon, 2007). We also hope that the use of other nuclear, protein-coding genes will be explored with the goal of providing crustacean systematists with at least half a dozen on which to base phylogenies. In anticipation of an increase in demand for nuclear sequences in crustacean systematics, we also recommend that care be taken to obtain good quality DNA from fresh material whenever possible.

Acknowledgments

We thank Stephen Morgan, Jennifer Fisher, Kristin Hultgren, Darryl Felder, and Allison Manning for providing some of the specimens used in this study. The manuscript benefited from critical comments by Caryl Chlan, Scott France, Darryl Felder, Kristin Hultgren, Christoph Schubart and two anonymous reviewers. This study was funded by the National Science Foundation (0315995) and a travel grant from the Smithsonian Tropical Research Institute, Panama.

References

Ahyong, S.T., O'Meally, D., 2004. Phylogeny of the Decapoda Reptantia: resolution using three molecular loci and morphology. Raffles Bull. Zool. 52, 673–693.
Akaike, H., 1974. A new look at statistical model identification. IEEE Trans. Autom. Contr. AC 19, 716–723.
Bessman, S.P., Capenter, C.L., 1985. The creatine–creatine phosphate energy shuttle. Annu. Rev. Biochem. 54, 831–862.
Brosing, A., Richter, S., Scholtz, G., 2007. Phylogenetic analysis of the Brachyura (Crustacea, Decapoda) based on characters of the foregut with establishment of a new taxon. J. Zoolog. Syst. Evol. Res. 45, 20–32.
Caterino, M.S., Cho, S., Sperling, F.A.H., 2000. The current state of insect molecular systematics: a thriving Tower of Babel. Annu. Rev. Entomol. 45, 1–54.
Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J.D., 2003. Multiple sequence alignment with the clustal series of programs. Nucleic Acids Res. 31, 3497–3500.
Cunningham, C.W., Blackstone, N.W., Buss, L.W., 1992. Evolution of king crabs from hermit crab ancestors. Nature 355, 539–542.
Ellington, W.R., 2001. Evolution and physiological roles of phosphagen systems. Annu. Rev. Physiol. 63, 289–325.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hardy, N.B., 2007. Phylogenetic utility of dynamin and triose phosphate isomerase. Syst. Entomol. 32, 396–403.

Hassanin, A., Leger, N., Deutsch, J., 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. Syst. Biol. 54, 277–298.

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B Biol. Sci. 270, 313–321.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Hultgren, K.M., Stachowicz, J.J., 2008. Molecular phylogeny of the brachyuran crab superfamily Majoidea indicates close congruence with trees based on larval morphology. Mol. Phylogenet. Evol.. doi:10.1016/j.ympev.2008.05.004.

Kim, W., Abele, L.G., 1990. Molecular phylogeny of selected decapod crustaceans based on 18S ribosomal RNA nucleotide sequences. J. Crust. Biol. 10, 1–13.

Kotlyar, S., Weihrauch, D., Paulsen, R.S., Towle, D.W., 2000. Expression of arginine kinase enzymatic activity and mRNA in gills of the euryhaline crabs *Carcinus maenas* and *Callinectes sapidus*. J. Exp. Biol. 203, 2395–2404.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinformatics 5, 150–163.

Marques, F.P.L., Pohle, G., 2003. Searching for larval support for majoid families (Crustacea: Brachyura) with particular reference to Inachoididae Dana, 1851. Invertebr. Reprod. Dev. 43, 71–82.

Marques, F.P.L., Pohle, G.W., Vrbova, L., 2003. On the larval stages of *Macrocoeloma diplacanthum* (Decapoda: Brachyura: Majidae), and a review of mithricine phylogenetic aspects. J. Crust. Biol. 23, 187–200.

Martin, J.W., Davis, G.E., 2001. An Updated Classification of the Recent Crustacea. Natural History Museum of Los Angeles County.

McLaughlin, P., Camp, D., Angel, M., Bousfield, E., Brunel, P., Brusca, R.C., Cadien, D., Cohen, A., Conlan, K., Eldredge, L., Felder, D., Goy, J., Haney, T., Hann, B., Heard, R., Hendrycks, E., Hobbs III, H.H., Holsinger, J., Kensley, B., Laubitz, D., LeCroy, S., Lemaitre, R., Maddocks, R., Martin, J., Mikkelsen, P., Nelson, E., Newman, W., Overstreet, R., Poly, W., Price, W., Reid, J., Robertson, A., Rogers, D., Ross, A., Schotte, M., Schram, F., Shih, C., Watling, L., Wilson, G., Turgeon, D., 2005. Common and Scientific Names of Aquatic Invertebrates from the United States and Canada: Crustaceans. American Fisheries Society, Bethesda, Maryland.

Meyer, R., Brown, T., Krilowicz, B., MJ, K., 1986. Phosphagen and intracellular pH changes during contraction of creatine depleted rat muscle. Am. J. Physiol. Cell Physiol. 250, C264–C274.

Meyer, R., Sweeney, H., Kushmerick, M., 1984. A simple analysis of the "phosphocreatine shuttle." Proceedings of the Biological Society of Washington. Am. J. Physiol. Cell Physiol. 246, C365–C377.

Neigel, J.E., Mahon, B., 2007. Molecular approaches in Crustacean evolutionary ecology. In: Duffy, J.E., Thiel, M. (Eds.), Evolutionary Ecology of Social and Sexual Systems: Crustaceans as Model Organisms. Oxford University Press, Oxford, pp. 48–67.

Nylander, J.A.A., 2004. MrAIC.pl. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University.

Perez-Losada, M., Bond-Buckup, G., Jara, C.G., Crandall, K.A., 2004. Molecular systematics and biogeography of the southern South American freshwater "crabs" Aegla (Decapoda : Anomura : Aeglidae) using multiple heuristic tree search approaches. Syst. Biol. 53, 767–780.

Perez-Losada, M., Jara, C.G., Bond-Buckup, G., Crandall, K.A., 2002. Phylogenetic relationships among the species of Aegla (Anomura: Aeglidae) freshwater crabs from Chile. J. Crust. Biol. 22, 304–313.

Pohle, G., Marques, F., 2000. Larval stages of *Paradasygyius depressus* (Bell, 1835) (Crustacea: Decapoda: Brachyura: Majidae) and a phylogenetic analysis for 21 genera of Majidae. Proc. Biol. Soc. Wash. 113, 739–760.

Porter, M.L., Perez-Losada, M., Crandall, K.A., 2005. Model-based multi-locus estimation of decapod phylogeny and divergence times. Mol. Phylogenet. Evol. 37, 355–369.

Regier, J.C., Shultz, J.W., 2001. Elongation factor-2: a useful gene for arthropod phylogenetics. Mol. Phylogenet. Evol. 20, 136–148.

Rychlik, W., Rhoads, R.E., 1989. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. Nucleic Acids Res. 17, 8543–8551.

Schubart, C.D., Neigel, J.E., Felder, D.L., 2000. Use of the mitochondrial 16S rRNA gene for phylogenetic and population studies of Crustacea. Crustac. Issues 12, 817–830.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.

Spears, T., Abele, L.G., Kim, W., 1992. The monophyly of brachyuran crabs: a phylogenetic study based on 18S rRNA. Syst. Biol. 41, 446–461.

Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G.R., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D., Birney, E., 2002. The bioperl toolkit: Perl modules for the life sciences. Genome Res. 12, 1611–1618.

Suzuki, T., Kamidochi, M., Inoue, N., Kawamichi, H., Yazawa, Y., Furukohir, T., Ellington, W.R., 1999. Arginine kinase evolved twice: evidence that echinoderm arginine kinase originated from creatine kinase. Biochem. J. 340, 671–675.

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods), version 4.0. Sinauer Associates, Sunderland, Massachusetts.

Townsend, J.P., 2007. Profiling phylogenetic informativeness. Syst. Biol. 56, 222–231.

Tudge, C.C., Cunningham, C.W., 2002. Molecular phylogeny of the mud lobsters and mud shrimps (Crustacea: Decapoda: Thalassinidea) using nuclear 18S rDNA and mitochondrial 16S rDNA. Invertebr. Syst. 16, 839–847.

Walliman, T., M, W., D, B., Nicolay, K., Eppenberger, H., 1992. Intracellular compartmentation, structure and function of creatine kinase isoenzymes in tissues with high and fluctuating energy demands: the "phosphocreatine" circuit for cellular energy homeostasis. Biochem. J. 281, 21–40.

Williams, S.T., Knowlton, N., Weigt, L.A., Jara, J.A., 2001. Evidence for three major clades within the snapping shrimp genus *Alpheus* inferred from nuclear and mitochondrial gene sequence data. Mol. Phylogenet. 20, 375–389.