Decapod Phylogenetics And Molecular Evolution

ALICIA TOON, MAEGAN FINLEY, JEFFREY STAPLES & KEITH A. CRANDALL

Department of Biology, Brigham Young University, Provo, UT 84602, U.S.A.

ABSTRACT

Decapoda is the most species-rich group of crustaceans, with numerous economically important and morphologically diverse species leading to a large amount of research. Our research groups are attempting to estimate a robust phylogeny of the Decapoda based on molecular and morphological data to resolve the relationships among the major decapod lineages and then to test a variety of hypotheses associated with the diversity of decapod morphological evolution. Thus, we have developed a database of molecular markers for use at different scales of the evolutionary spectrum in decapod crustaceans. We present potential mitochondrial and nuclear markers with an estimation of variation at the genus level, family level, and among infraorders for Decapoda. We provide a methodological framework for molecular studies of decapod crustaceans that is useful at different taxonomic levels.

1 MOLECULAR TAXONOMY

There are several competing hypotheses concerning the relationships of the major lineages of Decapoda based on morphological estimates of phylogeny. Early taxonomy of the decapods was largely based on the mode of locomotion; taxa were divided into the swimming lineages (Natantia) and the crawling lineages (Reptantia) (Boas 1880). Morphological and molecular studies suggest Natantia is paraphyletic; it is presently classified based on gill structure (Burkenroad 1963, 1981) dividing Decapoda into the suborders Dendrobranchiata (penaeoid and sergestoid shrimps) and Pleocyemata (all other decapod crustaceans). Relationships within Pleocyemata are still controversial and remain unresolved. As morphological data, both recent and fossil, and genetic data continue to accumulate, we are moving towards phylogenetic resolution of these controversial relationships. Here we present a progress report for the Decapoda Tree of Life effort and the tools with which we will continue our analysis of decapod crustacean phylogenetic relationships.

Several recent hypotheses based on combined analysis of morphological and molecular data or molecular data alone suggest that resolving the systematics of this group is a difficult task (see Fig. 1). There is agreement among these studies that Dendrobranchiata represents a basal lineage within the decapod crustaceans and that within Pleocyemata the Caridea and Stenopodidea are basal infraorders (Porter et al. 2005; Tsang et al. 2008). Molecular research also supports the removal of polychelids from Palinura following Scholtz and Richter (1995) and its establishment as a separate infraorder (Polychelida) (Tsang et al. 2008; Ahyong this volume). Relationships among reptant decapods remain unresolved by the addition of molecular data. Several recent phylogenetic analyses incorporating mitochondrial and nuclear data (Robles et al. this volume) or nuclear data alone (Tsang et al. 2008; Chu et al. this volume) suggest Thalassinidea are not monophyletic but rather

may represent several infraorders. The timeline of diversification among the reptant decapods or specifically whether Astacidea (Porter et al. 2005) or the Anomura/Brachyura lineages (Ahyong & O'Meally 2004; Tsang et al. 2008) are the most recently derived lineages remains a question of interest.

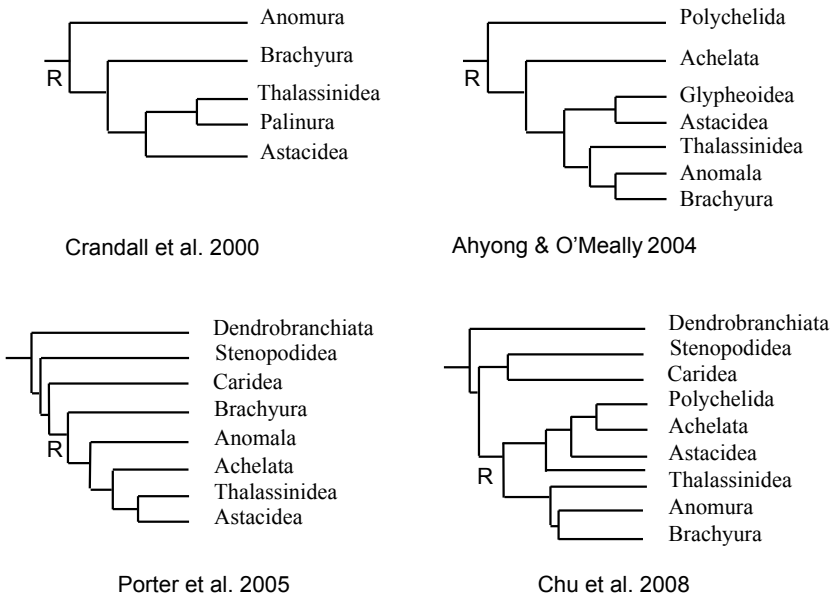

Figure 1. Hypotheses of decapod evolutionary relationships based on molecular data. R shows the position of the reptant decapods.

2 DEVELOPING GENETIC MARKERS FOR MOLECULAR PHYLOGENY

The order Decapoda includes roughly 175 families (extant and extinct) and more than 15,000 described species. Complicating things further are the estimated 437 million years since the origin of the Decapoda with the major lineages estimated to have been established by 325 million years ago (Porter et al. 2007). Constructing a molecular phylogeny across such breadth of taxa and depth of timescale requires serious consideration of markers that have enough variation to reconstruct relationships at the fine scale (at and within the family level) as well as being conservative enough to be used across infraorders representing these deeper timescales. Our approach is to accumulate molecular sequence data for different gene regions including both mitochondrial and nuclear genes, coding and non-coding. In this way, we will be able to maximize data at deeper nodes where alignment of sequence data is most difficult while retaining information among families and between the most recently diverged taxa.

There are two molecular approaches to amplifying sequence data for use in phylogenetic studies. (1) Isolation of RNA from tissues, coupled with reverse transcription-polymerase chain reaction (RT-PCR) to amplify target genes or gene fragments, reduces problems associated with amplification of pseudogenes (non-coding duplicated gene segments) and sequencing through large introns. The

main limitation of RNA work is that fresh tissues, or at least tissues collected in an RNA preserving agent such as RNA*later*, require rapid transfer to -80°C storage. (2) Phylogenetic work using genomic tissue extractions and amplifications are still favored over RNA techniques due to lower costs, ease of field sampling, and the ability to use previously collected specimens in ethanol. To reduce the risk of sequencing multiple copy genes or pseudogenes, gene fragments are first cloned to identify the number of copies that a primer set amplifies. Although this is not the focus of this paper, in the course of looking for useful phylogenetic markers we have sequenced a number of multigene families such as hemocyanin, actin and opsins. These markers may be phylogenetically useful if a single gene is isolated and amplified. They also have many uses when looking at genome evolution and the expression of these genes in Decapoda (e.g., Porter et al. 2007; Scholtz this volume). However, one must be certain that the same copy is being amplified across taxa for useful phylogenetic results.

Introns or highly variable regions need to be considered when sequencing as they can be large (greater than 1000 base pairs in length) and include repeat regions in some taxa making amplification and sequencing difficult. Often there is too much variation in the intron among taxa to be aligned and included in the analysis. Introns can be avoided by first identifying their position and then designing primer sets within the exon to remove the introns. Here we redesigned primers for elongation factor 2 (EF-2) and transmembrane protein (TM9sf4) to exclude regions of high variability of approximately 300 base pairs in EF2 and 500-1000 base pairs in TM9sf4. Although this reduced the total length of sequence amplified, the highly variable regions produce a greater noise to signal ratio at the higher phylogenetic relationships, our principle focus. Of course, these more variable introns might become very useful for population genetic and species level phylogenetic work, and we continue to explore their utility at these lower levels of diversity.

3 THE GENES AND THEIR DIVERSITY

3.1 *Mitochondrial genes: 12S, 16S and COI*

Mitochondrial ribosomal genes 12S and 16S and coding genes such as COI have been extremely useful in population genetic and systematic studies. Mitochondrial markers have been favored in studies for several reasons (see Schubart, this volume, for details and proposed primer sets for decapod mtDNA amplification). The high copy number of mitochondria in tissues makes them relatively easy to isolate. They are haploid and maternally inherited and consequently are one quarter the effective population size of nuclear genes (Moritz et al. 1987), thus allowing population level studies and systematic studies among recently diverged taxa. Possibly the most important reason to use mitochondrial genes is the availability of universal mtDNA primer sets that have minimized laboratory time in the initial setting up of a project. Finally, there is already an extensive set of nucleotide sequences from these genes in GenBank, as they have been the staple for crustacean molecular phylogenetic work since its inception.

To provide a comparison of gene utility, we have included uncorrected divergence estimates between pairs of taxa: between species, between genera, between families, and between infraorders/ suborders for a number of genes. We also included COI on each graph as a reference (see Figs. 2 - 5). The ribosomal mitochondrial genes show similar levels of divergence to each other across all comparisons. In 12S, divergence estimates range from 3.9% among *Euastacus* species, 18% among genera within Parastacidae, 18.6% among families of Astacidea, and up to 24.2% among infraorders of Pleocyemata. Divergence of 16S ranges from 3.5% among species, 17.6% among genera, 23.5% among families and up to 26.2% among infraorders of Pleocyemata. The coding

mitochondrial gene COI is highly variable among species, thus making it a good candidate at lower levels. High divergence estimates were found above and including the family level, suggesting that this gene may have problems of nucleotide saturation above this level. This gene may still be useful for phylogenetic inference for resolving deeper nodes, however it is important to test for saturation and consider this in the analysis (i.e., use a model of evolution that incorporates multiple mutations at the same site — see Palero & Crandall this volume). A disadvantage of mitochondrial markers is that they are effectively a single locus, and, when used alone, they may not represent the true species tree.

Another problem of some mitochondrial genes such as COI is the presence of pseudogenes (nuclear copies of mitochondrial genes) in some species of decapods (Song et al. 2008).

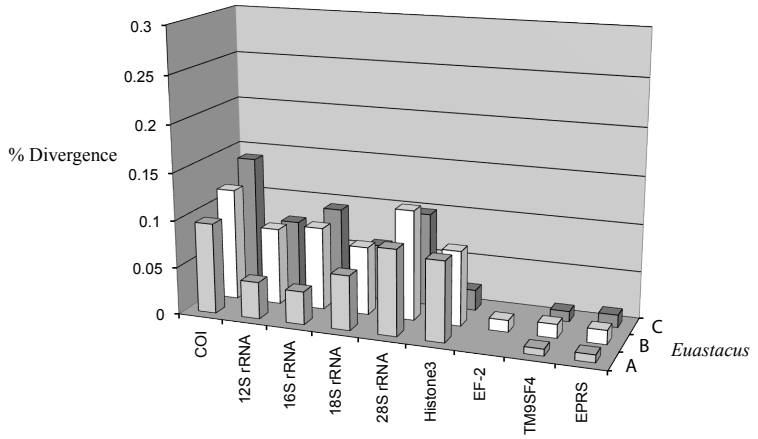

Figure 2. Pairwise divergence estimates between species of *Euastacus* (Astacidea) for mitochondrial and nuclear genes. Species are A: *E. eungella* and *E. spinichelatus*, B: *E. robertsi* and *E. eungella*, C: *E. robertsi* and *E. spinichelatus*.

3.2 *Nuclear genes*

Use of nuclear genes in addition to mitochondrial genes adds to the number of independent markers in a dataset, thus increasing the chances of reconstructing the true species phylogeny. In addition, a larger effective population size, and on average, a lower substitution rate (Moriyama & Powell 1997), results in nuclear genes evolving slower than mitochondrial genes. Consequently, they may be better at resolving deeper phylogenetic nodes (see Chu et al. this volume). There are several considerations when choosing nuclear markers. There are at least two copies of each gene, although this is not usually a problem for phylogenetic studies as variation within an individual is less than between species. However, as mentioned previously, many genes belong to multigene families where duplications have resulted in genes or domains with a similar nucleotide sequence. In order to establish single copy or at least the amplification of one dominate copy for new primer sets (EF-2, EPRS, TM9sf4) presented here, we analyzed 16-24 clones in several taxa representing Pleocyemata (Astacidea (*Homarus americanus*), Brachyura (*Cancer* sp.)) and Dendrobranchiata (*Penaeus* sp.). Low variation among some of the clones was observed. This could be attributed to *taq* polymerase error assuming an error rate of 1.6×10^{-6} to 2.1×10^{-4} per nucleotide per cycle (Hengen 1995) or to very low variation of a diploid gene.

Table 1. Gene regions and primer sets selected for reconstructing the phylogeny of decapod crustaceans. For each primer, details of position (3') and a reference sequence are given. NR (nested reaction) refers to the primers used in the first reaction (1) and subsequent hemi-nested reaction (2).

Gene Region	Primer Name	Primer Sequence (5' – 3')	NR	Position	Reference Sequence	Primer Reference
Mitochondrial Genes						
12S rRNA	12sf	GAA ACC AGG ATT AGA TAC CC		390	AY659990	Mokady et al. 1994
	12sr	TTT CCC GCG AGC GAC GGG CG		778	AY659990	Mokady et al. 1994
16S rRNA	16s-1472	AGA TAG AAA CCA ACC TGG		99	AF200829	Crandall and Fitzpatrick 1996
	16sf-cray	GAC CGT GCK AAG GTA GCA TAA TC		552	AF200829	Crandall and Fitzpatrick 1996
COI	LCO1-1490	GGT CAA CAA ATC ATA AAG ATA TTG		*		Folmer et al. 1994
	HCO1-2198	TAA ACT TCA GGG TGA CCA AAA AAT CA		*		Folmer et al. 1994
Nuclear Genes						
18S rRNA	18s 1f	TAC CTG GTT GAT CCT GCC AGT AG		*		Whiting et al. 1997, Whiting 2002
	18s b3.0	GAC GGT CCA ACA ATT TCA CC		*		Whiting et al. 1997, Whiting 2002
	18s a0.79	TTA GAG TGC TYA AAG C		*		Whiting et al. 1997, Whiting 2002
	18s bi	GAG TCT CGT TCG TTA TCG GA		*		Whiting et al. 1997, Whiting 2002
	18s a2.0	ATG GTT GCA AAG CTG AAA C		*		Whiting et al. 1997, Whiting 2002
	18s 9R	GAT CCT TCC GCA GGT TCA CCT AC		*		Whiting et al. 1997, Whiting 2002
28S rRNA	28s-rD1.2a	CCC SSG TAA TTT AAG CAT ATT A		*		Whiting et al. 1997, Whiting 2002
	28s-rD3a	AGT ACG TGA AAC CGT TCA GG		*		Whiting et al. 1997, Whiting 2002
	28s-rd3.3f	GAA GAG AGA GTT CAA GAG TAC G		*		Whiting et al. 1997, Whiting 2002
	28sA	GAC CCG TCT TGA AGC ACG		*		Whiting et al. 1997, Whiting 2002
	28s-rD4.5a	AAG TTT CCC TCA GGA TAG CTG		*		Whiting et al. 1997, Whiting 2002
	28S rD5a	GGY GTT GGT TGC TTA AGA CAG		*		Whiting et al. 1997, Whiting 2002
	28s-rD4b	CCT TGG TCC GTG TTT CAA GAC		*		Whiting et al. 1997, Whiting 2002
	28S B	TCG GAA GGA ACC AGC TAC		*		Whiting et al. 1997, Whiting 2002
	28s-rD5b	CCA CAG CGC CAG TTC TGC TTA C		*		Whiting et al. 1997, Whiting 2002
	28s-rD6b	AAC CRG ATT CCC TTT CGC C		*		Whiting et al. 1997, Whiting 2002
	28S rD7b1	GAC TTC CCT TAC CTA CAT		*		Whiting et al. 1997, Whiting 2002
	28s3.25a	CAG GTG GTA AAC TCC ATC AAG G		602	AY210833	this study
	28s4.4b	GCT ATC CTG AGG GAA ACT TCG		1594	AY210833	this study
H3	H3 AF	ATG GCT CGT ACC AAG CAG ACV GC		321	AB044542	Colgan et al. 1998
	H3 AR	ATA TCC TTR GGC ATR ATR GTG AC		694	AB044542	Colgan et al. 1998

Table 1 continued.

Gene Region	Primer Name	Primer Sequence (5' – 3')	NR	Position	Reference Sequence	Primer Reference
EF-2	EF2a IF2	TGG GGW GAR AAC TTC TTY AAC		824	EF426560	Porter ML pers. comm.
	EF2a 1R2	ACC ATY TTK GAG ATG TAC ATC AT		1236	EF426560	Porter ML pers. comm.
	EF2a-F978	TGG ANA CBC TGA ARA TCA A	1,2	978	EF426560	this study
	EF2-R1435	GTT ACC HGC TGG VAC RTC TTC	2	1435	EF426560	this study
	EF2-R1536	GAC ACG NWG AAC TTC ATC ACC	1	1536	EF426560	this study
EPRS	192fin1f	+GAR AAR GAR AAR TTY GC		6874	U59923	www.umbi.umd.edu/users/jcrlab/
	192fin2r	+TCC CAR TGR TTR AAY TTC CA		7316	U59923	www.umbi.umd.edu/users/jcrlab/
TM9SF4	3064fin6f	CAR GAR GAR TTY GGN TGG AA	1	1198	NM_014742	www.umbi.umd.edu/users/jcrlab/
	3064fin7r	AAN CCR AAC ATR TAR TA		1841	NM_014742	www.umbi.umd.edu/users/jcrlab/
	3064-F1204	+GAA TTT GGR TGG AAG CTG GT	2	1204	NM_014742	this study
	3064-R1697	+CTG GGN ATY TGG TTG GTT CG	1,2	1697	NM_014742	this study

" * " see primer reference for primer positions. **" + "** addition of M13 primers to the 5' end improves PCR amplification (Regier & Shi 2005)

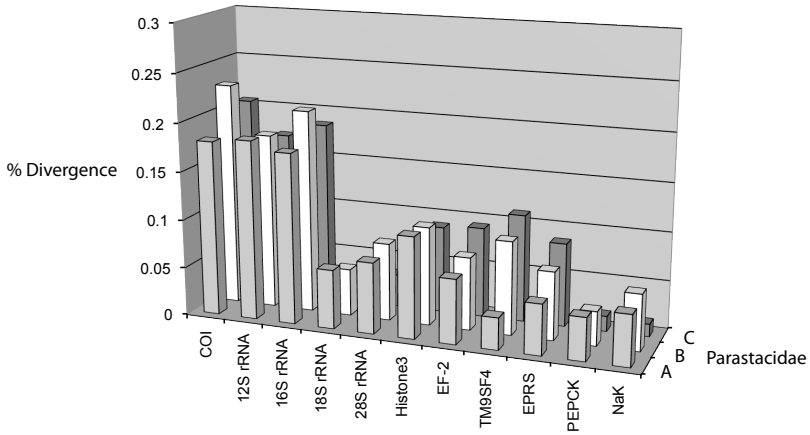

Figure 3. Pairwise divergence estimates between species of Parastacidae (Astacidea) for mitochondrial and nuclear genes. For genes COI, 12S, 16S, 18S, 28S, H3, EF-2, TM9SF4, EPRS the species are A: *Euastacus robertsi* and *Astacoides betsileoensis*, B: *E. robertsi* and *Parastacus defossus*, C: *A. betsileoensis* and *P. defossus*. Species for genes PEPCK and NaK are A: *Homarus gammarus* and *Nephropides caribaeus*, B: *H. gammarus* and *Nephropsis stewarti*, C: *N. caribaeus* and *N. stewarti*.

The ribosomal nuclear genes 18S rDNA and 28S rDNA have been extensively used in arthropod systematics including several decapod studies (e.g., Ahyong & O'Meally 2004; Porter et al. 2005; Mitsuhashi et al. 2007; Ahyong et al. 2007). Rates of evolution vary among and within these genes, making them valuable phylogenetic tools at different taxonomic levels (Hillis & Dixon 1991). We found divergence rates for 18S were consistently moderate among species (5.8 – 7.2%) and among infraorders (5.6%) within Pleocyemata but were higher among the suborders Pleocyemata and Dendrobranchiata (12.8% and 14.1%). Two hypervariable regions of 28S were identified and removed to avoid inflated estimates of divergence among poorly aligned repeat regions. 28S divergence estimates were higher than 18S among species (9.1 – 11.6%), within Pleocyemata (11.3%), and among the suborders (20.8 – 21.8%). Levels of divergence were lower for the intermediate taxon levels, among genera (3.4 – 8.0%) and among families (7.3 – 9.9%), and possibly represented a shorter nucleotide alignment due to indels (insertions or deletions) that are absent among species (within a genus).

Two nuclear protein coding genes that are currently used in arthropod systematics are histone 3 (H3) (e.g., Porter et al. 2005) and elongation factor 2 (EF-2) (e.g., Regier & Shultz 2001). Primer sets already developed for H3 (Colgan et al. 1998) amplify the target fragment across a range of decapod crustaceans and show moderate levels of divergence among species (2.2 – 8.4%), suggesting they are useful nuclear protein coding markers for relationships within a genus. It should be noted that *Euastacus* is relatively older than some decapod genera (see Breinholt et al. this volume) and consequently H3 may not be appropriate for phylogenetic analyses among recently diverged species. Divergence within and among families is also moderate (8.9 – 12.4%), with a higher level of divergence between *Euastacus robertsi* and *Calappa gallus* within Pleocyemata (17%).

Although we were able to amplify genomic fragments of the EF-2 gene with currently designed primer sets (see Table 1), an intron was located at base pair position 860 relative to mRNA in *Libinia emarginata* (GenBank accession AY305506). The intron may be useful for species/genera level studies, although preliminary analysis suggests it is less than 300 base pairs in caridean (Hip-

polytidae) and brachyuran (Calappidae, Leucosiidae, Goneplacidae, Majidae, Cyclodorippidae) decapods. A new forward primer was designed to exclude the intron, and GenBank sequences were downloaded and aligned to design reverse primers 400 – 500 base pairs downstream of the forward primer. Using different primer sets we were able to isolate two copies of EF-2. The two copies were more similar within an individual than between species of *Euastacus* crayfish. Two similar copies of EF-2 are present in *Drosophila melanogaster* (Lasko 2000). The divergence estimates for the longer fragment are presented in figure 2 and were low among species of *Euastacus* (1.3%). Percent divergence within Parastacidae (6.7% – 9.3%) and between families of Astacidea (13.6%) was moderate. High divergences were noted within Pleocyemata between *E. robertsi* and *C. gallus* (18.7%).

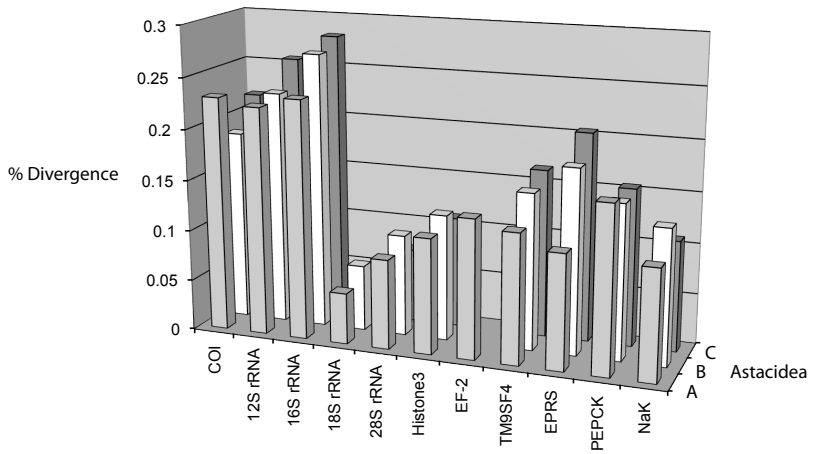

Figure 4. Pairwise divergence estimates among family representatives of Astacidea for mitochondrial and nuclear genes. For genes COI, 12S, 16S, 18S, 28S, H3, EF-2, TM9SF4, EPRS the species are A: *E. robertsi* and *Procambarus clarkii* (TM9SF4: *Orconectes virilis*) B: *E. robertsi* and *Nephropsis aculeata* (COI: *Homarus americanus*) C: *P. clarkii* (TM9SF4: *Orconectes virilis*) and *N. aculeate* (COI: *Homarus americanus*). Species for genes PEPCK and NaK are A: *H. gammarus* and *Cherax quadricarinatus* B: *H. gammarus* and *P. clarkii* C: *C. quadricarinatus* and *P. clarkii*.

The EPRS locus is a potentially useful nuclear gene for reconstructing phylogenetic relationships among the deeper nodes of decapod crustaceans. The EPRS locus encodes a multifunctional aminoacyl tRNA synthetase, glutamyl — prolyl — tRNA synthetase (Cerini et al. 1991). The two proteins are involved in the aminoacylation of glutamic acid and praline tRNA in *Drosophila* (Cerini et al. 1991; Cerini et al. 1997). Few phylogenetic studies have used EPRS, although a recent study of *Paramysis* (Crustacea: Mysida) demonstrates its usefulness in reconstructing relationships among genera of mysids (Audzijonyte et al. 2008). We found divergence levels were low among species of *Euastacus* (0.8% – 1.5%) but moderate for within the family Parastacidae (5.2% – 8.6%) and high between some families of Astacidea (11.3% – 20.5%). This locus showed high divergences within Pleocyemata between *E. robertsi* and *C. gallus* (33.9%) and between *E. robertsi* and *Penaeus* sp. (15.5% – 30.1%). The different levels of divergence at different taxonmic levels suggest this marker may be useful among genera up to order level for phylogenetic estimation.

Transmembrane 9 superfamily protein member 4, or TM9sf4, is a small molecule carrier or transporter. Our study is the first to present divergence estimates and phylogenetic results using

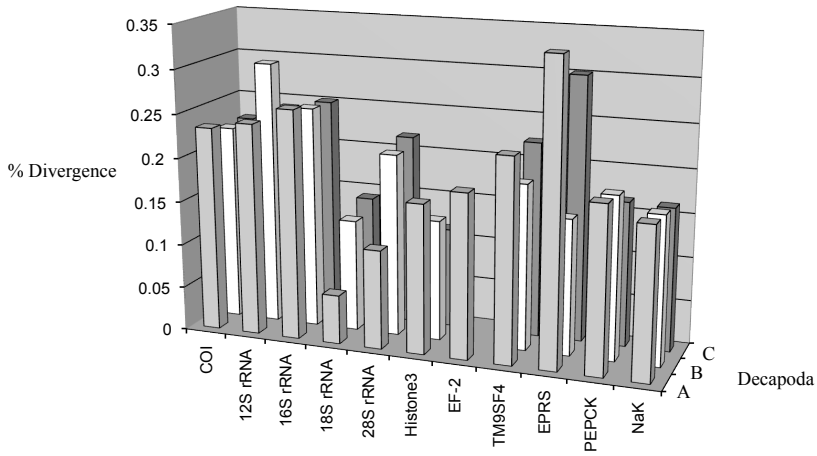

Figure 5. Pairwise divergence estimates among representatives of Decapoda for mitochondrial and nuclear genes. For genes COI, 12S, 16S, 18S, 28S, H3, EF-2, TM9SF4, EPRS the species are A: *E. robertsi* and *Calappa gallus* (COI: *Praebebalia longidactyla*) B: *C. gallus* (COI: *P. longidactyla*) and *Penaeus* sp. C: *E. robertsi* and *Penaeus* sp. Species for genes PEPCK and NaK are A: *H. gammarus* and *Calappa philargius* B: *C. philargius* and *Penaeus monodon* C: *H. gammarus* and *P. monodon*.

this gene. Uncorrected pairwise divergence results suggest it has potential as a valuable gene for reconstructing family to order level relationships. Divergence among species within *Euastacus* was low (0.7% – 1.5%) suggesting this marker may be less informative than other nuclear protein coding markers such as Histone 3 when reconstructing relationships among species. As with EPRS, this marker shows greater divergences (18.8% – 23%) at the deeper level (among infraorders/suborders) than Histone 3. High levels of divergence are often considered indicative of saturation; however, we found increasing divergence with increasing evolutionary distance suggesting saturation may not have been reached even among the deeper nodes, indicating the utility of this gene to infer phylogenetic relationships at these higher levels of divergence.

4 PHYLOGENY BASED SYSTEMATICS

Reconstructing the evolutionary relationships among decapod crustaceans using molecular data has taken two directions: using only protein coding genes, which are phylogenetically informative at deeper nodes, or incorporating as much molecular information available including both ribosomal RNA and protein coding genes in a family level supertree. We have taken the latter approach and reconstructed Decapoda relationships using a total of eight genes and 46 taxa including representatives of seven infraorders of Pleocyemata and a representative of Dendrobranchiata (*Penaeus* sp.) as an outgroup. Pleocyemata representatives include Astacidea, Achelata, Polychelida, Thalassinidea, Brachyura, Anomura and Caridea. Non-decapod crustaceans, *Euphausia eximia* (Euphausiidae: Euphausiacea) and *Lysiosquillina maculata* (Lysiosquillidae: Stomatopoda), were also included in the analysis as outgroups to all the decapods. Rather than focus on representing all lineages equally, we were interested in reconstructing relationships at many levels from among species within genera, among families and among infraorders within decapod crustaceans. Therefore, we focused on sampling the Astacidea to demonstrate the usefulness of these genes for reconstructing phylogenies at these various taxonomic levels.

Table 2. Taxonomy and accession numbers of decapod samples and outgroup included in this study. Accession numbers in bold were obtained from GenBank.

Taxon	Voucher ID	12S rRNA	16S rRNA	18S rRNA	28S rRNA	H3	EF-2	EPRS	TM9SF4
Decapoda Latreille, 1802									
Dendrobranchiata Bate, 1888									
Penaeoidea Rafinesque, 1815									
Penaeus sp. Fabricius, 1798	KCpen	EU920908	EU920934	EU920969	EU921005-EU921006	EU921075	—	—	EU921109
Pleocyemata Burkenroad, 1963									
Anomura MacLeay, 1838									
Galatheoidea Samouelle, 1819									
Aegla alacalufi (Jara & López, 1981)	KAC798	**AY050012**	**AY050058**	EU920958	**AY595958**	EU921042	EU921009	EU910098	EU921077
Eumunida funambulus (Miyake, 1982)	KC3100	EU920892	EU920922	EU920957	EU920984	EU921056	EU921032	EU910124	EU921089
Kiwa hirsute (Jones & Segonzac, 2005)	KC3116	—	—	EU920942	EU920987	EU921065	EU921035	EU910128	EU921097
Munidopsis rostrata (Milne-Edwards, 1880)	KC3102	EU920898	EU920928	EU920961	EU920985	EU921066	EU921034	EU910126	EU921100
Lomisoidea Bouvier, 1895									
Lomis hirta (Lamarck, 1810)	KAClohi	**AY595547**	**AY595928**	**AF436013**	**AY596101**	**DQ079680**	EU921040	EU910131	EU921098
Paguroidea Latreille, 1802									
Pomatocheles jeffreysii (Miers, 1879)	KC3097	EU920903	EU920930	EU920965	EU920983	EU921070	EU921031	EU910123	EU921105
Astacidea Latreille, 1802									
Astacoidea Latreille, 1802									
Astacus astacus (Linnaeus, 1758)	KC702	EU920881	**AF235983**	**AF235959**	**DQ079773**	**DQ079660**	EU921008	—	EU921078
Barbicambarus cornutus (Faxon, 1884)	KC1941	EU920883	EU920913	EU920951	EU920993	EU921045	EU921017	EU910106	EU921080
Orconectes virilis (Hagen, 1870)	KC709	EU920900	**AF235989**	**AF235965**	**DQ079804**	**DQ079693**	EU921041	—	EU921102
Procambarus clarkii (Girard, 1852)	KC1497	EU920901	**AF235990**	EU920952	EU920970	EU921067	EU921011	EU910100	—
Parastacoidea Huxley, 1879									
Astacoides betsileoensis (Petit, 1923)	KC1822	EU920882	EU920912	EU920955	EU920992	EU921044	EU921014	EU910103	EU921079
Cherax cuspidatus (Riek, 1969)	KC1175	**DQ006421**	**DQ006550**	EU920960	EU920996	EU921048	EU921010	EU910099	EU921083
Euastacus eungella (Morgan, 1988)	KC2671	**DQ006464**	**DQ006593**	EU920964	EU92100-EU921002	EU921055	EU921018	EU910109	EU921088
Euastacus robertsi (Monroe, 1977)	KC2781	**DQ006507**	**DQ006633**	EU920962	EU920988	EU921058	EU921019	EU910110	EU921091
Euastacus spinichelatus (Morgan, 1997)	KC2631	**DQ006512**	**DQ006638**	EU920963	EU920989	EU921059	—	EU910108	EU921092
Gramastacus insolitus (Riek, 1972)	KC640	EU920895	EU920926	EU920968	EU920994	EU921062	EU921007	EU910097	EU921094
Ombrastacoides huonensis (Riek, 1967)	KC611	EU920905	**AF135997**	EU920956	EU920995	EU921072	—	EU910096	EU921106
Parastacus defossus (Faxon, 1898)	KC1515	EU920902	**AF175243**	EU920953	EU920991	EU921068	EU921012	EU910101	EU921103
Parastacus varicosus (Faxon, 1898)	KC1529	EU920907	EU920933	EU920954	EU920990	EU921074	EU921013	EU910102	EU921108
Nephropoidea Dana, 1852									
Homarus americanus (Milne-Edwards, 1837)	KAChoam	**DQ298427**	**HAU11238**	**AF235971**	**DQ079788**	**DQ079675**	—	—	EU921095
Nephropsis aculeate (Smith, 1881)	KC2117	EU920899	**DQ079727**	**DQ079761**	**DQ079802**	**DQ079691**	—	EU910107	EU921101

Table 2 continued

Taxon	Voucher ID	12S rRNA	16S rRNA	18S rRNA	28S rRNA	H3	EF-2	EPRS	TM9SF4
Brachyura Latreille, 1802									
Calappoidea Milne-Edwards, 1837									
Cycloes granulose (de Haan, 1837)	KC3082	EU920887	EU920917	EU920943	EU920976	EU921050	EU921025	EU910116	EU921085
Calappa gallus (Herbst, 1803)	KC3083	EU920886	EU920916	EU920947	EU920977	EU921049	EU921026	EU910117	EU921084
Dorippoidea MacLeay, 1838									
Ethusa sp. (Roux, 1830)	KC3088	—	EU920925	EU920966	EU920980	EU921061	EU921029	EU910120	EU921093
Grapsoidea MacLeay, 1838									
Cyclograpsus cinereus (Dana, 1851)	KC3417	EU920884	EU920914	EU920945	EU920997	EU921046	EU921038	EU910130	EU921081
Leucosioidea Samouelle, 1819									
Ebalia tuberculosa (Milne-Edwards, 1873)	KC3085	EU920894	EU920924	EU920944	EU920978	EU921060	EU921027	EU910118	—
Praebebalia longidactyla (Yokoya, 1933)	KC3086	EU920904	EU920931	EU920946	EU920979	EU921071	EU921028	EU910119	—
Majoidea Samouelle, 1819									
Chorilia longipes (Dana, 1852)	KC3089	EU920889	EU920919	EU920948	EU920981	EU921052	EU921039	EU910121	EU921087
Raninoidea de Haan, 1839									
Cosmonotus grayi (White, 1848)	KC3092	EU920888	EU920918	EU920949	EU920982	EU921051	EU921030	EU910122	EU921086
Caridea Dana, 1852									
Palaemonoidea Rafinesque, 1815									
Anchistioides antiguensis (Schmitt, 1924)	KC3051	EU920880	EU920911	EU920936	EU920971	EU921043	EU921020	EU910111	—
Coutierella tonkinensis (Sollaud, 1914)	KC3068	EU920890	EU920920	EU920937	EU920975	EU921053	EU921024	EU910115	—
Crangonoidea Haworth, 1825									
Crangon crangon (Linnaeus, 1758)	KC3052	EU920885	EU920915	EU920938	EU920972	EU921047	EU921021	EU910112	EU921082
Bresilioidea Calman, 1896									
Discias sp. (Rathbun, 1902)	KC3108	EU920891	EU920921	EU920941	EU920986	EU921054	—	EU910127	—
Alpheoidea Rafinesque, 1815									
Hippolyte bifidirostris (Miers, 1876)	KC3059	EU920896	EU920927	EU920939	EU920974	EU921063	EU921023	EU910114	—
Eualus gaimardii (Milne-Edwards, 1837)	KC3056	EU920893	EU920923	EU920940	EU920973	EU921057	EU921022	EU910113	EU921090
Achelata Scholtz & Richter, 1995									
Palinuroidea Latreille, 1802									
Jasus edwardsii (Hutton, 1875)	KC3209	—	**DQ079716**	**AF235972**	**DQ079791**	EU921064	EU921036	EU910129	EU921096
Palinurus elephas (Fabricius, 1787)	KC3210	—	EU920929	EU920959	EU920999-EU921000	EU921069	EU921037	—	EU921104
Polychelida de Haan, 1941									
Polycheles typhlops (Heller, 1862)	KC3101	EU920906	EU920932	EU920950	EU921003-EU921004	EU921073	EU921033	EU910125	EU921107
Thalassinidea Latreille, 1831									
Callianassoidea Dana, 1852									
Lepidophthalmus louisianensis (Schmitt, 1935)	KAC1852	EU920897	**DQ079717**	**DQ079751**	**DQ079792**	**DQ079678**	EU921015	EU910104	EU921099
Sergio mericeae (Manning & Felder, 1995)	KAC1865	EU920909	DQ079733	DQ079768	DQ079811	DQ079700	EU921016	EU910105	EU921110

Table 2 continued

Taxon	**Voucher ID**	**12S rRNA**	**16S rRNA**	**18S rRNA**	**28S rRNA**	**H3**	**EF-2**	**EPRS**	**TM9SF4**
Outgroup									
Stomatopoda Latreille, 1817									
Lysiosquilloidea Giesbrecht, 1910									
Lysiosquillina maculata (Fabricius, 1793)	KC3832	EU920910	EU920935	EU920967	EU920998	EU921076	—	—	EU921111

The genes included in our analyses were 12S, 16S, 18S, 28S, H3, EF-2, EPRS and TM9sf4. A second analysis was run on the four nuclear protein-coding genes. Use of nuclear rRNA 18S and 28S data has been criticized for ambiguities noted in alignments (Tsang et al. 2008). The difficulties in aligning highly variable data may be overcome by using sophisticated methods of alignment employed in recently developed programs such as DIALIGN-T (Subramanian et al. 2005) and MAFFT (Katoh et al. 2002; Katoh et al. 2005). These programs produce more accurate alignments than ClustalW with increasing evolutionary distance (e.g., MAFFT, Nuin et al. 2006) or when gaps are present (indels) in the resulting alignment of sequence data (e.g., DIALIGN-T and MAFFT, Golubchik et al. 2007). To further improve the alignment, GBlocks can be used to identify and exclude ambiguous regions of sequence data (Castresana 2000; Talavera & Castresana 2007). We used MAFFT to align all gene fragments and subsequently ran each dataset through GBlocks (retaining half gap positions) to recover the most useful sequence data. As an example, this reduced the 28S MAFFT alignment from 4489 to 1254 base pairs. Our resulting alignment for the eight-gene dataset was 5104 nucleotides.

Table 3. Empirical base frequencies for each gene region and associated model parameters estimated from the sequence data in RAxML.

	A	**C**	**G**	**T**	**alpha**	**pinvar**
12S rRNA	0.3670	0.0981	0.1726	0.3622	0.6030	0.1934
16S rRNA	0.3399	0.1116	0.2027	0.3458	0.6235	0.2879
18S rRNA	0.2502	0.2342	0.2780	0.2377	0.9231	0.4940
28S rRNA	0.2501	0.2357	0.3161	0.1981	0.7772	0.2735
H3	0.2152	0.3172	0.2654	0.2022	1.0618	0.5882
EF-2	0.2364	0.2469	0.2655	0.2512	1.4067	0.4872
EPRS	0.2857	0.2159	0.2523	0.2460	1.6197	0.3690
TM9SF4	0.1587	0.2784	0.2455	0.3174	0.9592	0.4982

Maximum likelihood phylogenies were constructed with RAxML (Stamatakis 2006; Stamatakis et al. 2008) at the CIPRES portal assuming a GTR+G+I model and estimation and optimization of α-shape parameters, GTR-rates, and empirical base frequencies for each gene. We allowed the program to choose the number of bootstrap replicates, and for the eight-gene dataset, 150 bootstrap replicates were run before termination. For the smaller nuclear protein coding alignment, 250 bootstrap replicates were run before the program terminated. The estimated parameters are presented in Table 3.

The relationships within Astacidea were well resolved, with bootstrap support in 11 of 14 nodes supported by 95% or greater and all nodes supported greater than 80%. As a comparison, the ML phylogeny based on the four-gene dataset (nuclear protein coding) constructed a similar topology within Astacidea although the nodes were not as strongly supported. Only six nodes were supported greater than 95% with an additional five nodes supported greater than 70%. This result suggests that although the nuclear coding genes have power to resolve relationships within an infraorder, additional data from ribosomal genes adds to the information available for reconstructing relationships across the whole of decapod diversity. Our group continues to add genes and taxa to achieve our goal of reconstructing a robust phylogenetic estimate for the decapod crustaceans.

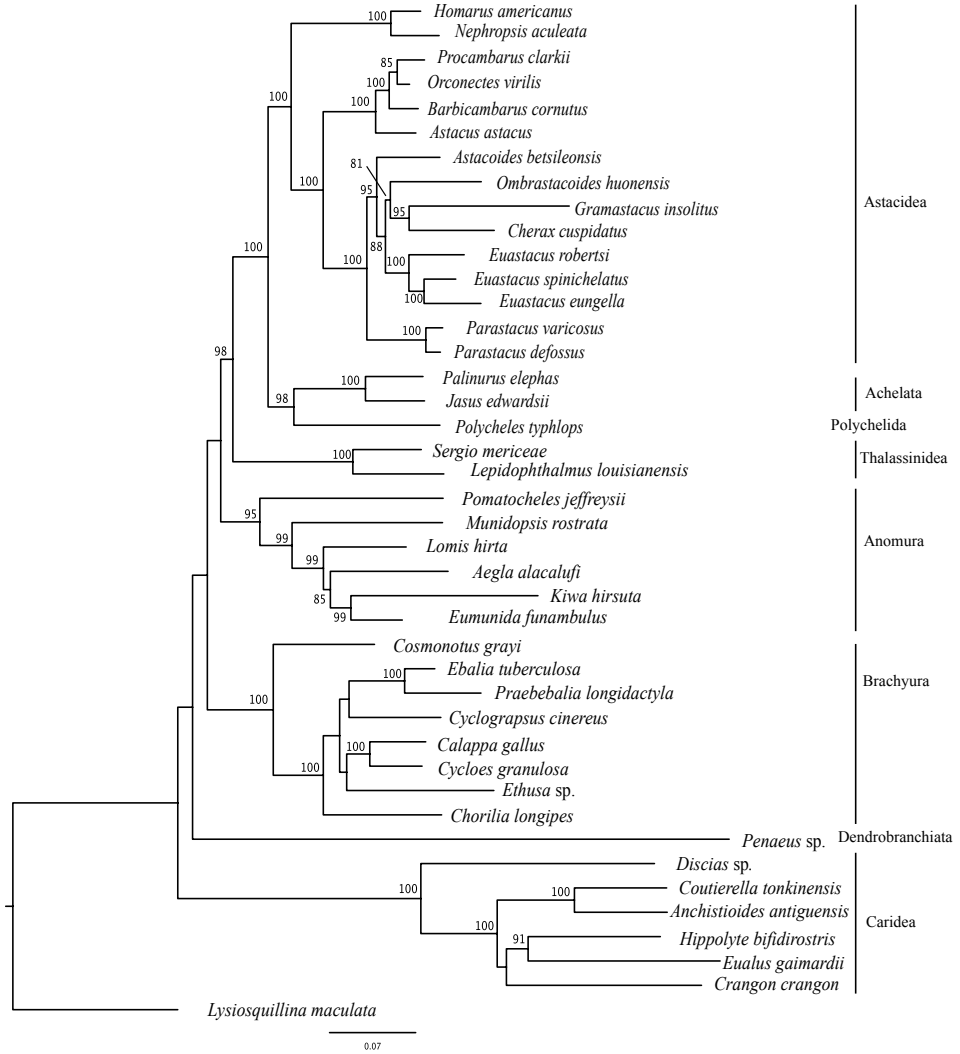

Figure 6. Maximum likelihood phylogeny based on two mitochondrial and six nuclear genes constructed in RAxML. Values at nodes represent bootstrap support greater than 70%.

ACKNOWLEDGEMENTS

We thank the wide variety of friends and colleagues who have helped us collect decapod crustaceans from around the world over the past 15 years. Likewise, this study was made possible by the exceptional undergraduates from Brigham Young University who have labored to collect DNA sequence data from decapod crustaceans. Our work was supported by Brigham Young University and a grant from the US NSF EF-0531762 awarded to KAC.

REFERENCES

Ahyong, S.T., J.C.Y. Lai, D. Sharkey, D.J. Colgan & Ng, P.K.L. 2007. Phylogenetics of the brachyuran crabs (Crustacea: Decapoda): the status of Podotremata based on small subunit nuclear ribosomal RNA. *Mol. Phy. Evol*. 45: 576-86.

Ahyong, S.T. & O'Meally, D. 2004. Phylogeny of the Decapoda Reptantia: resolution using three molecular loci and morphology. *Raffles Bull. Zool*. 52: 673-93.

Audzijonyte, A., Daneliya, M.E. & Vainola, R. 2008. Phylogeny of Paramysis (Crustacea: Mysida) and the origin of ponto-caspian endemic diversity: resolving power from nuclear protein-coding genes. *Mol. Phy. Evol*. 46: 738-59.

Boas, F.E.V. 1880. Studier over decapodernes slaegtskabsforhold. *Dan. Selsk. Skr*. 6: 26-210.

Burkenroad, M.D. 1963. The evolution of the Eucarida (Crustacea, Eumalacostraca) in relation to the fossil record. *Tulane Stud. Geol*. 2: 1-17.

Burkenroad, M.O. 1981. The higher taxonomy and evolution of Decapoda (Crustacea). *Trans. San Diego Soc. Nat. Hist*.19: 251-68.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol*. 17: 540-52.

Cerini, C., Kerjan, P., Astier, M., Gratecos, D., Mirande, M. & Semeriva, M. 1991. A component of the multisynthetase complex is a multifunctional aminoacyl-transfer RNA-synthetase. *Embo J.* 10: 4267-77.

Cerini, C., Semeriva, M. & Gratecos, D. 1997. Evolution of the aminoacyl-tRNA synthetase family and the organization of the Drosophila glutamyl-prolyl-tRNA synthetase gene - intron/exon structure of the gene, control of expression of the two mRNAs, selective advantage of the multienzyme complex. *Eur. J. Biochem*. 244: 176-85.

Colgan, D.J., McLauchlan, A., Wilson, G.D.F., Livingston, S.P., Edgecombe, G.D., Macaranas, J., Cassis G. & Gray, M.R. 1998. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Aust. J. Zool*. 46: 419-37.

Crandall, K.A. & Fitzpatrick, J.F. 1996 Crayfish molecular systematics: using a combination of procedures to estimate phylogeny. *Syst. Biol*. 45: 1-26.

Folmer, O., Black, M. Hoeh, W., Lutz, R. & Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotech*. 3: 294-99.

Golubchik, T., Wise, M.J., Easteal, S. & Jermiin, L.S. 2007. Mind the gaps: evidence of bias in estimates of multiple sequence alignments. *Mol. Biol. Evol*. 24: 2433-42.

Hengen, P.N. 1995. Methods and reagents - fidelity of DNA polymerases for PCR. *Trends Biochem. Sci*. 20: 324-25.

Hillis, D.M. & Dixon, M.T. 1991. Ribosomal DNA - molecular evolution and phylogenetic inference. *Q. Rev. Biol*. 66: 411-53.

Katoh, K., Kuma, K., Toh, H. & Miyata, T. 2005. Mafft version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 33: 511-18.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. 2002. Mafft: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Research* 30: 3059-66.

Lasko, P. 2000. The *Drosophila melanogaster* genome: translation factors and RNA binding proteins. *J. Cell. Biol*. 150: 51-56.

Mitsuhashi, M., Sin, Y.W., Lei, H.C., Chan, T.Y. & Chu, K.H. 2007. Systematic status of the caridean families Gnathophyllidae Dana and Hymenoceridae Ortmann (Crustacea : Decapoda): a preliminary examination based on nuclear rDNA sequences. *Invertebr. Syst*. 21: 613-22.

Mokady, O., Rozenblatt, S., Graur, D. & Loya, Y. 1994. Coral-host specificity of red sea lithophaga bivalves: interspecific and intraspecific variation in 12S mitochondrial ribosomal RNA. *Mol. Mar. Biol. Biotech*. 3: 158-64.

Moritz, C, Dowling, T.E. & Brown, W.M. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst*. 18: 269-92.

Moriyama, E.N., & Powell, J.R. 1997. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *J. Mol. Evol*. 45: 378-91.

Nuin, P.A.S., Wang, Z. & Tillier, E.R.M. 2006. The accuracy of several multiple sequence alignment programs for proteins. *BMC Bioinformatics* 7: 471.

Porter, M.L., Cronin, T.W., McClellan, D.A. & Crandall, K.A. 2007. Molecular characterization of crustacean visual pigments and the evolution of pancrustacean opsins. *Mol. Biol. Evol*. 24: 253-68.

Porter, M.L., Perez-Losada, M. & Crandall, K.A. 2005. Model-based multi-locus estimation of decapod phylogeny and divergence times. *Mol. Phy. Evol*. 37: 355-69.

Regier, J.C. & Shultz, J.W. 2001. Elongation factor-2: a useful gene for arthropod phylogenetics. *Mol. Phy. Evol*. 20: 136-48.

Regier, J.C. & Shi, D. 2005. Increased yield of PCR product from degenerate primers with nondegenerate, nonhomoloogous 5' tails. *BioTechniques*, 38: 34-38.

Scholtz, G. & Richter, S. 1995. Phylogenetic systematics of the reptantian Decapoda (Crustacea, Malacostraca). *Zool. J. Linn. Soc.-Lond*. 113: 289-328.

Song, H, Buhay, J., Whiting, M.F. & Crandall, K.A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Nat. Acad. Sci.* In review.

Stamatakis, A. 2006. Raxml-vi-hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-90.

Stamatakis, A., Hoover, P. & Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML Web-Servers. *Syst. Biol*., In press.

Subramanian, A.R., Weyer-Menkhoff, J., Kaufmann, M. & Morgenstern, B. 2005. Dialign-T: an improved algorithm for segment-based multiple sequence alignment. *BMC Bioinformatics* 6.

Talavera, G., & Castresana, J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol*. 56: 564-77.

Tsang, L.M., Ma, K.Y., Ahyong, S.T., Chan, T.-Y. & Chu, K.H. 2008. Phylogeny of Decapoda using two nuclear protein-coding genes: origin and evolution of the Reptantia. *Mol. Phy. Evol*. 48: 359-368.

Whiting, M.F., Carpenter, J.C., Wheeler, Q.D. & Wheeler, W.C. 1997. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol*. 46: 1-68.

Whiting, M.F. 2002. Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zool. Scr*. 93-104.