

Results from Prior NSF Support (Regier and Shultz, last 5 years)

Regier, J.C., Friedlander, T.P., Mitter, C., and Peigler, R.S. "Multiple Nuclear Genes for Lepidopteran Phylogenetics," DEB-9509174. 9/1/95 - 2/28/97. \$90,000. Relevance to current proposal: Enabled primer development of new protein-encoding nuclear genes. Publications and Manuscript Submitted: 1. Attacini: Ref. [27], 21 taxa sampled, 42 genes sequenced. 2. Period gene development: Ref. [71], 26 taxa sequenced, 26 genes sequenced. 3. Non-ditrysian Heteroneura: Ref. [29], 37 taxa sampled, 37 genes sequenced. 4. Lasiocampidae: Ref. [73], 13 taxa sampled, 13 genes sequenced. 5. Sphingidae: Ref. [72], 23 taxa sampled, 41 genes sequenced. 6. Saturniini: (manuscript in preparation), 87 taxa sampled, 151 genes sequenced. Student Support: Dr. Tim Friedlander (postdoc associate), Kurt Horst (honors undergrad)

Mitter, C., Mitchell, A., and Regier, J.C. Doctoral Dissertation Improvement Award: "Evolution of the True Cutworms (Lepidoptera: Noctuidae) inferred from Nuclear Genes," DEB-9623799. 6/1/96 - 5/31/98. \$10,500. Relevance to current proposal: Supported a talented graduate student. Publications, etc.: 1. Taxonomic limits of synonymous nucleotide changes: Ref. [60], 49 taxa sampled, 49 genes sequenced. 2. More taxa versus more genes: Ref. [61], 77 taxa, 154 genes sequenced. 3. Ph.D. dissertation and future publications: Ref. [59], 209 taxa sampled, 418 genes sequenced. Student Support: Mr. Andrew Mitchell (Ph.D. rec'd, 1998)

Shultz, J.W. and Regier, J.C. "Resolving Arthropod Phylogeny using Multiple Nuclear Genes," DEB-9629791. 9/15/96 - 02/28/00. \$125,000. Relevance to current proposal: Enabled primer development for EF-1 α and Pol II and their application to arthropod phylogeny. Publications and Manuscripts Submitted: 1. Pancrustacea I: Ref. [74], 21 taxa sampled, 32 genes sequenced. 2. Centipedes: Ref. [82], 10 taxa sampled, 10 genes sequenced. 3. Cambrian explosion: Ref. [75], 40 genes sequenced, 40 taxa sampled. 4. Myriapoda: Ref. [76], 48 taxa sampled, 96 genes sequenced. 5. Pancrustacea II: Ref. [83], 35 taxa sampled, 66 genes sequenced. 6. Opiliones: Ref. [84], 27 taxa sampled, 54 genes sequenced. 7. Arachnida: (manuscript in preparation), 32 taxa sampled, 64 genes sequenced. Student Support: Ms. Kelly Kissane (M.S. candidate)

Regier, J.C. and Shultz, J.W. "Testing Hypotheses of Early Arthropod Evolution," DEB-9981970. 4/1/2000 - 3/31/2003. \$200,000. Relevance to current proposal: Enabled primer development for EF-2 and their application to arthropod phylogeny. Publications, etc.: 1. EF-2: Ref. [77], 23 taxa sampled, 69 genes sequenced. 2. Enolase: GenBank accessions, 6 taxa sampled, 6 genes sequenced. Student Support: Chris Desjardins (Ph.D. cand.)

Wilson, H.M., Regier, J.C., and Shultz, J.W. Phylogeny of Millipedes (Diplopoda) using Anatomical, Fossil and Molecular Evidence, DEB-0075605. 8/1/2000 - 7/31/2003. \$210,000. Relevance to current proposal: Enabled expanded taxon (41 species) and gene sampling (EF-1 α , Pol II, EF-2) of Myriapoda.

Results of Prior NSF Support (Last 5 Years) for Joel W. Martin

1. Biotic Surveys and Inventories Program

DEB 9972100, Survey of the Shallow Water Marine Cryptofauna of Guana Island, BVI

Awarded to: T. L. Zimmerman and J. W. Martin

Duration: July, 1999 - June, 2002 (current)

Amount: \$183,374, plus a supplement of 98,364 awarded in September, 2000

We have spent two seasons in the field; our third will be this summer (2001). To date, we have prepared an extensive CD-ROM identification disk with over 500 color photographs of Caribbean marine invertebrates and have delivered it to officials in the BVI and to NSF (D. Causey). We have 14 manuscripts in preparation dealing with marine invertebrates of this region, plus the following that have appeared: [58], [57], and [26]. The web site is URL http://nhm.org/~tzimmerm/bvi_800/0bvi-idx.htm.

Dissertation Research: **PEET: Monographic Research in the Leptostraca, Brachyura, and Conchostraca NSF-DEB-9978193, \$9000 9/1/99 — 8/30/2004 PI s J.W. Martin, D.K. Jacobs \$750,000**

This award is for training three doctoral students in crustacean systematics and monography. Two of the students (T. Haney, S. Trautwein) are currently enrolled at UCLA; the fourth student (S. Boyce) will begin in the fall of 2001. We have also partially supported one postdoctoral researcher (R. Wetzer) and a fourth doctoral student (T. Zimmerman). Progress has been solid and steady. Peer-reviewed publications include [35], [34], [35], [56] Martin et al. (in press), [36], [55], and Hickman and Zimmerman (2000). Our web site, the first posted site for any of the PEET projects funded in 1999, can be viewed at URL <http://www.nhm.org/~peet/>.

Results of Prior NSF Support (Last 5 Years) for Jeff Thorne.

Evolution of evolutionary rates: a genomic perspective, NSF Award No.

INT-990934, 5/15/00-5/14/03, PI = J.L. Thorne. \$35,730. Summary: This project supports travel by Thorne and students to Japan (location of collaborator H. Kishino). Progress on Bayesian divergence time estimation methods is detailed below on pages 11-13 below. Publications:[46, 80]

Genomic dissection of a nematode-plant interaction: A tool to study plant biology. NSF Award No.

DBI0077503, 9/01/00 - 8/31/03, PI= D.M. Bird, 4 Co-PIs include J.L. Thorne, \$2,607,829. Summary: This is a genomic study of parasitic nematode species. Thorne's role in this project is to assist with sequence analysis and study the possibility of horizontal gene transfer events. Publications: No publications have yet resulted from this grant because it is only recently been initiated. However, two manuscripts are in preparation.

Results of Prior NSF Support (Last 5 Years) for PI Cunningham.

Dissertation Research: Comparative Phylogeography and Extinction-recolonization

Dynamics of the North Atlantic NSF-DEB-9972707, \$9000 6/15/99 — 5/31/2000.

JP Wares (PhD. 2000) carried out the largest comparative study to date of North Atlantic marine phylogeography to date and has led to four publications. [98, 99, 101] [100]

Dissertation Enhancement: Testing the Independent Origin of Compound Eyes in

Ostracoda (Crustacea): A Step Beyond Character Mapping NSF-DEB-9975062 — \$14451 6/15/99 — 5/31/2000. Grad Student Todd Oakley, PhD 2001. We have used phylogenetics (> 3KB per taxon) and studies of opsin gene family relationships and gene expression to provide the strongest evidence to date supporting the independent origin of an arthropod compound eye [68, 69].

Collaborative Research: Hermits, Kings, and Convergence: Integrating Molecules and Morphology to study the Phylogenetic Relationships of the Anomura (Crustacea, Decapoda) NSF-DEB-96-

15461, \$150,000 (including two REU supplements) 3/1/97-3/1/2000. We identified the first major mitochondrial gene rearrangements in the Pancrustacea, including 2 protein coding genes [37](we identified one tRNA movement in error and have corrected this mistake as an erratum to MBE). We have used these rearrangements together with 2750 bp of sequences for 26 taxa from four mtDNA and nuclear genes to identify 5 independent origins of the crab-like form [62], and an additional 1792 bp of sequences for 22 taxa from three genes to investigate evolution within the crab family Lithodidae to investigate the transition to the crab-like form [107]. This grant supported the following additional publications [18-20, 44, 67, 93]

Collaborative Research: Monographic Studies of the Hydractinioidea

NSF-DEB-9978131 \$553,490 10/01/99-6/30/05, with \$10,000 supplement from International Programs. Three PhD students have begun their monographic research (A. Frese, M. Miglietta, A. Lindner), and 12 American scientists attended a course we taught on Hydrozoan Systematics (summer 2000). PEET helped fund a text on hydrozoan taxonomy by our European collaborators[8], who, with PEET funding, have founded an online database of hydrozoan taxonomic bibliography <http://siba2.unile.it/ctle/hydro/index.php3>. Our (free) Duke hydrozoan sequencing service has sequenced 75 16S mtDNA sequences to aid hydrozoan taxonomists around the world in species identification and systematics to help circumvent the lack of characters to distinguish hydrozoan species.

(Cunningham Prior support, continued) SGER: Documenting the Effect of the 1997 El Niño Southern Oscillation Event on a Broadly-Dispersing Barnacle Species at a Major Biogeographical Boundary. NSF-DEB-96-15461, \$40,000 6/1/98-9/30/1999. We completed a comparative phylogeographic study of asymmetric migration events across Pt. Conception in three invertebrates [102]. We are preparing a manuscript documenting a significant effect on barnacle migration by the 1997 El Niño, and which will include a detailed study of a hybrid zone in the barnacle *Balanus glandula*.

PROJECT DESCRIPTION

Although systematists have long been fascinated by the incredible diversity of arthropods with more species than the rest of the Metazoa combined a clear understanding of the patterns and processes that gave rise to that diversity is largely missing. For example the first molecule-based, higher-level phylogenetic study that included arthropods was completed 13 years ago [25], but many relationships remain controversial, and it is fair to say that, to date, molecules have added little to our understanding of arthropod phylogeny. This lack of phylogenetic resolution has profound implications for some of the biggest questions in arthropod evolution, including the nature and number of lineage-specific transitions to the terrestrial environment; the evolution of developmental mechanisms controlling segmentation and the origin(s) of eyes; and, more generally, the rate of morphological diversification from Cambrian-explosion times onward [14, 33].

Recent molecular systematic studies of arthropods and related groups have yielded some notable hypotheses worthy of further investigation, including multiple lines of evidence that arthropods form a monophyletic group with other moulting animals — the Ecdysozoa [3, 22, 31, 54, 77] to the exclusion of the segmented Annelida. This surprising result has forced developmental biologists to reconsider the evolution of segmentation throughout the Metazoa. Another notable finding is that the terrestrial Atelocerata (i.e., Hexapoda + Myriapoda) are paraphyletic, so that hexapod insects and myriapods (e.g., centipedes and millipedes) represent a case of startling convergence [7, 32, 77], and that hexapods and crustaceans form a monophyletic Pancrustacea.

Despite these success stories, many fundamental questions remain unanswered. For example, the relationships of the three major arthropod clades (i.e., Pancrustacea, Chelicerata, and Myriapoda) and of the classes of Crustacea [32, 77, 87, 89], including the monophyly of some (e.g., Maxillopoda). These questions remain despite a recent study [77], (Figure 1) by two of the co-PIs which analyzed about 4 kb / taxon of three conservative, protein-encoding nuclear genes for a broad, but sparse, sampling of arthropods, including all widely recognized crustacean classes. We interpret their failure to robustly recover these deep relationships as indicating that substantially more data are needed for more taxa. Our proposal herein to sample approximately 60 kb (a 15-fold increase) for each of 85 taxa (a greater than fourfold increase) is designed to test this hypothesis. These questions are of broad evolutionary interest because crustaceans have the greatest diversity in bauplans of all arthropods.

The uncertainty over origins, coupled with the availability of new and powerful tools, i.e., phylogenomics, presents new possibilities. For example, the emerging view that hexapod insects are nested within a paraphyletic Crustacea [32, 77, 105] has led to the exciting possibility of identifying the closest extant relatives to the crustacean lineage that invaded land and founded the hexapod insects. Secondly, phylogenomics, and its promise of robustly resolving the arthropod tree, also offers the possibility of maximizing the utility of the fossil record for estimating rates and dates, particularly when extended to fossil-poor groups, e.g., Cephalocarida and Remipedia.

In the course of our proposed research, we will answer outstanding questions about the evolution of arthropods. This will be accomplished through the development of a large set of new genetic markers and of novel analytical methods for the estimation of divergence dates. We will also develop and

distribute user-friendly software to implement Bayesian methods of estimating divergence dates. Our research will have other benefits for the systematics community and beyond. These include the distribution of our host of primer sets to other arthropod systematists, the production of a user-friendly software package for estimating divergence times from sequence data, and our planned educational outreach to high school students.

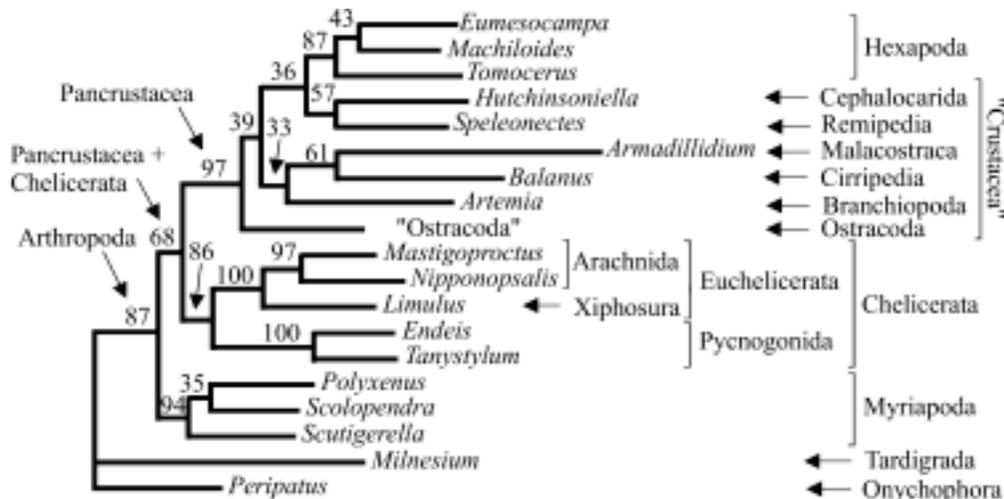


Figure 1: Arthropod relationships based on ML analysis of EF-1alpha + Pol II + EF-2 nucleotide sequences from 19 species. GTR model; site-specific rates; 3 character partitions/gene; nt3 excluded. BP values above branches.

Genomic objectives

¥ Search for phylogenetically useful gene regions (ca. 500 nt in length) among the 1523 genes reported to have strict orthologs between *H. sapiens*, *D. melanogaster*, and *C. elegans* [49].

- Preliminary screen of targeted sequences with computer to confirm orthology; assay overall sequence divergence; identify conserved primer sites.
- Screen promising PCR primers in the lab to confirm amplification for a set of test taxa; sequence PCR products from test taxa and do phylogenetic analysis to see whether expected relationships ("concordance groups") are recovered.
- Make these primers freely available to other workers in arthropod and invertebrate systematics.

Conceptual objectives:

- ¥ Reconstruct the higher-level phylogeny of the Arthropoda. Specific goals within the Arthropoda include - Confirm the monophyly of Pancrustacea (Crustacea + Hexapoda);
- Establish the relationships of the Pancrustacea, Myriapoda and Chelicerata;
 - Establish paraphyly/monophyly of the Crustacea;
 - Identify the crustacean sister taxon to the Hexapoda;

Quantitative objectives

¥ Use Bayesian methods to develop a timeline for arthropod evolution based on fossil occurrences and sequence data. This will include estimates of divergence times between lineages known to have existed at the time of the Cambrian explosion.

¥ Develop user-friendly implementations of these methods for general use.

Educational objectives

¥ Develop a program of outreach to high school students that imparts the excitement of studying arthropod evolution.

OVERVIEW AND INTRODUCTION TO OUR TEAM

We propose to build upon the work of co-PIs Regier and Shultz (plus C. Mitter and other colleagues) who have pioneered the use of nuclear protein-coding genes for higher-level arthropod phylogenetics. They have extensive experience selecting genes based on their phylogenetic utility and in amplifying and sequencing them from total RNA extractions. Our project will differ in three fundamental ways from their previous and ongoing efforts.

First, we will focus much more heavily on Crustacea, and to this end are joined by co-PI Jody Martin (co-author of "An Updated Classification of the Recent Crustacea" [55]), who has chosen the additional crustacean groups and who will arrange for their collection and identification.

Second, we will *not* emphasize the collection of long sequences from a small number of genes, as is the case for the three genes currently being investigated by Regier and Shultz. Instead we will focus on generating much more data in "directly sequenceable" bits, i.e., 120 segments each approximately 500 bp in length (acceptable range 300-800 bp). This approach has theoretical justification [15] and allows the generation of data at a faster pace.

Third, we will assess our multiple genes in a Bayesian framework together with evolutionary statisticians (co-PI Thorne and foreign collaborator H. Kishino, see supporting letter and CV) in order to estimate times of divergence for major clades identified in our phylogenetic studies. Their methods are designed to take full advantage of multiple genes. They allow departure from rate constancy for each gene along each branch. Funding of this proposal will enable the implementation of a user-friendly software program for general distribution.

The overall project will be coordinated by PI Cliff Cunningham who has worked on decapod molecular systematics, and whose laboratory has experience amplifying and sequencing invertebrate nuclear protein-coding genes [e.g.68]. Like Regier and Shultz, he has published on a variety of issues in molecular systematics, including the efficacy of phylogenetic methods and how best to treat multiple data partitions such as those we propose to collect here. Cunningham also founded the Duke phylogenetic consulting center, to bring phylogenetic expertise to molecular biologists seeking, for example, to distinguish between orthology and paralogy in their study of gene families.

We plan to operate as a coordinated team. This includes an agreement to co-publish results and to share all of our nucleic acid extracts. Sharing extracts will also be necessary because Duke and Maryland will be sequencing different genes for the same set of taxa.

TIMELINE

Years 1 and 2, *major initiatives*

- LA, MD ¥ Collect the crustaceans and hexapods needed to complete our expanded sampling (estimated time = 1 year).
- MD, Duke ¥ Screen for candidate genes, develop primers, and evaluate genes for phylogenetic utility using test taxa already available (estimated time = 1 year).
- NC St. ¥ Develop and distribute user-friendly software for Bayesian estimates of divergence times. Work with other co-PIs to apply methods to existing data relevant to project goals (estimated time = 2 years).
- LA ¥ Develop and implement K-12 teaching module for arthropod evolution (estimated development time = 1 year; implementation thereafter).

Years 2-4.5, *major initiatives*

- MD, Duke ¥ Amplify and sequence the 120 gene fragments identified by the screening process for the remaining taxa.

Year 4.5-5, *major initiatives*

- ¥ Analyze all data and publish results.

Duke (PI Cunningham); LA= Los Angeles County Museum (co-PI Martin); MD = University of Maryland (Shultz) and University of Maryland Biotechnology Institute (Regier); NC St. = North Carolina State University (co-PI Thorne).

BRIEF REVIEW OF HIGHER-LEVEL ARTHROPOD MOLECULAR SYSTEMATICS

Higher-level arthropod molecular systematics has been reviewed on multiple occasions, including several recent articles by co-PIs Regier and Shultz [74, 75, 77, 83]. Basically, three sources of molecular characters have been used: nuclear rDNA, mitochondrial DNA sequences and gene order, and nuclear single-copy protein-coding genes. The main conclusions derived from analysis of these data types are assessed below with regard to several major questions in arthropod phylogeny. Of course, inferences about arthropod phylogeny from morphology have a much longer history than those from molecules, and, indeed, the hypotheses we are aiming to test in this proposal are largely derived from morphology. However, a review of arthropod evolutionary morphology is beyond the scope of this proposal. Nevertheless, we are fully cognizant of morphological hypotheses; co-PIs Shultz and Martin are morphological experts in their groups, terrestrial arthropods and crustaceans, respectively. Both Shultz and Martin have played the central roles in developing our taxon sampling scheme, and both are keenly interested in relating the paleontological record to molecular dating techniques.

Three Higher-Level Hypotheses are generally supported by Molecular Data.

Outgroups to the Arthropoda

The placement of the arthropods with other moulting animals (*i.e.*, Ecdysozoa [3], and especially onychophorans and tardigrades) has been a success story for molecular phylogenetics. This result was first reported from a study of 18S rDNA [3] and was confirmed with expanded taxon sampling, although node support was low [12]. In addition, sequence analysis of Hox genes, in particular Abd-B [9], has been interpreted to support Ecdysozoa, although the short report is difficult to evaluate. Finally, two other "higher order" characters are consistent with Ecdysozoa, the absence of the δ -thymosin gene [54] and the presence of an insertion in the coding region of EF-2 [31, 77]. Within the Ecdysozoa, molecular studies have generally placed the Onychophora and the Tardigrada as most closely related to the Arthropoda, so representatives of both will be used as outgroups in our proposed studies.

Appearance of Pancrustacea, demise of Atelocerata, but lingering doubts about crustacean and hexapod monophyly

Sequence analysis of 18S + 28S [32] and of three nuclear protein-coding genes (see Figure 1 [77]), both separately and in combination, strongly support the monophyly of Pancrustacea (= Hexapoda + Crustacea), although neither data set is informative of crustacean monophyly. Based on extensive taxon sampling, Pancrustacea is further supported by a single rearrangement in mitochondrial gene order relative to the arrangement in numerous non-pancrustaceans [7]. Mitochondrial gene rearrangements are, for the most part, rare in arthropods [7], but see [6, 12, 37]. **The strong support that molecular data provide for Pancrustacea is perhaps the clearest evidence to date of the utility and potential in expanded studies such as ours. This is particularly noteworthy because morphologists had consistently supported the conflicting group Atelocerata (= Hexapoda + Myriapoda), a hypothesis that is now undergoing serious reassessment [23, 30].** One of the evolutionary implications of the Pancrustacea hypothesis is that the immediate ancestors to hexapods and myriapods must have separately invaded land. A complicating factor in our current understanding of Pancrustacea is that while the major hexapod clade consisting of true insects is clearly monophyletic relative to other pancrustacean groups, the positions of some basal hexapod groups (*i.e.* Protura, Collembola) have not been fully resolved using molecular data. (See "Monophyly of the Hexapoda" below).

Monophyly of Euchelicerata and Myriapoda.

Both ribosomal [32] and nuclear protein-coding genes (see Figure 1) [77, 83] support the morphologically-supported monophyly of Euchelicerata (= Arachnida + Xiphosura), and, less strongly, its sister group relationship to Pycnogonida to form Chelicerata, a group whose monophyly is still debated among morphologists. Our sampling of Pycnogonida should provide a rigorous test of the Chelicerata hypothesis. The monophyly of Myriapoda is also strongly supported by analysis of nuclear protein-coding genes (see Figure 1) [77, 83], despite lingering doubts among morphologists [47].

Three Outstanding Issues to be Addressed in this Study

Relationships among Pancrustacea, Myriapoda, and Chelicerata

As summarized above, there is widespread agreement within the molecular systematic community on the monophyly of Pancrustacea, Myriapoda, and Euchelicerata, with only the position of Pycnogonida not fully resolved. However, support for their inter-relationships is weak, both for protein-coding (Figure 1) [25, 26, 77, 83]) and ribosomal sequences (reviewed by Giribet and Ribera [32]). The protein-coding genes recover Pancrustacea + Chelicerata using a variety of analytical techniques, occasionally with modest bootstrap support (Figure 1). However, given that this grouping is counter to the morphology-based and widely-cited Mandibulata hypothesis (= Hexapoda + Myriapoda + Crustacea), further work is clearly needed. With modest bootstrap support already observed, it is reasonable to expect robust resolution of this problem when presented with a greatly expanded data set, as herein proposed.

Monophyly of the Hexapoda

Although morphologists do not seriously question hexapod monophyly, the recent finding that crustaceans, not myriapods, contain the sister group of hexapods raises questions about the strength of this long-held conclusion. Perhaps the position of the basal hexapod clades (i.e., Ellipura = Collembola + Protura) is of greatest interest in this regard. Molecular studies to date have not been conclusive. Ribosomal sequences sometimes recover a monophyletic Hexapoda and sometimes don't (reviewed by Giribet and Ribera [32]). Various crustacean groups are occasionally found to split the hexapods. Similarly ambiguous results are found with histone and ubiquitin sequences [13, 103]. The combined analysis of 3 protein-coding nuclear genes (4 kb/taxon) studied by Regier and Shultz (Figure 1) [77, 83] generally supports the monophyly of Hexapoda, albeit with limited taxon sampling. In their studies [77, 83], recovery of Hexapoda was sensitive to method of analysis. Clearly, more data and more taxa are needed.

Crustacean Relationships and the Sister Group to the Hexapoda

With some exceptions (e.g., the inclusion of Phyllocarida within Malacostraca; [87]), molecular studies of higher-level crustacean relationships have not provided robust resolution despite impressive efforts by Spears, Abele, and others [1, 32, 86-90]. According to the most recent revision of crustacean classification co-authored by co-PI Martin [55], 5 or 6 classes of crustaceans are recognized, although virtually nothing is agreed upon concerning their inter-relationships, nor their relationship to hexapods. Interestingly, Brusca has suggested that hexapods arose from within the Pericarida [10, 55], and his hypothesis will be tested through our sampling scheme.

Malacostraca (now with Phyllocarida) is widely considered to be monophyletic, and this is strongly supported by molecules [77, 83, 89]. Remipedia and Cephalocarida each contain only a handful of species and, given their distinctive morphologies, are likely to be monophyletic. Monophyly of Branchiopoda has not been seriously questioned by morphologists [104] and has been confirmed by a recent 18S rDNA study [88]. Maxillipoda are widely recognized as likely to be polyphyletic, and it is common to refer to the constituent subclasses instead (e.g., Ostracoda, Copepoda). Our taxon sampling includes representatives of nearly all of the major clades considered by various workers (at one time or another) to be "maxillopodan," while avoiding some of the diverse but clearly derived and purely parasitic taxa (such as rhizocephalans, ascothoracids, and tantulocarids) and groups that we are unlikely to find in sufficient numbers (e.g., the facetotectans).

Especially vexing among maxillopodan groups (above) is the Ostracoda, thought by some workers to warrant a separate class and by others to be members of the Maxillopoda; they are perhaps the most speciose of all living crustaceans [55]. We will target representatives of the two largest extant subclasses the Myodocopa and Podocopa [55]. Remarkably, we also have in our frozen collections some (few) specimens of the extremely rare "living fossil" genus *Manawa* (considered by some workers to be a living representative of the extinct Paleocopa, but now treated among the Podocopa: Platycopida).[55]

Why not whole mitochondrial genomes?

We have chosen to focus on multiple nuclear loci instead of whole mitochondrial genomes for three reasons: first, phylogenetic information from mitochondrial gene order is now complete for higher-level arthropod relationships, and the result is a single character in support of Pancrustacea (see discussion above); second, there is growing evidence that sequence data from whole mitochondrial genomes can strongly support incorrect nodes with regard to several major evolutionary questions. For example Naylor and Brown [65] showed that a whole mitochondrial genome phylogeny found 100% bootstrap support placing the cephalochordate lancelet as an outgroup to a group composed of sea urchins and chordates. In contrast, in a series of studies of mammal relationships, nuclear protein coding genes appear to converge on a single answer when sample sizes are large [43, 53, 64], that contradict several results from whole mitochondrial genomes [42, 63]. Co-PI s Regier and Shultz have found independent concordance with regard to higher level arthropod relationships between the three genes they have studied [77, 83], and it is also likely that most mitochondrial genes simply evolve too rapidly to recover late Mesozoic-to-Paleozoic divergences [91]. Third, and most importantly, there is simply much more sequence data available in the nucleus.

Selecting Candidate Nuclear Protein-Coding Genes.

The complete genomic sequence of *Drosophila melanogaster* and of a second ecdysozoan (*C. elegans*) is a major boon for higher-level arthropod phylogenetics. Although Regier, Mitter and his colleagues began searching for protein coding genes for arthropod phylogenetics nearly a decade ago [28], the availability of whole genomes not only reveals many more candidate genes, but also makes assessing paralogies and copy number much simpler. The extensive genomic comparisons that accompanied the publication of *D. melanogaster* and *H. sapiens* genomes have led to the identification of over 1523 strict orthologs [97] (defined as reciprocal matches when BLASTed against the other genome). Although orthology needs to be tested by phylogenetic analysis, as described by Hillis [38] and others, these 1523 genes represent a good starting point for identifying candidate genes.

Our approach is based on methods perfected over years of experience by Regier and his colleagues, and is similar to the approach used successfully by a recent study of eutherian mammal phylogeny [64]. There will be five steps to our screening procedure (specific aspects of molecular methodology will be described below in Additional Methods) .

- 1- We will use phylogenetic analysis to ensure the orthology of the putative strict orthologs shared between *H. sapiens*, *D. melanogaster* and *C. elegans* [97]. To ensure detection of paralogy, this analysis will include all genes with significant BLAST results in all three genomes. To maximize the amplification of only orthologous genes, we will reject genes with either multiple copies in *D. melanogaster*, or whose closest paralogs in the *D. melanogaster* genome are closer than 60% at the amino acid level. This will in minimize the chances of amplifying paralogs (based on the experience of Regier and Shultz). Although we cannot ensure *a priori* that recent gene duplications haven t taken place elsewhere in the arthropods, this will reduce the possibility of amplifying misleading paralogs [38], i.e. those that have undergone non-synonymous changes.
- 2- Genes with confirmed orthology across the three genomes will then be examined for regions of conservation. Experience has shown that assessment of divergence between *D. melanogaster* and *H. sapiens* is more useful as an initial criterion than comparisons to *C. elegans*, which will occur subsequently. This is because of the rapid rate of molecular evolution that characterizes the *C. elegans* genome, despite its putatively closer relationship to *Drosophila*. We will search for genes that show >60% amino acid sequence identity over >170 residues between *D. melanogaster* and *H. sapiens*, but we will be particularly interested in those with ≥70% identity. This range of pairwise differences has been shown to be appropriate for resolving deep nodes in arthropods (Mesozoic-Paleozoic) using EF-1 α , EF-2, and RNA polymerase II [25, 26].

- 3- The alignments of conserved amino acids will be augmented by BLAST searches with any orthologous non-dipteran arthropod sequences in GenBank. For each candidate region, fully degenerate flanking primer pairs will be designed only for those conserved amino-acid sequences whose fully degenerate codons display < 256-fold degeneracy. For more specific amplification, we will also identify a third internal primer to be used for nested re-amplification of the original PCR product. (The Regier lab has found nested reamplification an important procedure when using degenerate primers to amplify across Panarthropoda.) This internal primer must be > 300 nt, but less than 800 nt, from the furthest external primer, assuming no introns. Nested primer sites can oftentimes be conveniently positioned to overlap one of the original primer pair, thereby reducing the effort at locating suitable sites and maximizing the amplicon length.
- 4- To ensure consistent amplification of selected gene regions, we will test each primer set using RT-PCR procedures and total RNA isolated from 12 test taxa that were selected to represent the diversity of Panarthropoda. These test taxa include 6 Crustacea (Malacostraca: *Armadillidium vulgare*; Branchiopoda: *Artemia salina*; Remipedia: *Speleonectes tulumensis*; Cephalocarida: *Hutchinsoniella macracantha*; Ostracoda: *Skogsbergia lernerii*; Cirripedia: *Semibalanus balanoides*), 2 Chelicerata (Xiphosura: *Limulus polyphemus*; Arachnida: *Nipponopsalis abei*) 2 Myriapoda (Diplopoda: *Polyxenus fasciculatus*; Chilopoda: *Scolopendra polymorpha*), 1 Hexapoda (*Archaeognatha: Machiloides banksi*); and 1 Tardigrada (*Milnesium tardigradum*) as outgroup.
- 5- For each primer set that amplifies all 13 test taxa, amplification products from these sequences will be isolated, sequenced, and aligned relative to each other. Three criteria will be used to indicate phylogenetic utility:
- Near-absence of non-synonymous polymorphism within individuals (NOTE: Synonymous intra-individual polymorphism is not uncommon throughout the Arthropoda, for example, perhaps 25% of taxa sequenced by Regier and Shultz for EF-1 α , EF-2, and RNA polymerase II display some polymorphism, typically only a few per cent of the total characters. This finding is consistent with the hypothesis of frequent duplication and rapid extinction of gene copies: [50]. Extensive non-synonymous variation within species may indicate deep paralogy which can mislead our analysis).
 - The ability to recover in phylogenetic analyses three monophyletic concordance groups strongly supported by a consensus of morphological and/or molecular studies (see review below):
 - Panarthropoda (Crustacea + Hexapoda; Note that this does *not* assume crustacean monophyly, which will be tested by our analysis.);
 - Euchelicerata (Arachnida + Xiphosura); and
 - Myriapoda (Diplopoda + Chilopoda).
 - The mean and ranges of pairwise differences for amino acids across all test taxa will be evaluated to ensure they are similar to that observed for the three genes already known to be useful in resolving deep arthropod relationships [77](mean = 0.17-0.23, max = 0.25-0.35).

Feasibility of screening procedures: We expect to easily identify enough candidate loci for screening using these procedures. For example, a study of Kingdom-level phylogeny has 15 candidate genes that satisfy a criterion of >70% sequence identity between *D. melanogaster* and *H. sapiens* (see below) [66]. All of these will be tested for orthology by phylogenetic analysis as described above.

Locus	Acc. # Fly	Acc. # Human	Length (aa s)	Identity (%)
78 kdalton glucose-regulated protein	AAF48095	M19645	630	80
Actin	AAF57294	X13839	377	94
Calmodulin-dependent protein kinase	AAD20442	AAA51459	478	78
Casein kinase II alpha	AAF45439	M55265	327	82
Cell division control 2	AAF52932	X05360	288	73
Enolase	AAF51344	X56832	433	71

Extracellular signal-regulated kinase 1	AAA28677	M84490	343	81
F1-atpase alpha	P35381	X59066	528	83
F1-atpase beta	AAF59391	X03559	483	85
Heat shock protein (HSP 70)	AAF55150	M11717	611	81
Phosphatase 2A	P23696	J03805	308	94
Phosphatase I alpha	AAF56306	J04759	308	89
Proliferating cell nuclear antigen	AAF57493	M15796	259	71
Vacuolar ATPase subunit A	AAB02270	L09234	613	91
Vacuolar ATPase subunit B	AAF54836	X62949	481	90

ADDITIONAL METHODS

Acquisition of Specimens. All of our samples have either been collected and extracted, or will be collected by co-PI s Martin and Shultz in the first year of the study (see taxon list on page 15). Because our sequencing depends on RT-PCR, our material is stored in 100% ethanol at -85°C. Specimens are collected alive using methods appropriate to the taxon and placed into a large volume of 100% ethanol. Although long-term storage at -85°C is preferred, the Regier lab has obtained excellent results from ethanol-preserved specimens that had been stored at room temperature for over a year. The Cunningham lab routinely stores specimens in RNA-later (AMBION).

Generation of DNA Sequence Data Sets. Total RNA is isolated and amplified by RT-PCR using gene-specific primers. PCR fragments are gel-isolated, reamplified using an internal primer (or two), and re-gel-isolated. PCR fragments are sequenced from M13 sites added during primer construction to the 5' end of all PCR primers. Automated sequencing occurs on ABI model 3700 and 3100 sequencers. The sequencer files, including chromatograms are analyzed in the Staden software package [92]. Consensus sequences for each gene are aligned using the interactive, on-screen, color editor of the Genetic Data Environment [85]. Nucleotide data sets are prepared using Genetic Data Environment. Protein data sets are prepared using MacClade 3.0 [51].

Phylogenetic Inference Procedures. In our phylogenetic analyses using PAUP* [94], we test optimization criteria, evolutionary models, and data sets. Recovery of concordance groups (listed above in our screening procedure for candidate loci) and bootstrap support for those concordance groups are tests of suitability. As with the publications from the Maryland group, we will show a representative topology that has optimal concordance group recovery, but we also summarize the results of all analyses. ML and MP are the optimization criteria. For nucleotides under MP, we use unordered transformations and 6-parameter parsimony [17, 93] with and without nt3.

Under ML, we will use two classes of best-fit models. The first is site specific rate models [95] that assign rates to designated character classes (e.g., all nt, nt1 + nt2, nt1 with leucine- and arginine-encoding positions removed + nt2). Although Modeltest [70] does not implement siterate models, the Akaike [4] criterion can be used to choose among site specific rate models [11]. The second class of models are those that accommodate among-site rate variation to fit a gamma distribution [106] with invariant sites [81] estimated (with and without nt3), and will be chosen using Modeltest [70].

In collaboration with T. Buckley, PI Cunningham has just completed an extensive comparison of both classes of maximum likelihood models using several large data matrices with strongly corroborated nodes (> 15,000 characters each)[11]. In addition, Buckley and Cunningham have compared Bayesian Montecarlo Markov Chain phylogenetic inference (MrBayes)[39] to conventional nonparametric ML bootstrapping [11], and will continue to do so with the datasets generated by this project. These MCMC approaches are in their infancy and will be evaluated carefully.

We will also analyze amino acids, using unordered and using the Protpars stepmatrix distributed with MacClade 3.0 [51]. For likelihood analysis of amino acids, we use *protml* within the MOLPHY package[2] to calculate in batch mode the lnL scores for up to 60,000 of the best parsimony trees.

Analyzing multiple data partitions. Choice of data set remains a contentious issue among systematists [9] and therefore we will analyze and present results for total nucleotides, for various nucleotide subsets, and for amino acids. Using recovery of concordance groups as a gauge, Regier and Shultz, like others [61, 77, 78], find that combining characters from multiple genes generally leads to the best estimate, even when data sets may appear to be in conflict (e.g., when $P < 0.05$ in the ILD test [24] as implemented in PAUP*; [94]). Although PI Cunningham formerly advocated the ILD as a combinability test at p values > 0.01 [16], he now believes that the test is much too conservative in this regard, even at p values < 0.01 .

In an analysis with 120 gene fragments, considering each gene separately is simply not feasible, and gene fragments will be combined. Although we will combine genes, this will be done in the context of partitioning of synonymous and non-synonymous substitutions, which has been necessary to recover concordance groups with the 3 nuclear protein coding genes analyzed by co-PI s Regier and Shultz (Figure 1) [77]. We realize that this conclusion is dependent on taxon sampling and that synonymous changes can be informative at deep levels. However, in higher-level arthropod phylogeny, where inter-node distances may represent hundreds of millions of years, concordance group recovery invariably decreases when nt3 where most synonymous change is localized is included. Nt3 evolves 10 - 50 times faster than nt2 [71, 77] and can account for at least 90% of total character change. Nt3 base composition, unlike that of nt1 and nt2, is always non-homogeneous when tested using PAUP* [94]. Indeed, we can get further improvement in concordance group recovery by also removing that subset of nt1 characters that encode leucine- and arginine-encoding codons, for which synonymous change is also possible [77]. The significance of alternative hypotheses with different data sets will be tested using the method of Kishino and Hasegawa [45], as well as with recent improvements to this test.

Addressing concerns about paralogy. As described above in our screening procedures for choosing genes (Step #5, page 9) Regier and Shultz routinely encounter very closely related paralogs for the three genes they have sequenced to date. This finding is consistent with the hypothesis of frequent duplication and rapid extinction of gene copies: [50]. Deep paralogy is of much greater concern, and can lead to misleading phylogenetic inference [38]. Although our screening procedure is designed to minimize this possibility, there are some circumstances, where lineage sorting has left only one, but distinct paralogs in different lineages, that cannot be addressed by our screening. Fortunately, our intent to collect 120 gene fragments per taxon will minimize the misleading effect of any undetected paralogous genes.

INTEGRATING FOSSIL AND MOLECULAR DATA IN A BAYESIAN FRAMEWORK TO GENERATE A TIMELINE FOR ARTHROPOD EVOLUTION

A comprehensive picture of arthropod evolution will be incomplete without estimates of the timeline of arthropod evolution. The rich fossil record of arthropods provides invaluable insight into the evolutionary history of this group [5]. However, even the fossil record of arthropods is incomplete. Although paleontological and molecular data are sometimes viewed as separate and conflicting sources of information, sequence and fossil information can be exploited in a complementary fashion. We propose to take a two-pronged approach to developing a timeline for arthropod evolution. Our morphologists (co-PI s Martin and Shultz) will use the rich fossil record of arthropods [5] to estimate minimum dates of divergence for as many nodes as possible in our phylogeny. These estimates will be used as constraints on the Bayesian models that will be developed by co-PI Thorne and his collaborator Hirohisa Kishino (see accompanying letter of support).

Bayesian approaches to divergence times in the Arthropoda

Conventionally, estimates of divergence times are based upon the assumption of a molecular clock. Clock-based methods suppose the rate of sequence evolution is constant over time and they require fossil information to calibrate this rate. In reality, rates of molecular evolution change over time. For example, mammals and bacteria certainly do not evolve at the same rate. Moreover, fossil evidence

cannot usually determine the exact date at which two evolutionary lineages diverged and therefore there is uncertainty attached to estimated rates of molecular evolution.

Recently, a number of improvements to traditional clock-based techniques have been introduced [21, 40, 46, 79, 96]. In this project, divergence times of arthropods will be inferred according to these new methods. Particular emphasis will be placed on the non-clock (non-clock) method developed by co-PI Thorne and collaborator Kishino (see supporting letter from Kishino [46, 96]). With the non-clock method, the constant rate assumption of a molecular clock is relaxed. In particular, this non-clock approach allows rates of molecular evolution to change over time. The idea is that rates of evolution largely depend on biological factors (e.g., mutation rate, generation time, natural selection and genetic drift). Therefore, closely related lineages tend to evolve at similar rates whereas more distantly related lineages tend to evolve at more different rates.

With the non-clock method, the rate of molecular evolution at the end of a branch on a phylogenetic tree depends on the rate at the beginning of the branch. Given the rate at the beginning of the branch, the logarithm of the rate at the end of the branch is assumed to have a normal distribution. The mean of this normal distribution is such that the expected rate at the end of the branch is equal to the rate at the beginning. The variance of this normal distribution is the product of the time duration of the branch and a rate variation parameter. A value of 0 for the rate variation parameter corresponds to the constant rate assumption of a molecular clock and the departure from a molecular clock increases as the value of the rate variation parameter increases.

Because it is a Bayesian approach, the non-clock method requires prior distributions for divergence times, the rate variation parameter, and the rate at the root of the tree. The method combines this prior information with information from the sequence data to yield posterior estimates of divergence times. Constraints on node times are now also incorporated into the method. When a lineage is known from fossil evidence to be at least a certain age, this can now be properly taken into account. Likewise, nodes can be constrained to be no older than a specific age. The presence of constraints appears to greatly improve divergence time estimates [46]. Additional stratigraphic information could also potentially be incorporated into analyses [41].

Because this method for estimating divergence times has been reported elsewhere [46, 96], we omit further details here. However, we do need to explain a new and unpublished extension of the method here. In the past, the non-clock method has been implemented only for the analysis of single gene data sets. Some recent projects for estimating divergence times with clock-based methods have been based on data sets where each taxon is represented by sequences from multiple genes [48]. Because each set of gene sequences in an interspecific phylogeny can be assumed to share a common set of divergence times, a data set consisting of multiple genes contains more information about divergence times than does a data set consisting of only a single gene. The non-clock method can now be applied to multiple gene data sets. In the current implementation, all genes are assumed to share a common set of divergence times but the evolutionary rates of genes change independently of one another over time. Planned future implementations will enable a tendency for genes to have *a priori* correlated changes of rates of molecular evolution. Another limitation that will be relaxed in the near future is the assumption that all genes deviate the same amount from the molecular clock. In other words, the assumption to be relaxed is that all genes share the same value of the aforementioned rate variation parameter. We are optimistic that current and future versions of our multiple gene non-clock method will substantially improve estimates of arthropod divergence times. In addition, these approaches will shed light on how rates of molecular evolution in arthropods have changed over time.

One role of co-PI Thorne and the postdoctoral associate will be to help to carefully analyze the sequence data generated by other investigators in this project. Particular emphasis will be on estimating arthropod divergence times and exploring patterns of rate evolution that are evident in arthropod evolution. The postdoctoral associate will also be responsible for producing more user-friendly versions of the software that Thorne has written. Although the current, non-user-friendly software has not been advertised, Thorne has recently received roughly one request per week for his divergence time estimation software. He freely distributes this software to all who request it. However, many users of these

programs have been understandably frustrated by the incomplete documentation, the specialized data format, and the lack of a user interface. Thorne simply does not currently have the resources to devote to substantial improvements of this software. As well as analyzing the arthropod sequence data, the postdoctoral associate will be charged with improving the software. We are not unrealistic enough to expect that this software will be as professional as packages such as PAUP* and we have no intent to duplicate the many tasks that other software performs so well. Instead, the software will focus on divergence time estimation and exploration of rate evolution. Recently, Drs. David and Wayne Maddison have announced the "Mesquite" project[52]. The goal of this project is to create a modular software system for studying evolution. Individual scientists can author modules that perform tasks that interest them and can rely upon modules written by others. We plan to explore the feasibility of converting the non-clock method software into modules for this Mesquite system.

EDUCATING A NEW GENERATION IN ARTHROPOD EVOLUTION AND DIVERSITY

Because of their incredible diversity and ubiquity arthropods are of special fascination to the public at large. A key element of our proposal is to develop a serious educational component to our study of arthropod phylogeny. This program will be based at the Los Angeles County Museum, and will provide a model that can be applied elsewhere on the country.

A well-planned, hands-on After School Program (ASP) targeting local high school students is already in place at the Natural History Museum of Los Angeles County. The stated objective of the Museum's ASP is to meet the needs of students in inner-city Los Angeles by providing free, educational, meaningful and safe activities during after school and off track hours, while enhancing and expanding their academic and career horizons in science and history. The ASP is implemented by Museum educators who work with Museum curators and community representatives to develop programs and collaborative efforts designed to use the Museum's resources in ways that effect measurable changes in children's attitudes and knowledge about science, culture, and history.

This unique and innovative program trains local high school students from predominantly minority communities and schools. The high school students in turn serve as mentors for younger (elementary and middle school) students, both at the Natural History Museum and in local schools. The high school students serve as a bridge between the Museum and the community, by making the Museum and its resources more accessible to the community and by serving as voices of the community in the development of the Museum's exhibits and programs. This program is therefore an incredibly effective way of communicating science to local school children, using fellow students whom they respect as the primary vehicles for science education. As one indication of the cascading and successful effects of the program, below are figures from the July-December 2001 ASP season:

<u>Attendance</u>		<u>Contact Hours</u>	
High School students (interns)	23	High School students	2,487
Middle School students	65	Middle School students	488
Elementary students	300	Elementary students	600
Families	250	Families	500
Total Attendance	627	Total Contact Hours	4,075

The impact of this program already has had a tremendous effect on the lives of these inner-city teens; they come from a community where the dropout rate is 10.9 percent — more than double the California state average. Their community is nationally and internationally known for its violence, substance abuse, teen pregnancy, high unemployment, and poor academic performance; and they come from families where no one has graduated from college.

In anticipation of the currently proposed project, we have designed a new component of the Los Angeles ASP curriculum that targets Arthropod Diversity and Evolution. The program will enable students to work alongside active museum curators, postdoctoral researchers, and collections managers, and will teach students to identify, classify, and curate scientific specimens while exploring related issues such as adaptation, biological classification, biodiversity, and tissue extraction. Selected specimens

collected as part of the ASP will be used by researchers in Los Angeles and by the Duke and UM teams; tissues taken from these specimens will eventually form part of the arthropod genomic database.

Activities of the Arthropod Diversity and Evolution ASP will include:

- Tours of the museum displays that highlight or include arthropods
- Behind-the-Scenes tours of arthropod (Crustacea, Entomology) collections
- Interactions with Museum curators and collection managers
- Active participation in field work, including collection of some specimens actually used in the research project.

The connection between Los Angeles students and students in the areas served by Duke and the University of Maryland is twofold. First, students in the Los Angeles ASP and participating undergraduates and graduates at Duke and UM will be able to track each specimen they collect for the project. All specimens will have the student collectors name attached, and students will thereby be able to follow the life of a biological specimen used for genomic analysis. Second, student-designed web sites designed at Duke, but incorporating student ideas and suggestions and components from UM and Los Angeles, will provide updates on progress of the overall project, names of all participants, forthcoming publications, new developments in the field, and lists of any additional specimens needed for the project.

In Los Angeles, co-PI Martin and R. Wetzer will present talks on arthropod and especially crustacean diversity. PIs Cunningham and Regier will also present talks to the students at least once during the cycle (see below) during working visits to Los Angeles. Additionally, students will be exposed to dissection of readily available arthropods (e.g., crayfishes and lubber grasshoppers available through biological supply houses such as *Carolina*). Field trips will be designed to expose students to modern collecting techniques and will include (1) trips to local terrestrial arthropod high diversity sites, such as the Biona Wetlands, targeting insects and arachnids, and (2) trips to Catalina Island to target marine arthropods, with emphasis on crustaceans.

We are requesting funds to hire an Education Specialist for one year to oversee this program, after which time we plan to obtain matching (non-NSF) funds to continue his employment. This person has already been identified; he is Antonio Solorio, a gifted bilingual education specialist currently working as the After School Program Assistant Coordinator at the Natural History Museum. We are also requesting funds for 20% time for his immediate supervisor (our current full time ASP Coordinator) to devote time to this program, and funds to offset the costs involved in three student field trips to collect arthropods at \$1,000 per field trip.

The proposed timeframe for the Los Angeles ASP program on arthropod diversity and evolution is designed to correspond and comply with the LAUSD (Los Angeles Unified School District) schedule, and is as follows:

- January — February, 2002 (first 8 week session)
- March — April, 2002 (second 8 week session)
- July — August, 2002 (third 8 week session)

Concluding statement

Large-scale, but targeted, genomic sequencing offers a promising approach to resolving higher-level arthropod phylogeny. The arthropod tree, once completed, will be both a final product and a starting point for the renewed interest in arthropod evolution. It will guide classical paleontologists in their search for intermediates; it will enable molecular paleontologists to better date key events in the history of life; and it will help all to make more rational assessments of the "tempo of life." It will guide evolutionary morphologists in identifying synapomorphies, symplesiomorphies, and parallelisms. It will provide functional biologists with a framework for understanding the evolution of developmental mechanisms.

Taxa to be used in the proposed study. Asterisks indicate that specimen is extraction-ready.

MYRIAPODA

CHILOPODA (centipedes): *Anopsobius neolandicus**, *Ballophilus australiae**, *Cormocephalus monteithi**, *Craterostigma tasmanianus**, *Scutigera coleoptrata**

DIPLOPODA (millipedes): *Docodesmus trinidadensis**, *Glomeridesmus trinidadensis**, *Polyxenus fasciculatus**, *Rhinotus purpureus**, *Stemmiulus insulanus**

PAUROPODA: *Allopauropus proximus**

SYMPHYLA (garden "centipedes"): *Scutigera sp.**, *Symphylella sp.*

CHELICERATA

PYCNOGONIDA (sea spiders): *Endeis leavis**, *Tanystylum orbiculare**

XIPHOSURIDA (horseshoe crabs): *Carcinoscorpius rotundicauda**, *Limulus polyphemus**

ARACHNIDA (spiders, scorpions, etc.): *Aphonopelma chalcodes**, *Centruroides culpturatus**, *Mastigoproctus giganteus**, *Nipponopsalis abei**

HEXAPODA

PROTURA: *Acerella barberi*, *Eosentomon vermiforme*

COLLEMBOLA (springtails): *Hypogastrura sp.**, *Metisotoma grandiceps**, *Orchosella imitari**, *Podura aquatica**, *Sminthurus sp.**, *Tomocerus sp.**

DIPLURA (two-tailed bristletails): *Campodea sp.*, *Eumesocampa frigilis*, *Metajapyx subterraneus**, *Parajapyx isabellae*

MICROCORYPHIA (jumping bristletails): *Machiloides banksi**, *Pedetontus saltator**

ZYGENTOMA (bristletails, silverfish): *Ctenolepisma lineata**, *Nicoletia meinerti**

PTERYGOTA (winged insects): *Acheta domesticus**, *Allocapnia sp.**, *Calopteryx maculata**, *Forficula auricularia**, *Hexagenia limbata**, *Periplaneta americana**, *Drosophila melanogaster**

CRUSTACEA

BRANCHIOPODA: Anostraca (fairy shrimp): *Artemia salina**, *Streptocephalus peali**.

Conchostraca (clam shrimp): *Limnadia lenticularis**, *Lynceus gracilicornis**. Notostraca

(tadpole shrimp): *Triops longicaudatus**. Cladocera (water fleas): *Bythotrephes longicaudatus*, *Daphnia magna**

CEPHALOCARIDA: *Hutchinsoniella macracantha**

MALACOSTRACA: Phyllocarida: *Nebalia hessleri**; Hoplocarida (mantis shrimp, etc.):

*Neogonodactylus oerstedii**. Anaspidacea: *Anaspides tasmaniae*; Peracarida (isopods,

amphipods, etc.): *Armadillidium vulgare**, *Americhorchestia megalophthalma**. Lophogastrida:

Lophogaster sp. Decapoda (crabs, lobsters, true shrimp, etc.): *Penaeus setiferus**, *Stenopus hispidus**; *Callinassa sp.*, *Libinia emarginata**

MAXILLOPODA (fish lice, barnacles, copepods, etc.)

Branchiura: *Argulus nobilis**. Cirripedia: *Lepas ansifera**, *Semibalanus balanoides**, *Chthamalus*

*fragilis**, *Loxothylacus texanus**. Copepoda: *Eurytemora affinis**, *Bryocamptus sp.*, *Diacyclops sp.*,

Mystacocarida: *Derocheilocaris typica*. Pentastomida: *Porocephalus crotali*

OSTRACODA (seed shrimp): *Skogsbergia lerner**, *Harbansus paucichelatus**, *Cypridopsis sp.*,

*Manawa staceyi**

REMIPEDIA: *Speleonectes tulumensis**

OUTGROUPS

ONYCHOPHORA (velvet worms): *Euperipatoides rowelli**, *Peripatus sp.**

TARDIGRADA (water bears): *Milnesium tardigradum**, *Echiniscus viridissimus**, *Macrobiotus islandicus**, *Testechiniscus sp.**