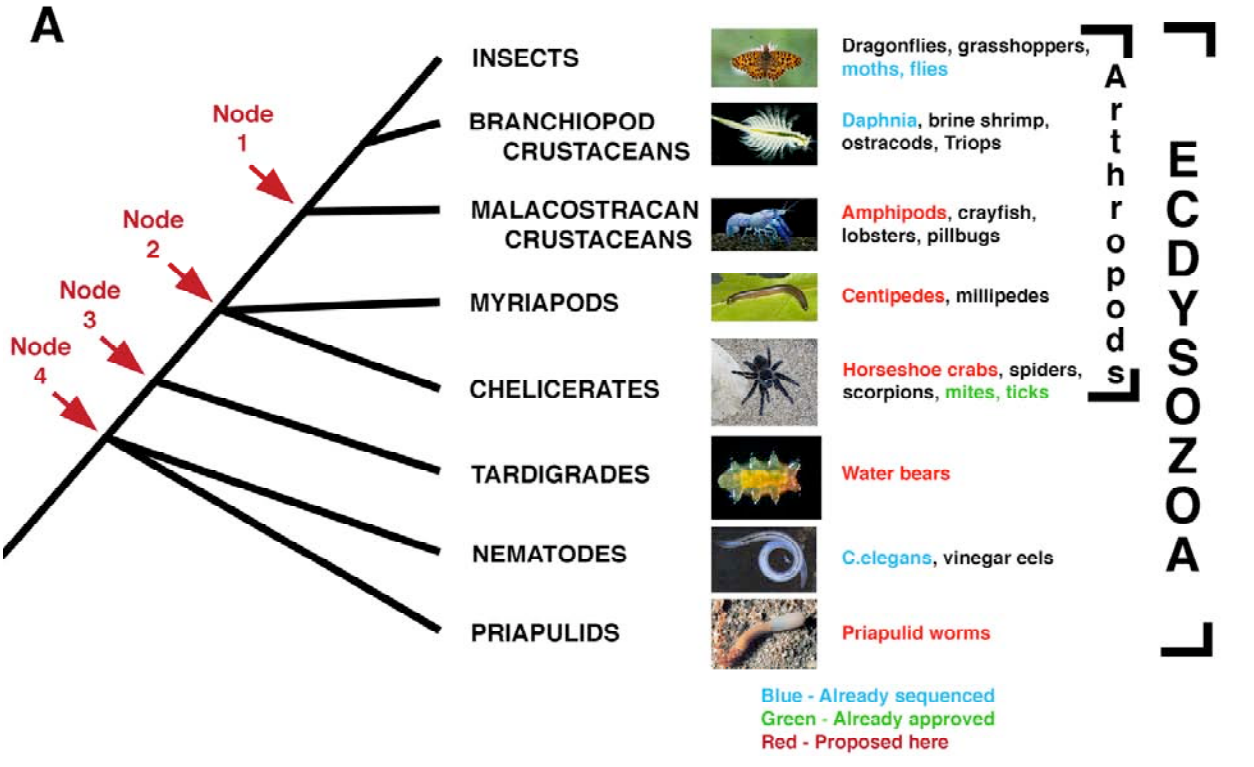
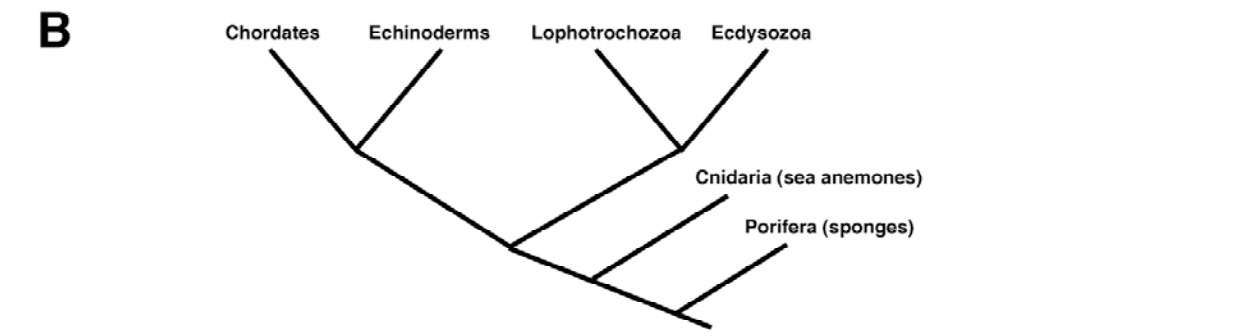
Ecdysozoan Sequencing Proposal

Nipam Patel (University of California, Berkeley, USA), Dan Rokhsar (Joint Genome Institute, USA), Max Telford (University College London, UK), Graham Budd (University of Uppsala, Sweden), Michael Akam (Cambridge University, UK), Ariel Chipman (Hebrew University, Israel), Bob Goldstein (University of North Carolina, USA), Mark Blaxter (Edinburgh, UK), Michael Eisen (University of California, Berkeley)

Purpose: To better understand the composition of the transcriptome and proteome at different evolutionary nodes within the Ecdysozoa (see Figure 1A), and to ultimately reconstruct the genome of the ancestral Ecdysozoan. These data will improve the annotation and interpretation of the genomes of the Ecdysozoan genetic model systems, *Drosophila melanogaster* and *C. elegans*. In turn, this will improve our understanding of the evolution and function of genes that are found in other organisms, including humans. Finally, the data will also help us understand macroevolutionary mechanisms, and will be particularly useful in investigating the evolution of development.

Figure 1 Evolutionary Relationships





Scope of Work:

The five species for analysis have been chosen by consulting the community currently working with various Ecdysozoa.

Species	Genome Size (in Mb)	Sequence Coverage	# cDNAs (paired end reads)
Jassa slatteryi Amphipod crustacean	690	5 X	100,000
Strigamia maritima Geophilimorph centipede	290	5 X	100,000
Limulus polyphemus Chelicerate (horseshoe crab)	270	5 X	100,000
Hypsibius dujardini Tardigrade	70	5 X	100,000
Priapulus caudatus Priapulid worm	500	5 X	100,000

Achieving an appropriate "phylogenetic spread" to adequately begin reconstructing the nodes of Ecdysozoan evolution was the most important consideration; in addition several of these species are potentially experimentally tractable for future genetic, developmental, and physiological studies. Many Ecdysozoan genomes are quite large, but we attempted to chose animals with relatively small genomes, but avoided targeting the tiniest genomes we could find within the different groups for reasons that will be discussed below.

In addition, we propose the generation of a BAC library for four of these species (*Jassa slatteryi*, *Strigamia maritima*, *Limulus polyphemus,* and *Hypsibius dujardini*; a BAC library is already under construction for *Priapulus caudatus*). We also propose 6X BAC end sequencing from all the species.

In total: 9,050 Mb genomic shotgun sequence + 500,000 cDNAs (paired end reads; approx. 750 Mb) + 6X BAC end reads (approx. 60 Mb)
= 9,860 Mb (equivalent to about 3X coverage for a single human-sized genome)

Supply of DNA and cDNA libraries

In general, methods for isolating high quality high molecular weight genomic DNA from arthropods has proven relatively straightforward. The basic protocol for *Drosophila* has already been used successfully to isolate high molecular weight DNA from both *Jassa* and *Priapulus*, and we believe that similar approaches will be successful in other Ecdysozoa. While we do not anticipate any serious problems, we will certainly work to resolve any that are encountered, and

will work with the sequencing centers to ensure that high quality DNA is available for both the whole genome sequencing and BAC library construction.

For EST sequencing, the user community will provide the appropriate cDNA libraries. The libraries will meet criteria that will be negotiated with the sequencing centers (for example, at least 95% of the clones must have inserts, average insert size should be at least 1.0 kb, etc.). The libraries will be normalized to increase the usefulness of such deep sequencing. We anticipate that sequencing will be done in batches, and at appropriate steps, an analysis of the return of novel sequences will be made to a determine if further sequencing of the library is warranted.

Evolutionary Considerations:

The bilaterian animals are divided into two main clades, the Deuterostomes and the Protostomes. The Deuterostomes include two particularly well known phyla, the Chordata (which includes humans) and the Echinodermata (sea urchins, starfish etc.) (Fig. 1B). The Protostomes are further subdivided into two large clades, or Superphyla, known as the Ecdysozoa and the Lophotrochozoa (Fig. 1B); in turn the Lophotrochozoa include a number of phyla including the Mollusca (snails, clams, etc.) and Annelida (earthworms, leeches, etc.), while the Ecdysozoa include a phyla such as the Arthropoda (insects, crustaceans) and the Nematoda (roundworms such as C. elegans).

Of all of these animal groups, it is the Ecdysozoa that contains the vast majority of described animal species, this is largely because of the diversity of arthropods. Even taking into account various estimates of "undiscovered" species, it is still the Ecdysozoa that rank as the most species-rich group, and this clade also contains some of the most extreme variations in morphology and life history. Within the Ecdysozoa lie two of the most powerful genetic model systems, *Drosophila melanogaster* and *C. elegans*. Not only have the genomes of these two organisms been completely sequenced, but multiple species that are closely related have also been completely sequenced to complement the work in the two model species. While these two species have been, and will continue to be, premier genetic systems for mechanistic and detailed studies in many fields of biology, it has become clear that they have many unusual traits that make them quite different from other members of the Ecdysozoa, and have biased many of our views on animal evolution.

In particular, the genomes of *Drosophila* and *C. elegans* are somewhat unusual. They are extremely compact, which indeed probably contributed to their adoption as genetic model systems, and certainly played a role in the decision to sequence their genomes. But many studies indicate that they may not be representative of a typical extant Ecdysozoan, and also not representative of the ancestral Ecdysozoan. For example, vertebrates contain 12 *WNT* family members (*WNT* genes are a family of important cell-cell signaling molecules whose founding member is the *wingless* gene of *Drosophila*). Only three family members are found in *Drosophila*, and *C. elegans* also contains a small set of highly derived *WNT* family members. This data was used to suggest that the diversification of *WNT* genes occurred in the vertebrate lineage. Analysis of the phylogenetically basal cnidarian, the sea anemone *Nematostella vectensis* (see Fig. 1B for phylogenetic position of cnidaria), reveals that cnidaria have 11 of the 12 *WNT* family members found in vertebrates (Kusserow et al., 2005). This certainly changes how we view the evolution of the *WNT* family members and suggests that the small number in *Drosophila* was due to a gene loss. Another key developmental gene, *noggin*, is also absent

from *C. elegans* and *Drosophila*, but present in vertebrates and cnidaria (Rokhsar et al, unpublished data). The question remains, however, where in evolution these gene losses occurred, and points to the limitation in using *Drosophila* and *C. elegans* to assess the evolutionary history of genomes. A reciprocal BLAST analysis of a cDNA survey in the crustacean, *Parhyale hawaiensis*, revealed that of 2500 coding sequences, 92 had a clear homolog in humans (e-value $< e^{-20}$), but no clear match in *Drosophila* (e-value $< e^{-10}$). Similar types of studies reveal that the loss of introns in *Drosophila* and *C. elegans* genes also represents an evolutionarily derived state. The secondarily derived breakup of the *Hox* complex in various nematodes and *Drosophila* species is one more example of some of the unusual genomic events in these two lineages, and in this case we know that this differs from the "less-derived" lineages in the Ecdysozoa (for example, the *Hox* genes in the crustacean, *Daphnia*, are organized into a single compact cluster reminiscent of the clusters found in vertebrates. All this points to the need for better genome sampling within the Ecdysozoa.

We also feel that the past tendency to choose almost exclusively small genomes for sequencing within the Ecdysozoa may also contribute to the biased views we have developed regarding genome evolution. While the monetary considerations are very important, a continued focus on small genomes may lead us to mistakenly view secondarily derived characteristics of these small genomes as instead ancestral characteristics. The evolutionary pressures, whatever they might be, to reduce genome size in certain lineages has led to similar phenomena of gene loss, shortening of exons, and reduction in introns. These traits then make it difficult for us to recognize the genomic changes that are unique to the vertebrate lineage. It also complicates the assignment of gene orthology and comparisons of gene functions between humans and other animals, including the genetic model systems we use as proxies for deducing the role of these genes in normal development and in human disease. As described above, the consequences of genome compaction and other specializations that go along with rapid generation times in *Drosophila* and nematodes are just beginning to be analyzed, and data from other Ecdysozoans, particularly those that have not undergone extreme reduction in genome size, will be critical in making proper evolutionary assessments.

Phylogenetic Positions of Species for Sequencing:

In this proposal, we suggest that whole genome sequencing of five Ecdysozoan species will provide the coverage needed to develop a more accurate assessment of genome evolution in Ecdysozoa, and by extension a more accurate analysis of transcriptome and proteome relationships across all animals. We describe our choices using a "node-based" approach, describing how each species helps to fill in an important gap in our knowledge (refer to the tree diagram shown in Figure 1A). Below we describe the specific nodes accounted for by each species we have selected, and then follow this with a more detailed description of the animals and how they will also be useful to other areas of biological investigation.

As mentioned previously, sequencing has been completed for twelve species of *Drosophila*. Several other Diptera, such as the mosquito, are also finished. A few additional insects have also been sequenced, or will soon be sequenced. While arguments can be made for the need to sequence additional insect species whose lineages branch near the base of the insect radiation, these insects tend to have very large genomes (for example, grasshoppers and silverfish both have genomes close to the size of the human genome), and thus we have not proposed any additional insect species for sequencing as part of this proposal. Current phylogenetic analyses suggests that the sister group to the insects is the branchiopod crustaceans, which include

Artemia (brine shrimp) and *Daphnia* (water fleas). The sequencing of *Daphnia pulex* has been recently completed, and for that reason we are not suggesting another branchiopod crustacean as part of this proposal.

Node 1

Understanding the next node along the tree (node 1 in Figure 1A) requires sequence from a malacostracan crustacean. The malacostracan crustaceans include crabs, lobsters, isopods, and amphipods. No species in this group has been sequenced or is currently approved for sequencing. Species in this group tend to have quit large genomes (typically 1 to 5 Gb), but we have picked a member with one of the smaller genomes. The amphipod crustacean, *Jassa slatteryi*, has a genome size of 690 Mb, which is only 10 Mb larger than the smallest recorded genome within the malacostracan (a caprellid shrimp). While here we have picked a genome on the smaller end of the scale for malacostracan crustaceans, it would appear that no known members of this group have gone through the kind of genome compaction seen in *Drosophila* and *C. elegans*.

The reconstruction of this node is particularly important because the descendant clade of insects plus crustaceans, sometimes referred to as pan-crustacea, represents the majority of Ecdysozoan species and the majority of the morphological, developmental, ecological, and behavioral diversity for the Ecdysozoa. Understanding the diversification events that have occurred from this node will add significantly to our understanding of evolutionary processes, especially the genomic underpinnings of this remarkable radiation.

Node 2

Data from both a myriapod and a chelicerate will be needed to understand node 2. At the moment, current phylogenies place myriapods and chelicerate as sister taxa, or have myriapods as being closer to the pan-crustacea than chelicerates. Either way, node 2 represents the ancestor of all arthropods, and both a myriapod and a chelicerate sequence are needed. We propose to sequence the centipede, *Strigamia maritima,* which has a genome of approximately 290 Mb. This would be the first myriapod to be sequenced. As for a chelicerate, we recommend *Limulus polyphemus*, the horseshoe crab, which has a genome of 270 Mb. Two other chelicerates have already been put forward and approved previously for whole genome sequencing. These are the deer tick, (*Ixodes scapularis*) and a spider mite (*Tetranychus urticae*). The deer tick has a genome of about 2,000 Mb and sequencing has yet to begin. The spider mite has a remarkable small genome of 70 Mb, which is far from the typical chelicerate genome of several Gigabases. We suggest that the spider mite will have many of the other issues associated with animals that have evolved such a compact genome, and it has a very derived mode of development which make it far from typical for chelicerates. For these reasons we are proposing that *Limulus* be sequenced as a representative chelicerate.

Node 3

Recent molecular studies suggest that one of the closest out-groups to the arthropods are the tardigrades. The small animals appear to have quite small genomes, and we suggest that the 70 Mb genome of *Hypsibius dujardini* be sequenced.

Node 4

The most basal branching members of the Ecdysozoa appear to be the nematodes and the priapulid worms (the exact topology of the relationship between these two remains a matter of

debate). In addition to C. elegans, a number of other nematode worms have also been sequenced, but as described above, their small genomes and derived lifestyles make it important to also have sequence from an animal that might not be as evolutionarily distant from the rest of the Ecdysozoa. To accomplish this, we suggest whole genome sequencing of the 500 Mb genome of the priapulid worm, *Priapulus caudatus*.

Further Details and Implications for Understanding the Evolutionary Diversification of Ecdysozoa

As described previously, the data obtained will be valuable to all scientists who use genomic analysis to study animal transcriptome and proteome evolution. In addition, the genome information will be invaluable to the thousands of laboratories that use various Ecdysozoa in their experiments. For example, amphipod crustaceans such as *Jassa* are often used to test and monitor water quality. These animals are particularly sensitive to insecticides, and whole genome data will allow for much more detailed analyses (through microarrays for example). Likewise, arthropods such as Porcelein crabs are used to understand the mechanisms of adaptation to stress (fluctuations in water temperature and salinity) and microarray data is now being used to monitor the reaction to these conditions, especially as they are likely to change as a result of global climate shifts. A better understanding of the evolutionary origins of nematodes may well help direct research programs designed to fight diseases caused by these animals. Tardigrades show remarkable tolerance to freezing and desiccation, and whole genome data would help stimulate research into the remarkable properties of these animals.

One research area that will certainly benefit from additional genome data from Ecdysozoa is the field of evolution and development, or "evo-devo". The community of researchers in this field have made excellent use of the genomic data available to them so far to try and understand the mechanisms of developmental and morphological evolution. For example, within the insects, whole genome analysis in the beetle (*Tribolium castaneum*) and the bee (*Apis millifera*), have led to the rapid identification of the genes that play a role in axis specification, and the discovery of both protein coding changes and cis-regulatory changes that have played a role in the evolutionary diversification of embryonic patterning these lineages. Again, many different research communities will benefit from these Ecdysozoan genome sequences, not just the evo-devo community. While the evo-devo community may be small by some standards, it is exceptionally active and well-poised to make immediate use of the genome data; and the evo-devo community is just a small subset of the researchers who will make use of the ecdysozoa in future research of all kinds. In providing more details about each species further below, we also describe briefly how each animal is important to the evo-devo research community.

Specific evo-devo areas that could be addressed; an example of the use of the data for future experimental investigation.

Below we provide just a few examples of the specific kind of investigations that could be undertaken with the sequence information.

1) A better understanding of the mechanisms of neurogenesis, and its evolutionary origins and history. Studies of neurogenesis (and axonogenesis) in *Drosophila* have provided us with many insights into the mechanisms that are at work in other animals, including humans. The earliest stages of *Drosophila* neurogenesis, however, are very distinct from those in vertebrates (although similar genes are used). Recent findings suggest that chelicerate and myriapod neurogenesis is quite different from that seen in insects, such as *Drosophila*, and might be more reflective of an

ancestral arthropod mechanism which in turn would inform us about possible ancestral vertebrate mechanisms, and potentially those at work in extant vertebrates as well. In addition, malacostracan crustaceans such as *Jassa* display a remarkable ability to have neural stem cells switch back to an ectodermal fate, and then once again back to being neural stem cells. Understanding how this works might be quite useful in regenerating neural tissue in humans. Data on neurogenesis from Onychophoran and Priapulids would also be useful in polarizing the direction of change.

2) Arthropods possess a remarkably diverse range of appendage morphologies, indeed this variation has contributed significantly to their success. Appendage patterning has been addressed in some detail in insects (with *Drosophila* leading the way). Outside of the insects, however, much less is known about appendage development and what contributes to the extremes of morphological variation that are observed. Data from priapulids and onychophora will be valuable as they possess relatively simple limbs that are thought to be closer to the morphology of the ancestral state for the Ecdysozoa.

3) The evolution of segmentation and body regionalization are probably the most intensely studied "evo-devo" topics for the Bilateria, and are particularly relevant to understanding the evolution of Ecdysozoa (especially arthropods). Many questions remain as to the role of Hox genes in animal evolution, and nowhere is this issue more investigated and debated than in the arthropods. The Hox genes in arthropods also provides us with an excellent opportunity to approach problems of cis-regulatory evolution and its role in morphological innovation. Understanding the evolution of segmentation in various arthropods will help us understand how the system of segmentation evolved to work as it now does in *Drosophila*, as well as answer questions about how much variation in the process is present in other arthropods. In this regard, there is particular excitement in uncovering potential links between the segmentation of arthropods, annelids, and vertebrates.

In the end, understanding how these genes function and are regulated in a variety of arthropods will be important, but it will be the comparisons made between the different arthropods and Ecdysozoa that will answer numerous macroevolutionary questions about development. In addition, the sequence data obtained will also help improve our understanding of the phylogenetic relationships of all the Ecdysozoa, and how the Ecdysozoa are related to the other major animal groups. The data will also improve our understanding of the genome content and biological properties of several important evolutionary nodes. In particular, our understanding so far of the ancestral Ecdysozoa is heavily biased by *Drosophila* and *C. elegans*. While these are outstanding genetic systems, their modes of development and the composition of their genomes are highly derived (for example they are both missing many genes that were presumably present even in the ancestor of all bilatarians; and both appear to have reduced the size of their genomes). Thus, the broader coverage provided here will be very informative in reconstructing not only the ancestral Ecdysozoan node, but the ancestral Protostome node as well, which will ultimately provide us with better information on the lineage leading to humans. Furthermore, it is already apparent that some developmental processes are better conserved between humans and more "typical" arthropods that between humans and model systems such as *Drosophila* and *C. elegans*. For example, it appears that the mechanisms of segmentation in vertebrates shares more in common with chelicerates than with segmentation in *Drosophila*. Such insights clearly have implications for the study of human health an

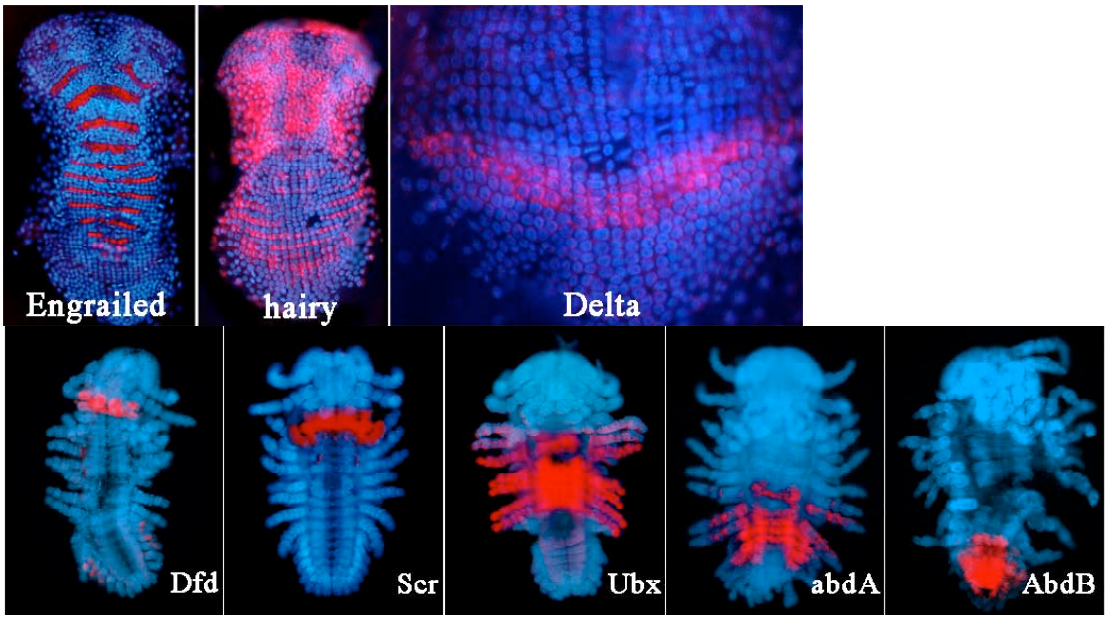
SOME SPECIFIC DETAILS FOR EACH ORGANISM

Jassa slatteryi (malacostracan amphipod crustacean; 690 Mb genome)

Adult male *Jassa slatteryi*

Adult male *Parhyale hawaiensis*

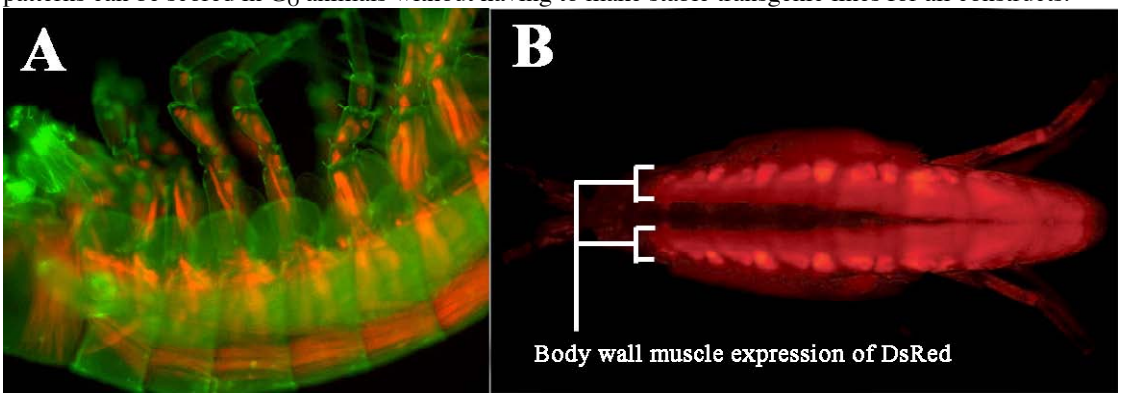
The amphipod crustaceans, *Jassa slatteryi* and *Parhyale hawaiensis* are becoming established "new model" organisms for developmental studies within the arthropods. Detailed knowledge of *Drosophila melanogaster* development, the diversity of extant body plans, and a large body of existing comparative work currently make the arthropods the premier group for studying many questions in development and evolution. Crustaceans in particular show a remarkable range of diversity, and provide an extremely useful outgroup comparison to all the insects. Amphipods such as *J. slatteryi* have a phylogenetically strategic position with respect to current developmental models and to existing sequenced and in progress genomes, and have excellent potential as experimental organisms for studying developmental processes. *J. slatteryi* and *P. hawaiensis* are relatively closely related amphipods – from cDNA sequence comparisons, we estimate that they are roughly the same molecular distance apart as *D. melanogaster* and *D. virilis*. While most techniques have been worked out with *P. hawaiensis*, the 3.6 Gb genome of *P. hawaiensis* would make it hard to justify a whole genome sequencing effort at the current time. Thus we have chosen *Jassa* instead, with its 690 Mb genome, as our selection for whole genome sequencing. In the future, the two organisms will probably be used in concert, although the availability of whole genome sequence data would probably start to shift the emphasis. At the moment, *Parhyale* is easier to raise in the lab than *Jassa*, but *Jassa* culturing methods are still being perfected. For now, obtaining adults of *Jassa* from the wild is simple (they live all along the Northern California coastline) and an inbred colony has been maintained for three years that can be used as the source of genomic DNA. Embryonic development appears to be largely identical between *Jassa* and *Parhyale*, indeed if not for the smaller size of early *Jassa* embryos, they are not easy to distinguish morphologically from *Parhyale* embryos.



Examples of gene expression data from amphipod crustaceans

Simple and robust methods have been developed that allow gene expression patterns to be characterized at all stages of amphipod development, and several techniques for experimental manipulation are already in use including mRNA injections (Gerberding et al., 2002), cell lineage tracing (Gerberding et al., 2002; Browne, 2005), and cell ablation. Most significantly, techniques for the generation of transgenic animals has been established for *Parhyale* (Pavlopoulos and Averof, 2005) and should be applicable to *Jassa*, as the transposable element has been shown to work in a variety of ecdysozoans and lophotrochozoans. This technique will allow for several types of functional analyses including gene misexpression using constitutive and inducible promoters, and the testing of cis-regulatory elements. Gene knockdown in *Parhyale* has been achieved by dsRNA injections(RNAi), siRNA injections, and morpholino injections (Gerberding, unpublished data, and Liubicich, and Patel, unpublished data). These gene knockdown techniques will be instrumental in determining gene function in all manner of experimental contexts.

Minos transformation of *Parhyale*. In (A), side view of a stable transgenic line showing DsRed expression in muscle (the green is cuticle autofluorescence seen in animals close to hatching). (B) shows a dorsal view of a G_O animal (directly injected at the one cell stage with same Minos construct shown in A) showing that expression patterns can be scored in G_O animals without having to make stable transgenic lines for all constructs.



DsRed muscle expression in stable, transgenic Parhyale using Minos

Other malocostracan crustaceans that have been used by researchers include lobsters and crabs, but these species are much more laborious to raise, reproduce far more slowly (often taking years to reach sexual maturity), and most have genomes as large or larger than even *Parhyale*. One species that is broadly reared, the green crab, has a genome of 1 Gb, but it is difficult to rear in comparison to amphipods, and no methods have been established for manipulating embryos in this species.

From a phylogenetic view, any malocostracan crustacean would be useful in reconstructing node 1, which is an especially important node as it lies at the base of the major ecdysozoan radiation. We think that *Jassa* provides an animal with a reasonable sized genome (again, just about the smallest known for any malocostracan crustacean, but still not one that has undergone the extreme compaction seen in *Drosophila* and *C. elegans*). It evolutionary closeness to *Parhyale* , and possession of typical morphological and developmental characteristics for malacostracan crustaceans, also makes it an attractive organism for future experimental analysis.

Technical Information:
Based on FACS analysis, we estimate the size of the *Parhyale hawaiensis* genome to be 3,600 Mb, and *Jassa slatteryi* to be 690 Mb. We consider our FACS analyses to be quite high quality, and all control samples produced clear results from which the amphipod genome sizes could be reproducibly estimated. We have no information on overall genome G+C content for these amphipods, but from cDNA sequences and a sample of intron sequences we observe a G+C content comparable to that of *Drosophila melanogaster*. Preliminary analysis of intron sequences suggests that polymorphism for these amphipods is around 2%, but we have been inbreeding both species for several years. We do not have details regarding repeat structures but expect to gain some insights into this issue from the sequencing proposed here.

Interests of Proposers:

Early blastomere fates	Matthias Gerberding. Tübingen, Germany Michael Akam and Cassandra Extavour. Cambridge, UK (C. Extavour will soon establish her own lab at Harvard, USA)
Segmentation	Nipam Patel, Univ. of Calif., Berkeley, USA Ernst Wimmer, Göttingen, Germany
Homeotic genes	Michalis Averof, IMBB, Crete, Greece Nipam Patel, Univ. of Calif., Berkeley, USA
Appendage development	Max Telford, University College, London, UK Michalis Averof, IMBB, Crete, Greece
Neurogenesis	Bill Browne and Mark Martindale. Univ. of Hawaii, USA Nipam Patel, Univ. of Calif., Berkeley, USA
Phylogeny	Max Telford, University College, London, UK
Genomics	Dan Rokhsar, JGI and UC Berkeley, USA

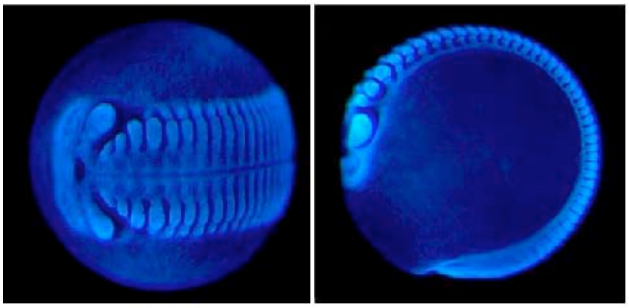
***Strigamia maritima* (Geophilimorph centipede, Myriapod, 290 Mb genome)**

Adult *Strigamia* with eggs

Strigamia maritima is a member of the Myriapoda, a large group that includes both centipedes and millipedes, and which represents one of the four major sub-lineages of arthropods. Myriapods are generally accorded the rank of a class within the phylum Arthropoda, and diverged from other arthropod groups at least 400 million years ago. Until recently the myriapods have been considered sister taxon to the insects, but molecular data consistently place them more basally, as sister group to either the Pancrustacea (insects + crustaceans) or to the chelicerates. At present, there are no genome projects for any member of the Myriapoda. Thus *S. maritima* would be an important addition to the range of arthropod species currently being sequenced.

S. maritima is also becoming established as a model for addressing questions relating to the evolution of segmentation, segment number variation and the differentiation of body regions in arthropods. As a representative of a poorly studied basal branch of the arthropods, it promises also to be valuable in studies of many other aspects of arthropod genome evolution and character evolution -for example, the convergent evolution of characters adapting arthropods to life on land.

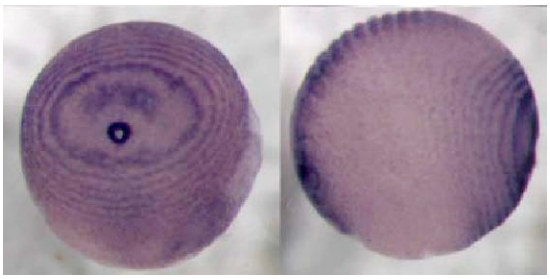
S. maritima was first selected for detailed study by Wallace Arthur and colleagues, for two principal reasons. First, this species shows variation in segment number both within and between populations. However , individuals hatch from the egg with the full adult number of segments. This combination of characters is common to many Geophilomorph centipedes, but exceptional among arthropods more widely. It makes Geophilomorphs particularly well suited for studies of variation in segment number. Kettle and Arthur (2000) defined a geographic cline in segment numbers of *S. maritima* in the UK, and Arthur's group have begun detailed population studies across Europe.



Strigamia embryos

Second, *S. maritima* grows at unusually high population densities. It is an intertidal species, living in shingle below debris at the high tide line. In favored localities, population densities

reach hundreds per square meter. Thus large population samples can be obtained readily, without concern of oversampling. The species has not been induced to breed in the laboratory, but during the summer breeding season, thousands of eggs can be collected in a single day from a few square meters of beach. Large samples of eggs and adults have been collected from a single beach in Brora (NE Scotland) for each of the last six years by the groups of Wallace Arthur and Michael Akam.



Odd-skipped in situ in *Strigamia*

Preliminary experiments have shown that egg injection is feasible (Akam group, unpublished), but more work needs to be done on conditions for embryo culture before the feasibility of manipulative experiments (e.g. RNAi) can properly be assessed.

No other myriapod combines these advantages. Centipedes of the genus *Lithobius* have been used for developmental work. These have the advantage that they can be bred in the lab (with limited success), but the embryos are much harder to handle than those of *Strigamia*, and the genome size is substantially larger (about 2pg). The only millipede that has recently been used for developmental work is *Glomeris*, and this cannot be bred permanently in culture. The genome size of *Glomeris* is not known.

The community of scientists currently working on *S. maritima*, includes the groups of Michael Akam (Cambridge, UK – segmentation and early pattern formation),Wallace Arthur (National University of Ireland, Galway – geographic variation in segment number), Angelika Stollewerk (Mainz, Germany; neurogenesis) and Alessandro Minelli (Padua, Italy; segment specification, systematics and evolution of centipedes). This number will grow as members of the Akam and Arthur groups set up independent research labs. Two additional laboratories are planning projects in collaboration with the Akam group, to study characters convergently acquired by myriapods and insects – Michalis Averof (Crete, Greece – tracheal development) and Helen Skaer (Cambridge, UK – Malpighian tubule development).

Current genomic resources include established protocols for generating high quality DNA from adult animals and embryonic RNA from eggs. There is a cDNA library made from mixed embryonic stages (average insert size 0.8 Kb, number of primary clones 1.2 X 10^6). This library has been used successfully for screening genes, and for a preliminary, small EST screen. Recent published work on *S. maritima* includes biogeographic work, morphological and molecular analyses of the segmentation process, a description of trunk *Hox* gene expression, and a description of neural development. Ongoing and planned projects include a description of early cleavage and blastoderm formation, an analysis of gastrulation and mesoderm formation, and a study of head formation and differentiation, as well as the projects otherwise mentioned above.

***Limulus polyphemus* (Horseshoe crab, Chelicerate, 270 Mb genome)**

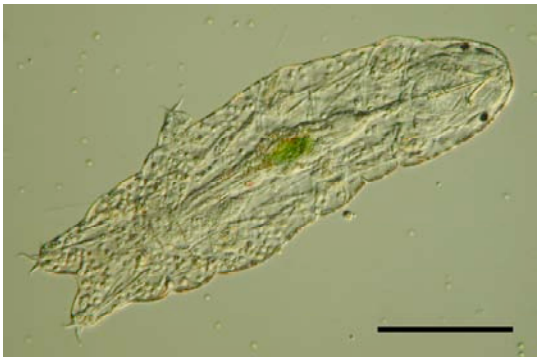
Limulus adult (from http://tolweb.org/Limulus_polyphemus/14737)

The horseshoe crab, *Limulus polyphemus*, has been an experimental system for over a century, and has been intensively studied because of the remarkable properties of its hemolymph. Within the chelicerates, it is thought to retain many ancestral characteristics in terms of morphology and development; indeed it is considered a sort of "living fossil". While not raised from generation-to-generation in the lab, it is readily available all along the East Coast of the USA, and within its breeding season, mass quantities of eggs are commercially available (from Woods Hole Marine Biological Labs for example). Fertilized embryos can be stored at 4 degrees C, which halts their development, but can be restarted again anytime within the next 12 months by simply warming them up.

Genome sequencing has been approved for two chelicerates, the deer tick *Ixodes scapularis*, and the spider mite *Tetranychus urticae*. The former was chosen for its connection to Lyme disease and the latter for its importance to agriculture. While both have several strong attributes, they have extremely small embryos and other limitations to their usefulness as embryological systems. They are also very derived in their body morphology, and preliminary studies in the spider mite suggest that its development is quite different from more typical chelicerates. The extremely small genome size (70 Mb) of the spider mite also suggests that its genome may be very derived, and thus not representative of chelicertates in general. Two spiders, *Achaeranea tepidariorum* and *Cupiennus salei* have also been used in developmental studies, but both have considerably larger genomes than *Limulus* (1.4 and 4 Gb respectively). RNAi methods have been used with some limited success in *Cupiennus*, but these spiders are relatively difficult and expensive to rear.

For these reasons, *Limulus* is the species of choice for this proposal and would be a useful addition within the chelicerates. In addition, its embryos are amenable to standard procedures for examining the patterns of gene expression and while the eggs are not transparent, the embryos develop on the surface of the yolk and can be readily imaged throughout development.

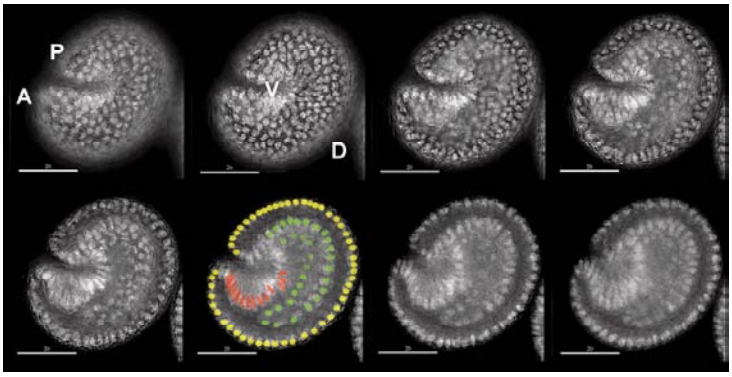
***Hypsibius dujardini* (Tardigrade; 70 Mb genome)**



Adult *Hypsibius dujardini*

Tardigrades (also known as water bears or moss piglets; phylum Tardigrada) are a widely distributed component of meiofaunal assemblages in marine sediment and fresh-water/water-film ecosystems worldwide. While only ~750 species (mostly terrestrial) have been described, the true total species level diversity is likely to be much greater, particularly in the marine groups.

Tardigrades are renowned for their abilities to withstand extreme environmental conditions, and are well-studied models for anhydrobiosis and other phenotypes. In terms of genetics and development, after a few key publications in the 1920's and 30's, tardigrades were little studied until a freshwater, herbivorous species, *Hypsibius dujardini*, was adapted to lab culture by the Blaxter lab (Edinburgh, UK) in the late 1990's, and is now established in several others such as the Goldstein (NC, USA) and Aboobaker (Nottingham, UK) labs. This species is parthenogenetic, and has a simple direct lifecycle of about ten days from egg to reproductively mature adult. Easy access to materials has led to advances in the understanding of early development, life cycle characteristics, adult anatomy, and molecular genetics. Currently there are 5000 ESTs and 1000 GSS in EMBL/GenBank. Antibody staining techniques have been developed, and the developmental lineage mapped using time-lapse video microscopy.



Early *Hypsibius dujardini* embryogenesis

Tardigrades are segmented, with four pairs of post-cephalic appendages (lobopod limbs ending in sclerotised claws) and four post-cephalic ventral ganglia. The epidermis is covered by a flexible, proteinaceous cuticle, which is ecdysed during growth (5 stages pre-adult in *H. dujardini*), and also periodically during adult life (every ~4 days in *H. dujardini*). They have a through gut that has at its anterior end a complex pharyngeal structure including a buccal tube with a pair of piercing stylets, and a triradiate myoepithelial pump. Tardigrades can thus be

thought of as being morphologically intermediate between the segmented, arthropodised (but pharynx-less) Arthropoda, and those Ecdysozoa that have a triradiate sucking pharynx (nematodes, onychophorans, priapulids). Their development is direct, with simple holoblastic cleavage, and thus resembles that of the model nematode *Caenorhabditis elegans* more than that of the fly *Drosophila melanogaster* and other arthropods. Identifying the genetic regulatory circuits underpinning early developmental pattern formation, pharyngeal development and limb outgrowth will inform models of ecdysozoan evolution. One system of particular interest is that controlling ecdysis. While *C. elegans* does not have 'canonical' ecdysone receptor genes, and controls moulting in a very different manner than does *D. melanogaster*, other nematodes do have EcR-like genes, and an EcR homologue (and other components of the presumptive tardigrade moulting control system) have been isolated from *H. dujardini*. Complete genome sequence will productively inform these analyses.

The *H. dujardini* genome size estimated from densitometric staining is 70 Mb, and current genome sequence data suggests the existence of densely packed genes with small introns. While the presence of such a small genome size could again indicate some derived characteristics in this species, the small size instead seems to be typical for Tardigrates, so will probably still be representative of this group of ecdysozoans. The repetitive DNA content of the genome, estimated from GSS data, is ~17%. Comparison of *H. dujardini* EST clusters to fly and nematode genomes reveals a fascinating pattern of conservation with some genes having homologues only in fly, and others only in nematode transcriptomes. The sequence divergence between *H. dujardini* and arthropods or nematodes does not preclude meaningful comparison. A tardigrade genome sequence is thus an essential component of efforts to reconstruct an ancestral ecdysozoan genome, and to illuminate the particular evolutionary trajectories of the other ecdysozoan phyla.

Priapulus caudatus **(Priapulid worm, 500 Mb genome*)***

Adult *Priapulus caudatus*

The priapulid worms are in a pivotal phylogenetic position at the base of the Ecdysozoa and are the ideal outgroup to both principal model invertebrate groups – the arthropods and the nematodes. The priapulids' primitive bodyplan, which has changed little in half a billion years, and evidence for particularly conservative genomic evolution are of particular significance in this regard. Their phylogenetic position will also make them a useful intermediate for comparative studies (genomic as well as developmental and morphological) between the model invertebrates and vertebrates, including humans.

The priapulid worms are marine predators that differ little, morphologically, from their fossil ancestors found in Early Cambrian deposits (>500 million years ago). Phylogenetically they are allied to the Ecdysozoa which also includes the arthropods and nematode worms. They are direct developers with radially cleaving embryos and their adult bodyplan – large size, possibly primitively segmented, with a triradiate pharynx and terminal mouth and anus – has been interpreted as being typical of the ancestral ecdysozoan. *Priapulus caudatus* breeds annually and an individual female typically produces hundreds of thousands of embryos. *In vitro* fertilization is routine making embryological studies with staged embryos relatively straightforward (G. Budd, Uppsala) and work is in progress on developing *in situ* hybridization techniques.

Long term objectives: 1) To provide genome sequence data (gene linkage, protein coding and potential regulatory regions) for interpolation between the characterized genomes of model invertebrates(including *Drosophila*, *Anopheles*, *Daphnia*, *Caenorhabditis*) and vertebrates with implications for understanding of the human genome. 2) To use the genome sequence of a priapulid worm to act as a comparator for evolutionary studies of the Ecdysozoa, especially of the arthropods and nematodes. Comparative developmental studies are more advanced in the arthropods and nematodes than in any other group. In particular, due to their phylogenetic position, priapulids can be used as the baseline for understanding the origins and evolution of many important arthropod novelties such as segmentation, appendages, cephalisation and body regionalization (tagmatization).

A normalised library has been made from mixed embryonic stages (G.Budd Uppsala) and a preliminary sequencing analysis of several hundred clones is in progress (M.Telford, UCL); an arrayed BAC library is also in production (D. Ferrier, Oxford). Preliminary analysis of approximately 400 clones from an older cDNA library shows that the genome of *Priapulus caudatus* has a tendency for slow change compared to other basal ecdysozoans including nematodes and nematomorphs (Webster et al. Evol. Dev. in press.). The fully sequenced mitochondrial genome of *P. caudatus* reinforces this interpretation and its gene order is identical

but for a single inversion to that seen in most arthropods. This genomic conservatism holds great promise for comparative genomic studies, as differences from sister-groups such as the arthropods have a greater chance of being conserved, informative, primitive characters. This contrasts with the nematodes where many differences from the arthropods are simply nematode idiosyncrasies and tell us little about the process of ecdysozoan evolution.

Interests of Proposers:
Phylogeny (Max Telford. University College London, UK), Comparative Genomics/Bioinformatics (Richard Copley. Wellcome Trust Centre for Human Genetics, Univ. Oxford, UK and Max Telford. University College London, UK), Embryology and Palaeontology (Graham Budd, Univ. Uppsala, Sweden), Evolution of Homeobox genes (David Ferrier, Dept Zoology, Univ. Oxford, UK), Comparisons with arthropods (Nipam Patel, Univ. of Calif., Berkeley, USA, Max Telford. University College London, UK, and Graham Budd, Univ. Uppsala, Sweden)

References

Abzhanov A, and TC Kaufman. 1999. Novel regulation of the homeotic gene Scr associated with a crustacean leg-to-maxilliped appendage transformation.*Development*. 126:1121-8.

Akam, M. 2000 Arthropods: Developmental diversity within a (super) phylum. *Proc. Natl. Acad. Sci. USA* 97: 4438-4441.

Arthur W, Chipman AD. 2005. How does arthropod segment number evolve? – some clues from centipedes. *Evol Dev 7:600-607*

Arthur W, Chipman AD. 2005. The centipede *Strigamia maritima*: what it can tell us about the development and evolution of segmentation. *Bioessays* 27: 653-60

Averof, M. and N. H. Patel 1997. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* 388:682-686.

Averof M and Cohen SM. 1997. Evolutionary origin of insect wings from ancestral gills. *Nature* 385: 627-630.

Brena C, Chipman AD, Minelli A, Akam M. 2006. The expression of trunk Hox genes inthe centipede *Strigamia maritima*: Sense and antisense transcripts. *Evol. Dev*. 8:252265

Browne, WE. 2003. The embryonic development of *Parhyale hawaiensis*. Ph.D. Thesis. The University of Chicago. Committee on Developmental Biology.

Browne, W. E., Price, A. L., Gerberding, M., and Patel, N. H. 2005. Stages of embryonicdevelopment in the amphipod crustacean, *Parhyale hawaiensis*. Genesis *42*, 124 149.

Carroll, SB., JK Grenier, and SD. Weatherbee. 2001 From DNA to Diversity. London,England: Blackwell Science. 214 pp.

Chipman AD, Arthur W, Akam M. 2004. A double segment periodicity underliessegment generation in centipede development. *Curr. Biol*. 14: 1250-5

Chipman AD, Arthur W, Akam M. 2004. Early development and segment formation inthe centipede *Strigamia maritima* (Geophilomorpha). *Evol. Devel*. 6: 78-89

Chipman AD, Stollewerk A. 2006. Specification of neural precursor identity in thegeophilomorph centipede *Strigamia maritima*. *Dev. Biol*. 290: 337-50

Colbourne, J. K., and Hebert, P. D. 1996. The systematics of North American Daphnia(Crustacea: Anomopoda): a molecular phylogenetic approach. *Philos Trans R Soc Lond B Biol Sci* 351:349-360.

Davis, G.K. and N.H. Patel 2002. Short, long and beyond: molecular and embryologicalapproaches to insect segmentation. *Annual Reviews of Entomology* 47:669-699.

Extavour CG. 2005. The fate of isolated blastomeres with respect to germ cell formationin the amphipod crustacean *Parhyale hawaiensis*. *Developmental Biology* 277:387

402. Fusco G. 2005. Trunk segment numbers and sequential segmentation in myriapods. *Evol. Dev. 7: 608-617*

Gerberding M, Browne WE, Patel NH. 2002. Cell lineage analysis of the amphipodcrustacean *Parhyale hawaiensis* reveals an early restriction of cell fates.*Development* 129:5789-5801.

Kettle C, Arthur W. 2000. Latitudinal cline in segment number in an arthropod species,*Strigamia maritima*. *Proc. R. Soc. Lond B* 267: 1393-7

Kettle C, Johnstone J, Jowett T, Arthur H, Arthur W. 2003. The pattern of segmentformation, as revealed by *engrailed* expression, in a centipede with a variablenumber of segments. *Evol. Dev.* 5: 198-207

Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, Martindale MQ, Holstein TW. 2005. Unexpected complexity of the Wnt gene family in a sea anemone. *Nature*. 433(7022):156-60.

Little, T. J., Colbourne, J. K., and Crease, T. J. 2004. Molecular evolution of daphniaimmunity genes: polymorphism in a gram-negative binding protein gene and analpha-2-macroglobulin gene. *J Mol Evol* 59:498-506.

Little, T. J., O'Connor, B., Colegrave, N., Watt, K., and Read, A. F. 2003. Maternaltransfer of strain-specific immunity in an invertebrate. *Curr Biol* 13:489-492.

Margulies, E. H., Vinson, J. P., Miller, W., Jaffe, D. B., Lindblad-Toh, K., Chang, J. L.,Green, E. D., Lander, E. S., Mullikin, J. C., and Clamp, M. 2005. An initial strategyfor the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing. *Proc Natl Acad Sci U S A* 102:4795-4800.

Patel, N.H., T.B. Kornberg, and C.S. Goodman 1989. Expression of *engrailed* duringsegmentation in grasshopper and crayfish. *Development* 107: 201-212.

Patel, N. H., Ball, E. E., and Goodman, C. S. 1992. Changing role of even-skipped duringthe evolution of insect pattern formation. *Nature* 357:339-342.

Pavlopoulos, A., and Averof, M. 2005. Establishing genetic transformation forcomparative developmental studies in the crustacean Parhyale hawaiensis. *Proc Natl Acad Sci U S A* 102*:* 7888-7893.

Peel AD, Chipman AD, Akam M. 2005. Arthropod segmentation: Beyond the Drosophilaparadigm. *Nat. Rev. Genet*. 6: 905-16

Price, A.L., and N.H. Patel (2006) Investigating divergent mechanisms of mesoderm development in arthropods: the expression of *Ph-twist* and *Ph-mef2* in *Parhyale hawaiesis*. *Molecular and Developmental Evolution*. JEZ PartB: DOI: 10.1002/jez.b.21135

Sagawa, K., Yamagata, H., and Shiga, Y. 2005. Exploring embryonic germ linedevelopment in the water flea, Daphnia magna, by zinc-finger-containing VASA as a marker. *Gene Expr Patterns* 5:669-678.

Scholtz G, Patel NH, Dohle W. 1994. Serially homologous engrailed stripes aregenerated via different cell lineages in the germ band of amphipod crustaceans(Malacostraca, Peracarida). *International Journal of Developmental Biology* 38:471
478.

Shiga, Y., Yasumoto, R., Yamagata, H., and Hayashi, S. 2002. Evolving role ofAntennapedia protein in arthropod limb patterning. *Development* 129: 3555-3561.

Stollewerk A, Chipman AD. 2006. Neurogenesis in myriapods and chelicerates and itsimportance for understanding arthropod relationships. *Int. Comp. Biol.* in press

Watanabe, H., Tatarazako, N., Oda, S., Nishide, H., Uchiyama, I., Morita, M., and Iguchi, T. 2005. Analysis of expressed sequence tags of the water flea Daphnia magna.*Genome* 48: 606-609.