

Angling for our ancestors: fishing with comparative genomics in deep waters

Eric Gaidos
Associate Professor of Geobiology
School of Ocean and Earth Sciences and Technology and
NASA Astrobiology Institute

It is in our nature to want to know our ancestors. Where did they come from, what were they like and how did events in their lives shape our present existence? The tools of molecular biology and genomics now give us the ability to query our deep evolutionary ancestry: not tens of generations back into history, but tens of millions of generations back into the abyss of geologic time. Charles Darwin first provided the mechanistic framework of evolution by natural selection within which a similarity between two species could be understood as a manifestation of inheritance from a common ancestor that possessed the common trait. More than a century later Emile Zuckerkandl and Linus Pauling, studying the amino acid structure of primate hemoglobins, realized that biological molecules, whose patterns are transmitted by genetic inheritance from one generation to the next, also contain a record of shared ancestry between the organisms that contain them [1]. Because all animal morphology and physiology (with the exception of certain aspects of the immune system) has a heritable molecular basis, and because molecular patterns – i.e., sequences of peptides or nucleotides - are more tractable to quantitative analysis, the study of molecular evolution has become an increasingly prominent aspect of the field of evolutionary biology. This “molecular evolution revolution” has occurred because of two technological advances: the advent of cost-effective high-throughput DNA sequencing, which has vastly accelerated both the rate at which new taxa and new molecular data for each taxon can be investigated, and the continued exponentiation in computational power, which has enabled the proliferation of fast computation tools to analyze such data. Complete or nearly-complete sequencing of entire genomes and their analysis has now become “routine”. Traits that can be compared now include presence or absence of genes, gene order (synteny), and chromosome structure.

Comparative evolutionary genomics across a range of taxonomic and evolutionary scales now offers a breathtaking opportunity for new insights into human origins and the human condition. One fundamental question is the origin of animal life itself (the Metazoa) sometime in the late Precambrian ca. 600 million years ago. By comparing the genes and genomes of extant organisms, properties of the last common ancestor of all animal life (or groups of animals) can be inferred, e.g. axial symmetry, sensory and neural systems, and developmental pathways. However, evolutionary reconstructions or phylogenetic analyses using the gene encoding the small subunit of the ribosomal RNA molecule – a favorite in deep evolutionary analyses by virtue of its high degree of conservation - have by and large failed to unambiguously resolve the relationships between the deepest (oldest) branches in the

animal tree of life [2]. A many-gene or “phylogenomic” approach has also produced unresolved polytomies [3]. Moreover, many studies have demonstrated the importance of including as many phylogenetically informative taxa as possible [4].



Jillian Ward, a doctoral student in the Department of Oceanography, is investigating the evolutionary genomics of the enigmatic organism, *Trichoplax adhaerens*. *Trichoplax* is an enigmatic mm-sized marine invertebrate with only a few cell types, an extremely simple body plan, and one of the smallest known animal genomes (50 million base pairs) [5]. It is found world-wide in tropical and

sub-tropical waters and is recovered on glass surfaces immersed in the water column until a biofilm develops (Figure 1). It is the only described species in the phylum Placozoa (for comparison the phylum Vertebrata contains nearly 60,000 described species). The location of placozoans in the tree of animal life is unknown, but a recent report places the phylum at the base of the extant animal tree of life based on its mitochondrial genome sequence [6]. If this location is correct it would mean that the simple structure and genome of this organism is more likely to be primitive, rather than derived (from a more complex ancestor). However, mitochondrial genes are subject to various potential artifacts. A more robust approach would use a large number of nuclear genes, and include information (in the form of a model) on how each gene has evolved in different parts of the tree. Gene evolution may be unusual in *Trichoplax* because of its global dispersal, asexual modes of reproduction, possible genome reduction, and the low GC content of its genome. Jillian’s approach to this problem is based on two recent advances in our understanding of placozoans: First, the Joint Genome Institute is sequencing the complete genome of one cultivated strain of *Trichoplax*. Second, molecular haplotyping

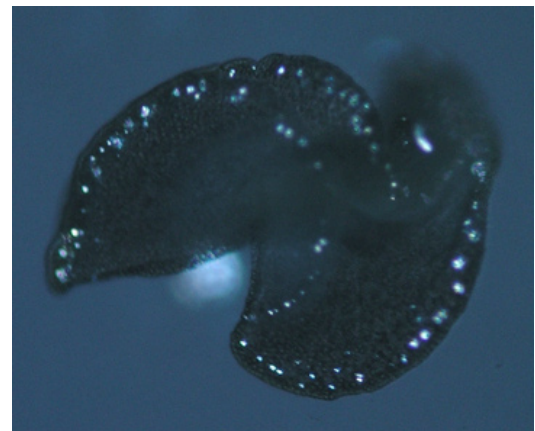


Figure 1. *Trichoplax adhaerens* seen in polarized light on a glass slide. The bright spots are birefringent granules, minerals whose formation is under strict biological control by the organism. This individual is about one mm across.

of cultivated isolates worldwide has revealed the presence of at least 8 different “species” [7]. Jillian is isolating different strains from Hawaii and elsewhere, and constructing libraries of expressed sequence tags (ESTs). EST libraries are sequences of expressed genes that have been advocated as an efficient means of generating data on many taxa for phylogenomic analysis [8]. The complementary DNA (cDNA) pool that is an intermediate product of EST library construction can also be probed for specific genes of interest using degenerate primers in a polymerase chain reaction (PCR). A data “matrix” of many gene sequences from multiple strains will be used to develop a model of gene evolution in the phylum, which can in turn be used in the phylogenetic analysis. Using these data, Jillian expects to also say much more about the biogeography and population genetics of Trichoplax. Ms. Ward is the recipient of a fellowship from the University of Hawaii High Performance Computing Center and support from the Director’s Discretionary Fund of the NASA Astrobiology Institute.



Dr. Gayle Philip, a postdoctoral fellow with the UH Lead Team of the NASA Astrobiology Institute, is examining existing molecular data for the major animal groups. Most analyses have analyzed single genes, or concatenated multiple genes into one single sequence from which a phylogenetic tree is inferred. Gayle will use an approach known as “supertree”

analysis, in which individual genes are used to construct trees, and then those trees are combined into a consensus “supertree” [9]. An advantage of this approach is that individual evolutionary models can be used for specific genes. Furthermore, the true complexities of different genes’ support for different parts of the tree can be represented as a graphical network [10].

This approach has been used in a previous analysis by Dr. Philip and her co-workers that was based on complete genomes of

representative organisms and questions the widely-held view that animals and fungi are more closely related to each other than to plants (Figure 2) [11]. While construction of a single, well-supported and unambiguous tree of animal life may not be possible, Gayle hopes to better understand which phylogenetic signals are associated with which genetic loci, and why.

The DNA sequencing and analysis capacity developed for the human genome project has now been turned to a large number of mammals and non-mammals of interest to medicine, conservation biology or evolutionary biology. New techniques (such as pyrosequencing) promise to accelerate the accumulation of data to even faster rates. The field is now limited only by bright ideas and bright minds to pursue them. Genes and species are the evolutionary warp and weft from which the tapestry of all life is woven. My junior colleagues and I are using evolutionary comparative genomics to understand the processes that weave the raw biological fabric upon which natural selection acts, and document the imprint of that selection on the living beings that currently (now often perilously) grace our planet.

References

1. Zuckerkandl, E., R.T. Jones, and L. Pauling 1960. “A comparison of animal hemoglobins by tryptic peptide pattern analysis” Proc. Natl. Acad. Sci. USA 46, 1349-1360.
2. Wallberg, A., M. Thollesson, J.S. Farris, and U. Jondelius 2004. “The phylogenetic position of the comb jellies (Ctenophora) and the importance of taxonomic sampling” Cladistics 20, 558-578.
3. Rokas, A., D. Kruger, and S.B. Carroll 2005. “Animal evolution and the molecular signature of radiations compressed in time” Science 1933-1938.
4. Hedtke, S.M., T.M. Townsend, and D.M. Hillis 2006. “Resolution of phylogenetic conflict in large data sets by increased taxon sampling” Syst. Biol. 55, 522-529.
5. Grell, K.G. and A. Ruthmann, “Placozoa”, in Microscopic Anatomy of Invertebrates, F. Harrison and E. Ruppert, Editors. 1991, Wiley-Liss: Hoboken, NJ.
6. Dellaporta, S.L., A. Xu, S. Sagasser, W. Jakob, M.A. Moreno, L.W. Buss, and B. Schierwater 2006. “Mitochondrial genome of Trichoplax adhaerens supports Placozoa as the basal lower metazoan phylum” Proceedings of the National Academy of Sciences of the USA 103, 8751-8756.
7. Pearse, J.S. and O. Voigt 2007. “Field biology of placozoans (Trichoplax): distribution, diversity, biotic interactions” Integrative and Comparative Biology 47.
8. Philippe, H. and M.J. Telford 2006. “Large-scale sequencing and the new animal phylogeny” Trends Ecol. Evol. 21, 614-620.
9. Sanderson, M.J., M.J. Purvis, and C. Henze 1998. “Phylogenetic supertrees: assembling the trees of life” Trends Ecol. Evol. 13, 105-109.
10. Huson, D.H. and D. Bryant 2006. “Applications of phylogenetic networks in evolutionary studies” Mol. Biol. Evol. 23, 254-267.
11. Philip, G.K., C.J. Creevey, and J.O. McInerney 2005. “The Opisthokonta and Ecdysozoa may not be clades: stronger support for the grouping of plants and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa” Mol. Biol. Evol. 22, 1175-1184.

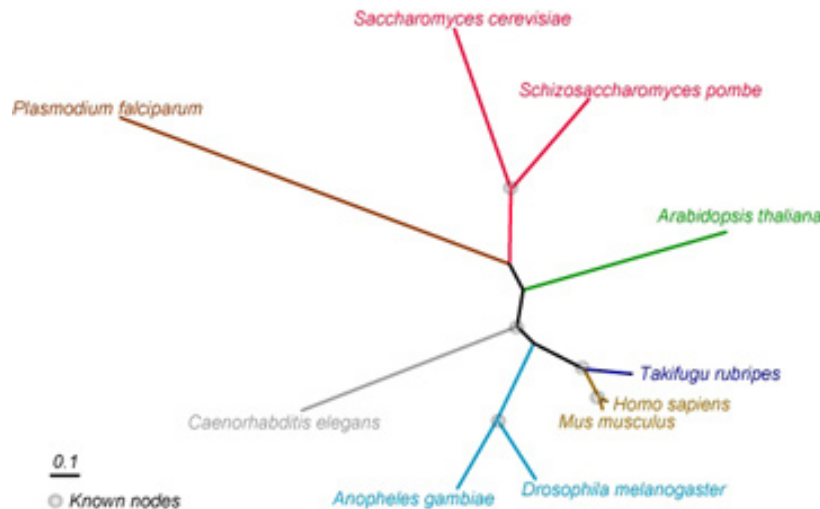


Figure 2. A phylogenetic tree constructed from 10 complete eukaryotic genomes and both supertree and concatenated sequence methods [11]. Known nodes are those supporting grouping of the mammals (H. sapiens and M. musculus), vertebrates (mammals and Takifugu), arthropods (Drosophila and Anopheles), animals (arthropods, vertebrates and C. elegans) and fungi (S. cerevisiae and S. pombe).