Building the Dipteran Tree: Co-operative Research in Phylogenetics and Bioinformatics of True Flies (Insecta: Diptera)

With an estimated 150,000 described species, the insect order Diptera (true flies), is one of the most diverse branches on the tree of life. The challenge of generating a robust phylogenetic framework for this impressive diversity has attracted the attention of some of the best systematists (e.g., Willi Hennig). Nonetheless, despite all efforts, the relationships between the main clades of the dipteran tree are still unknown or controversial. This is particularly unfortunate considering that the Diptera contain critical species for diverse fields in biology and medicine (e.g., *Drosophila, Anopheles*) and could be a model taxon for comparative evolutionary studies in genomics, development, neurobiology, and behavior. However, these would ultimately depend on having a reliable phylogenetic tree. We will reconstruct this tree in detail during our project based on an unprecedented large sample of morphological and molecular data. We will also integrate information from the exceptional fossil record of Diptera, from emerging genomic and developmental databases, and from new and existing specimen data and published phylogenetic trees.

Intellectual Merit. We will assemble the fly tree of life through a collaborative effort among a large, international team with morphological and molecular systematic expertise and complimentary expertise in the development of bioinformatics tools for natural history information. The task will be mastered in a three-tier approach. In the first tier, we will reconstruct a backbone tree based on the first comprehensive morphological character matrix and molecular data for 42 species representing all major dipteran lineages and incorporating important model organisms. The second-tier analysis will use a more comprehensive taxon sample comprising at least one species for all dipteran families, more morphological characters and a reduced amount of molecular data. In this tier we will also incorporate characters from well-preserved fossil species. The third tier will rely on the backbone trees reconstructed in the first two tiers for building a supertree for 1500-2000 species. Using these data, we will explore three approaches to reconstructing the phylogenies of megadiverse taxa. First, we will investigate support for a stable higher-level framework for fly phylogeny from a very large dataset collected for a small taxon sample (42 taxa, first-tier). The performance of this data set will then be compared with the one of the second-tier matrix built for a much larger taxon sample (240 taxa) but containing a smaller amount of character data. Due to the large number of terminals, the second-tier matrix will be used as a model for exploring the efficiency of different search strategies and parallel computing in large-scale-phylogenetics. In the third tier we will test different techniques for constructing 'supertrees' in an attempt to judge whether supertrees are a viable alternative to more traditional techniques for reconstructing the phylogeny of megadiverse taxa.

Broader Impacts. A robust phylogenetic hypothesis for Diptera will be extremely valuable because the group is a major component lineage of the astounding radiation of holometabolan insects. Molecular and morphological character systems and genes explored in this project should be useful in resolving similar-aged radiations in other insects. Ultimately a comprehensive new phylogeny for all of Diptera will provide a valuable framework for testing evolutionary hypotheses critical in comparative studies of dipteran development, behavior, genomics, and neurobiology. In order to efficiently communicate our results, we will concurrently build the 'Diptera Web.' This website will be the central hub for information on Diptera and present our trees, contain important electronic catalogs (e.g. World Catalog for Diptera), databases for Diptera species in culture, a comprehensive anatomy atlas, interactive keys, and links to collection homepages, FlyBase, and GenBank. An existing specimen database program will be modified to link DNA sequences with digital images and supplementary information on the specimen from which the DNA was extracted.