H.J., and P.R.P. Hofmann 2005, Dros. Inf. Serv. 88: 97-101; Tidon, R., and F.M. Sene 1988, Dros. Inf. Serv. 67: 90; Tidon, R., D.F. Leite, and B.F.D. Leão 2003, Biol. Cons. 112: 299-305; Tidon, R., 2006, Biol. J. Linn. Soc. 87: 233-247; Val, F.C., C.R. Vilela, and M.D. Marques 1981, In: *The Genetics and Biology of* Drosophila (Ashburner, M., H.L. Carson, and J.N. Thompson, jr., eds.), pp. 123-168, Academic Press, NY; Val, F.C., 1982, Pap. Av. Zool. 34: 309-347; Vilela, C.R., 1983, Rev. bras. Ent. 27: 1-114; Vilela, C.R., and G. Bächli 1990, Mitt. Schweiz. Ent. Ges. 63: 1-332; Wheeler, M.R., and H. Takada 1963, Ann. ent. Soc. Am. 56: 392-399.

Rapidly evolving genes show well-resolved but conflicting phylogenies: Evidence from *Drosophila simulans* complex.

Jeck, William R., Hung-Jui Shih, and Corbin D. Jones. University of North Carolina, Department of Biology, CB #3280, Chapel Hill, NC 27599, USA; TEL: 919-843-5162; FAX 919-962-1625; Corresponding author email: cdjones@email.unc.edu.

Introduction

The *D. simulans complex,* which includes *D. simulans, D. mauritiana,* and *D. sechellia*, is a model system for the experimental study of speciation, sexual selection, and adaptation. Remarkably, the phylogenetic relationship of these species to each other has not been adequately resolved (summarized in Kliman *et al*., 2000; Harr *et al*., 1998; Ting *et al*., 2000). Specifically, a clear molecular phylogeny has not emerged, despite obvious behavioral, morphological, and physiological differences among these species. For example, the male genitalia of the three species are readily distinguishable from each other (Ashburner, 1989). Furthermore, these are good biological species as F_1 males between these species are completely sterile. Two issues make resolution of the relationships among these species problematic. First, a number of genes do not resolve the species – alleles of one species are often nested within alleles of another. Second, even when monophylly is observed, the branching order of the species is often inconsistent among genes.

Ting *et al*. (2000) advocated “speciation” genes, those loci that contribute to divergence between species (*e.g*., those involved in gametogenesis, behavior, morphology), as ideal genes for constructing a phylogeny of closely related species. To demonstrate their point, Ting *et al.* (2000) used the rapidly evolving *OdsH* to try to resolve the *D. simulans* clade phylogeny. Their analysis of the gene tree of *OdsH* suggested that *D. sechellia* split from *D. simulans* first, followed shortly by *D. mauritiana*. This result was congruent with the overall result of Harr *et al*. (1998), but not with analyses of several other genes (Kliman *et al*., 2000). A possible explanation for the observed discrepancy between Ting *et al.* and Kliman *et al.* is that any rapidly evolving gene will resolve a gene tree, but that these gene trees may not correspond to the actual species tree.

In this study, we test the idea that the resolved tree produced by Ting *et al*. is typical of that produced by any rapidly evolving gene. We also show that the trichotomy of the *simulans* clade is unlikely to be resolved.

Materials and Methods

We queried the NCBI sequence repository for all nucleotide sequences from *D. mauritiana*, and combined with sequence data produced in our lab for a set to 114 unique genes. Whole genome sequences were taken from the AAA CAF1 releases for *D. sechellia, D. simulans* and *D. melanogaster*. Homologs to the 114 *D. mauritiana* genes were identified using BLASTn (Altshul *et al*., 1997). BLAST results within 10 kb were condensed and their bit scores combined. The region for each species having the highest bit score was then taken and considered to represent the gene of interest. All four sequences were initially aligned with Clustalw (Thompson *et al*., 1994) using all four species and then visually inspected. Alignments with anomalies or missing sequence were removed, leaving 78 alignments. These were analyzed using MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001) using the Generalized Time Reversable model and gamma distributed rate variation (GTR+G; Tavaré, 1986). We chose this general model rather than determining a specific model in order to be sure that trees and branch lengths among our genes were comparable.

To test our hypothesis that genes evolving rapidly result in better resolved phylogenies, we compared the rank correlation between each gene branch length to *D. melanogaster* from the common ancestor of the *D. simulans* clade with the posterior probability of the most credible tree using MATLAB (www.mathworks.com).

Results and Discussion

Interspecific DNA sequence data from 78 different genes did not resolve the *D. simulans-D. sechellia-D. mauritiana* trichotomy, although our data suggest that *D. simulans* and *D. mauritiana* are not the most closely related taxa (Figure 1). One explanation for this pattern is that the *D. mauritiana* and *D. sechellia* split from *D. simulans* during the same vicariance event. Alternatively, the ability to resolve the phylogeny is hampered by the large amounts of segregating genetic variation found in *D. simulans* (Begun *et al*., 2007). Ultimately, determining the phylogeny of every gene in the genome of these species via whole genome sequencing of *D. mauritiana* may be the only way to determine which of these species are more closely related to each other.

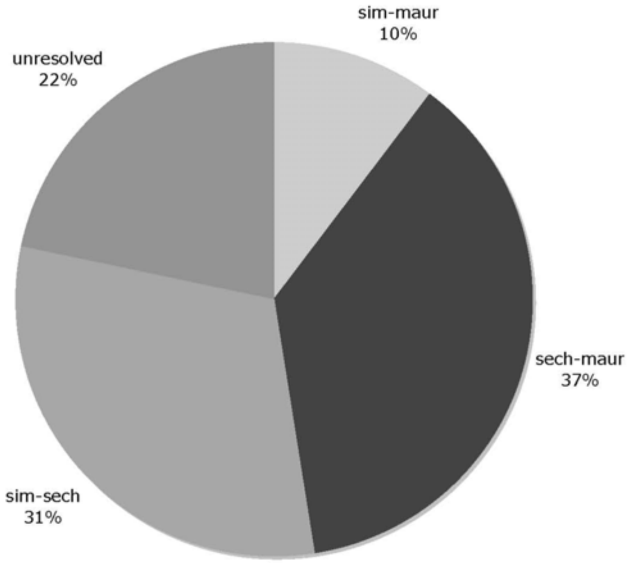

Figure 1. Sequence data from 78 different genes did not resolve the *D. simulans-D. sechellia-D. mauritiana* trichotomy. The chart shows the fraction of resolved trees indicating that that pair of species are the two most closely related among the three. In total, our data suggest that *D. simulans* and *D. mauritiana* are not the most closely related taxa.

Recently speciation genes, such as *OdsH,* were suggested to be the best indicators of species phylogeny. Indeed *OdsH* results in a well-resolved tree. *OdsH*, however, is also a rapidly evolving gene. Thus an alternative explanation is that rapidly evolving genes tend to result in better resolved gene trees, but not necessarily better species trees. We used out data to test this hypothesis. Branch length from the ancestor of the *D. simulans* clade to *D. melanogaster* was positively correlated with posterior probability of the most credible tree (Figure 2; Spearman's correlation coefficient of 0.47, *P* <

0.0001). These data suggest that rapidly evolving genes tend to resolve the clade, regardless of their role in speciation. Given that the only known "speciation gene" in the *D. simulans* group—*Odh*—is also rapidly evolving, our analysis suggests this gene's phylogenetic signal may result from its rapid rate of evolution and *not* from its role in speciation. Consistent with this view, the particular phylogenetic relationship suggested by this gene was not the most common among our data set.

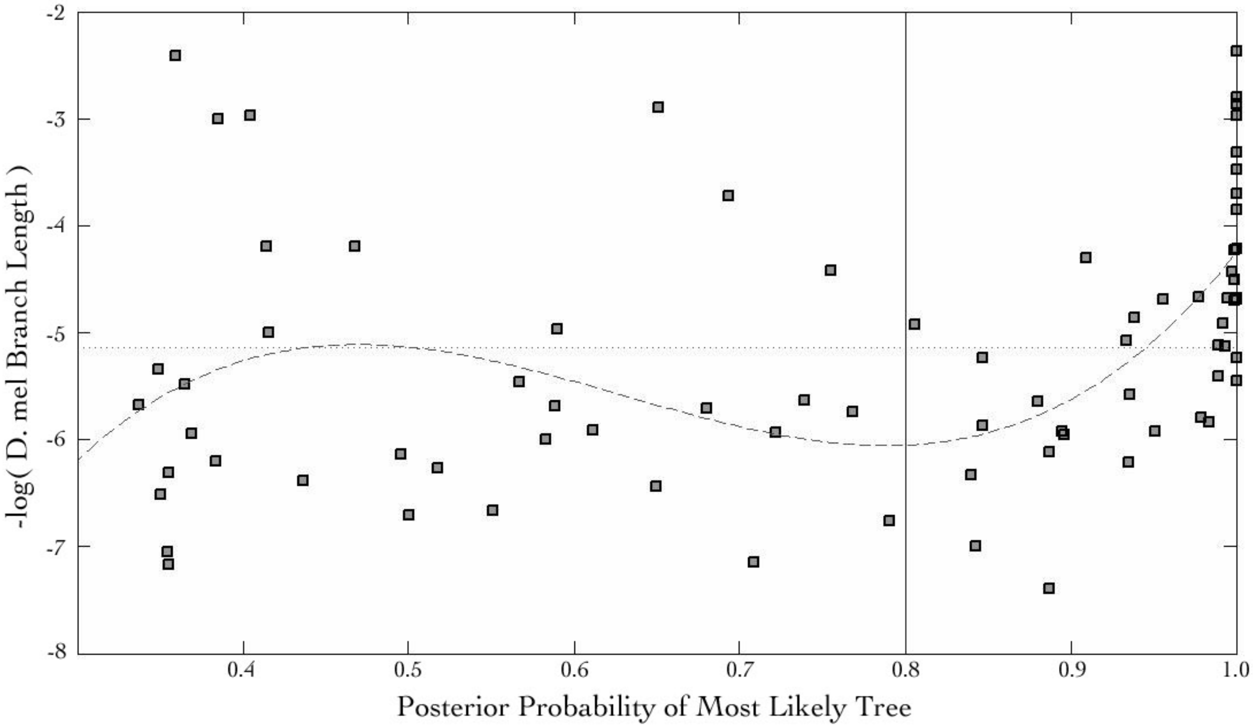

Figure 2. Rapidly evolving genes result in more credible trees. The most credible trees are positively correlated with genes that have the longest branch length from the ancestor of the *D. simulans* clade to *D. melanogaster* (Spearman's correlation coefficient of 0.47, $P < 0.0001$). A best fit polynomial is plotted to illustrate the general pattern. These data suggest that rapidly evolving genes tend to resolve a clade, regardless of their role in speciation.

Acknowledgments: This work was supported by funds from the National Science Foundation.

References: Kliman R.M., P. Andolfatto, *et al*., 2000, Genetics 156: 1913-31; Harr B., S. Weiss, *et al*., 1998, Curr. Biol. 8: 1183-6; Ting, C.T., S.C. Tsaur, *et al*., 2000, Proc. Natl. Acad. Sci. USA 97: 5313-6; Ashburner, M., 1989, Drosophila: *A Laboratory Handbook*; Altschul, S.F., T.L. Madden, *et al*., 1997, NAR 25: 3389-402; Thompson J.D., D.G. Higgins, *et al*., 1994, NAR 22: 4673-80; Ronquist, F., and J.P. Huelsenbeck 2003, Bioinformatics 19: 1572-4; Tavaré, S., 1986, Lectures on Mathematics in the Life Sciences 17: 57-86; Begun, D.J., A. K. Holloway *et al.,* 2007 PLoS Biol. 5(11): e310.