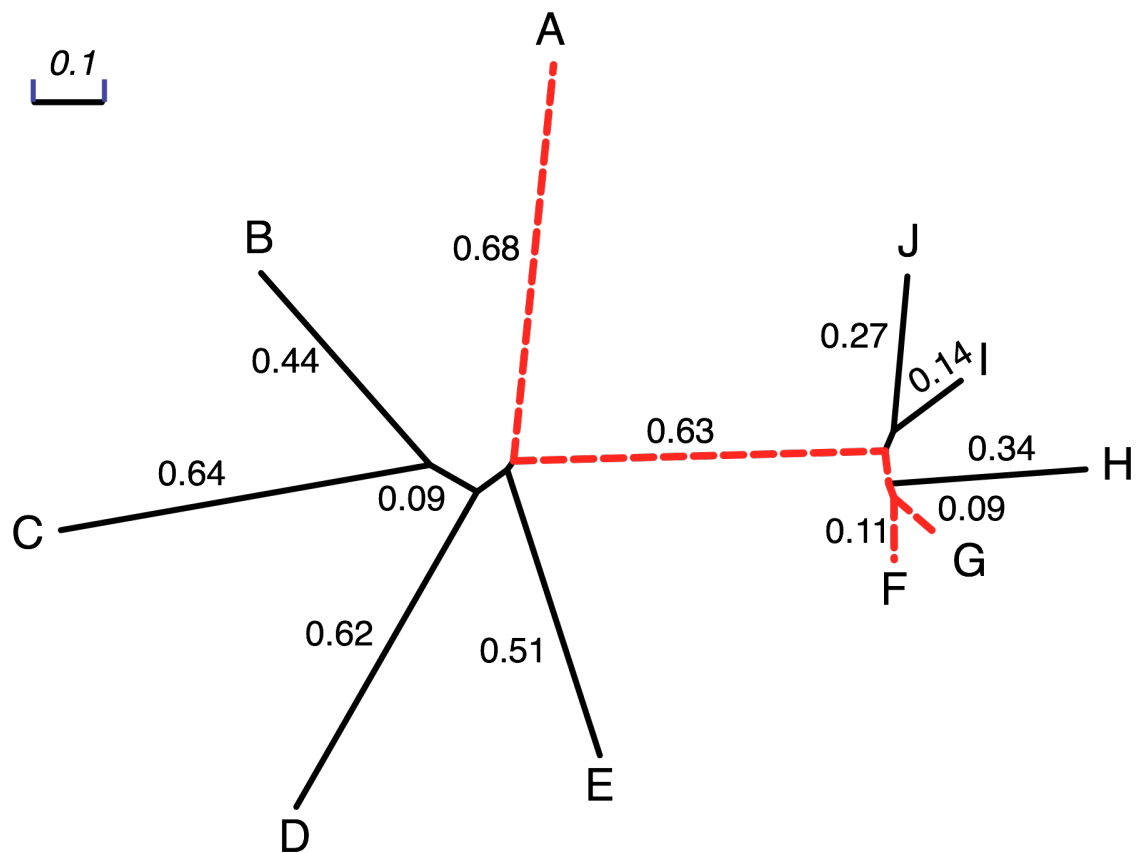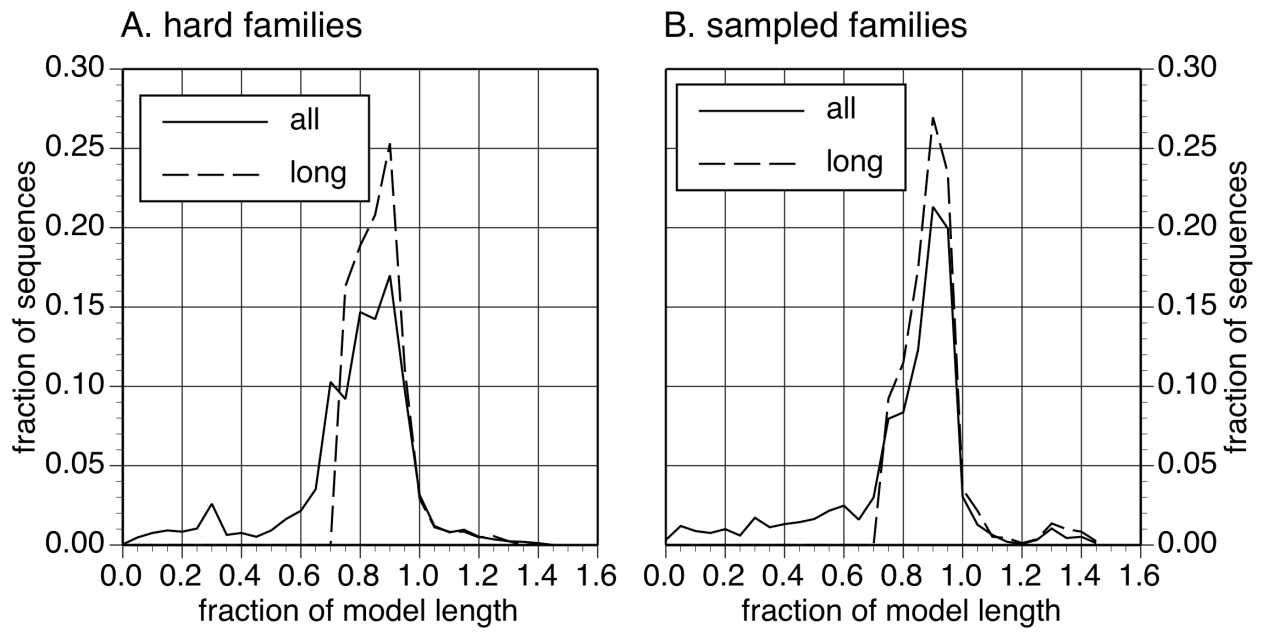**SUPPLEMENTAL FIGURES**

Supplemental Figure 1:

Measuring tree coverage. Tree coverage was measured as the weighted sum of the branch lengths of the sub-tree covered by the TPs found in any given search. This diagram shows an example of an unrooted tree for a family of 10 domains, with a topology described by 17 branches. To illustrate how tree coverage is calculated, assume that domain F was used as the query for a PSI-BLAST search, which identified domains F (self-recognition), G, and A. The "found" sub-tree (in red) is identified by traversing the tree from the query to the most distant TP hit identified by the search. Thus, the tree coverage for this search is the sum of the branch lengths of the 6 "found" branches (1.60). Separately, the tree size (i.e., the sum of all tree branch lengths) is calculated (4.76) and used to generate the weighted tree coverage for the search (0.34) so that comparisons across trees of various sizes and topologies can be made.

Distribution of homologous domain lengths. The lengths of homologous domains for the (A) 50 challenging and (B) 50 randomly sampled query families. Fifty homologous domains were sampled from each query family; each length is divided by the Pfam model length for the query family to calculate the fraction model length. Solid lines show the distribution of domain lengths in the *standard* library; dashed lines report domain lengths in the *long-domain* library.