**Supplementary information for**

**orthoSNAP: a novel tree splitting and pruning algorithm for retrieving subgroups of single-copy orthologs from gene family trees**

**Abstract**

Molecular evolution studies, such as phylogenomic studies and genome-wide surveys of positive selection, often rely on gene families of single-copy orthologs (SC-OGs). In contrast, large gene families with multiple homologs in one or more species—a phenomenon observed among several important families of genes such as transporters and transcription factors—are often ignored because identifying and retrieving SC-OGs nested within them is challenging. To address this issue and increase the number of markers used in molecular evolution studies, we developed orthoSNAP, a software that uses a phylogenetic framework to simultaneously split gene families into SC-OGs and prune species-specific inparalogs. We term SC-OGs identified by orthoSNAP SNAP-OGs because they are identified using a splitting and pruning procedure. From 46,645 orthologous groups of genes inferred using graph-based clustering of sequence similarity scores across four separate eukaryotic datasets, we identified 6,630 SNAP-OGs using orthoSNAP and 6,634 SC-OGs. Comparison of SNAP-OGs and SC-OGs revealed no differences between the two groups in their phylogenetic information content. orthoSNAP is useful for increasing the number of markers used in molecular evolution data matrices, a critical step for robustly inferring and exploring the tree of life.