

#### **Additional data file 4 - Diversity of kinesin protein domain architectures.**

Pfam and CDD searches were used to identify putative gene architectures for all 1624 putative kinesin proteins and mapped the relative taxon distribution. The domain architectures are noted adjacent to the last column using short domain names. The species are labelled across the top and are grouped into higher taxonomic units. The taxon distribution of architecture characters used for the parsimony tree analyses of the terminal branches of the eukaryotes are marked c (character) adjacent to the first column. While and characters excluded because they were not homologous are marked d(ex) (discounted and excluded) and taxon distributions that were altered with one or more taxa discounted because the architectural character is not homologous are marked d/c (discounted – amended and then counted as a character).

