Supplemental File 1: A full list of 486 kinesin-like sequences from 19 diverse organisms.

Supplemental File 2: A manually-edited alignment of the motor domains of 400 kinesin proteins from 19 diverse organisms.

All sequences included pass the e<1e-70 cut-off (see Methods). For phylogenetic inference included residues were: 1) Full 403x400 matrix: 33-62 83-100 167-186 188-202 204-206 216-238 465-485 592-604 616-630 794-808 810-817 822-869 942-963 1002-1014 1049-1065 1187-1206 1215-1216 1225-1242 1282-1363. 2) Smaller 293x400 matrix: 165-186 188-202 204-206 216-234 467-483 594-604 616-629 795-808 810-817 822-861 943-963 1003-1014 1049-1065 1189-1206 1215-1216 1225-1242 1282-1323.

Supplemental File 3: Hidden Markov models for prediction of kinesin family membership without phylogenetics.

File contains analytical HMMs for 19 kinesin groups: 14 kinesin families, 2 Kinesin-14 subfamilies, the Kinesin-4/10 'superfamily' and 2 phylum-specific kinesin groups. A quick method for using the HMMs follows:

1. Download and install HMMER2.0 (Sean Eddy, Washington University School of Medicine, USA) from: http://hmmer.wustl.edu/

Installing on Apple Mac OS-X or Linux is very straight-forward. For Microsoft Windows, it can be more tricky. See the HMMER website (http://hmmer.wustl.edu/) for more details on installation, and notes on all the other things that can be done with HMMER.

2. Download the 19 HMMs for the kinesin groups.

3. Get your sequence. HMMER can read several different file formats, including FASTA, SWISS-PROT and PIR. You can also put several sequences into the same file and they will be searched successively.

4. Open a terminal (or MS-DOS) window and go to the folder containing the HMMs and sequence(s). (If you aren't familiar with using a terminal window, this can be done with the command: cd <dir_name>).

5. The sequence(s) can be searched against the HMMs by typing:

hmmpfam <hmms_filename> <seq_filename>

To use the built-in cut offs add one of the following options:

    --cut_ga   (implements gathering cut-off for HMMs)

    --cut_nc   (implements noise cut-off for HMMs)

    --cut_tc   (implements trusted cut-off for HMMs)

We recommend the use of the gathering cut-off for most situations.

6. The output will appear in the terminal window (for a single sequence this should take only a few seconds). If no kinesin family specific HMM hit is found, then the kinesin is most probably an orphan.