# Obtaining Extremely Large and Accurate Protein Multiple Sequence Alignments from Curated Hierarchical Alignments

## Andrew F. Neuwald, Christopher J. Lanczycki, Theresa Hodges and Aron Marchler-Bauer

### **Supplementary Tables**

**Table S2**. Semantics of taxonomy annotated variables added to fasta-formatted sequences definition lines by the AddPhylum program.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | variable | type | semantics | example |
|  | *seqid* | string | sequence identifier | RPF34781.1 |
|  | *offset* | integer | start site offset | 0 |
|  | *taxid* | integer | species identifier | 2485151 |
|  | *phylum* | string | phylum | Actinobacteria |
|  | *kingdom* | character | major groupa | B |
|  | *description* | string with spaces | protein description | ATP-binding cassette… |

aThe letter designations for kingdoms are: metazoa, M; protozoa, E; plants, V; fungi, F; bacteria, B; archaea, A.