The estimation of the microsporidian last common ancestor protein set

Proportion of orthologous and lineage specific proteins (or the reduction and expansion in microsporidian genome)

We analyzed the proportion of orthologous and lineage specific proteins in 11 microsporidia species.

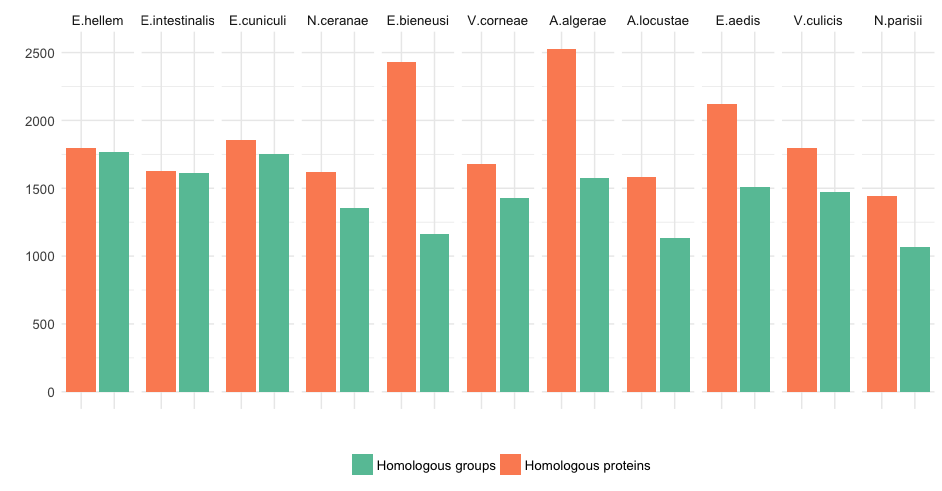


Figure 1‑1: The distribution of number of homologous proteins (orange) and number of homologous groups (green) in each microsporidia taxon.

Figure 1‑1 shows the relative number of homologous proteins and number of homologous groups in each microsporidia species. In some species, such as E.bieneusi or A.algerae, the number of homologous genes is substantially higher than the number of corresponding homologous groups. We check the number of in-paralogs for each microsporidia taxon in the homologous group. The result in Figure 1‑2 shows that there is no evidence for whole genome duplication in any species. But there are some instances where the homologous groups contain more than 10 co-orthologs for one microsporidia species showing the effect of gene dosage (Kondrashov and Koonin, 2004).

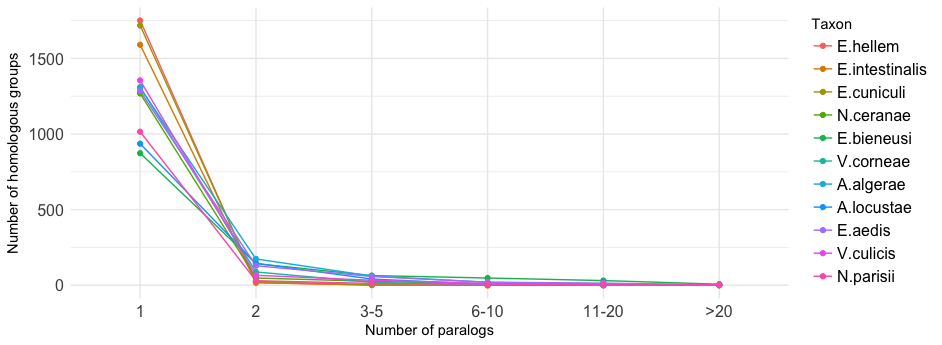


Figure 1‑2: The distribution of number of homologous groups as a function of number of in-paralogs. Colors denote different microsporidia taxa.

HamFAS:

Why did BlastKOALA and KAAS fail to annotate HamFAS-only protein?

One reason could be, that the orthology prediction approaches used by KAAS and BlastKOALA are not as sensitive as HaMStR. The second reason is due the their "secret" filter criteria to select candidate for doing annotation transfer.

We have observed, that there are in total 86 predicted KOs that are common between annotated and unannotated data sets. In which 44 cases are also supported by InParanoid. Examples:

**K00077:** sce:4314, B.subtilis:BSU15110, B.subtilis:BSU14440 (KEGG's representative sequences)

HamFAS: sce:6474 (unannotated protein) is orthologous with B.subtilis:BSU14440

Inparanoid: 2 separate OGs: (sce:4314, B.subtilis:BSU15110) and (sce:6474, B.subtilis:BSU14440)

**K00799:** ath:AT1G02930,..., ath:AT2G30870,..., sce:5364, sce:1884

HamFAS: sce:2310 - ath:AT1G02930,...

(sce:1884 has no ortholog with *A.thaliana* according to InParanoid)

**K00877:** S.pombe:4570, S.pombe:875, S.pombe:1336, sce:1877, sce:997

HamFAS: sce:487 - S.pombe:1336

Those proteins have been probably either not predicted as orthologs or discarded after filtering through KEGG annotation pipeline.