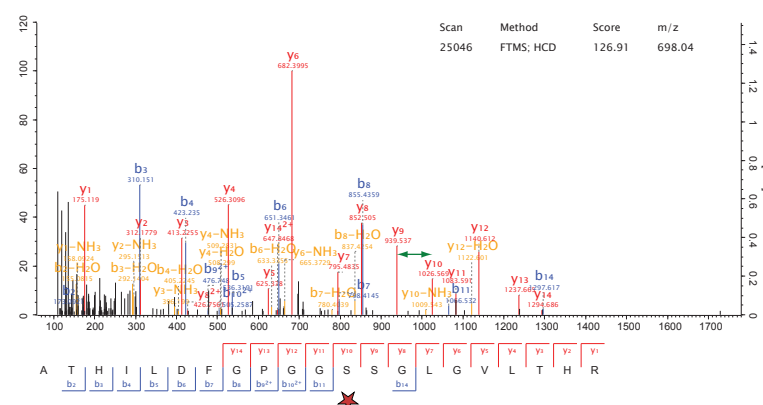
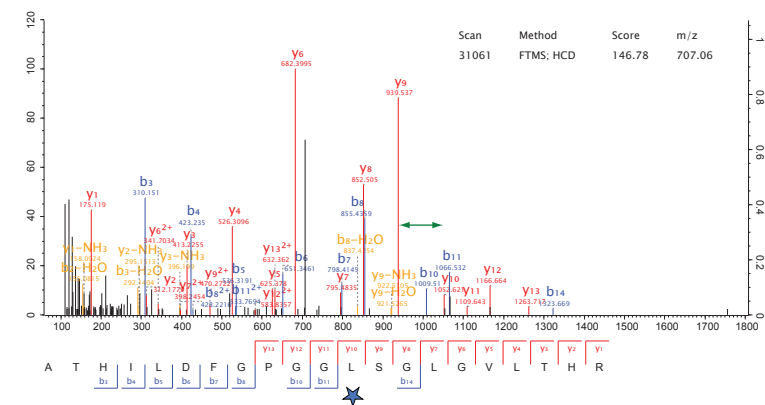
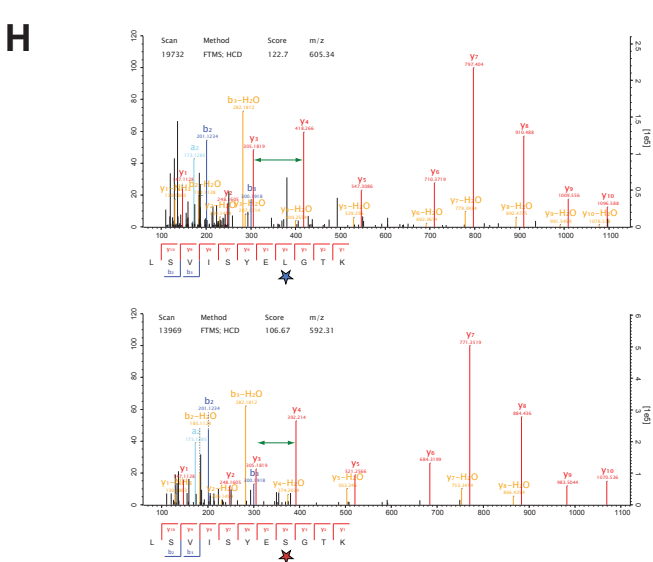
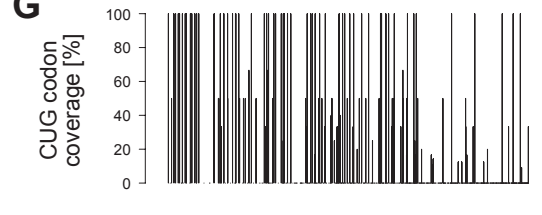
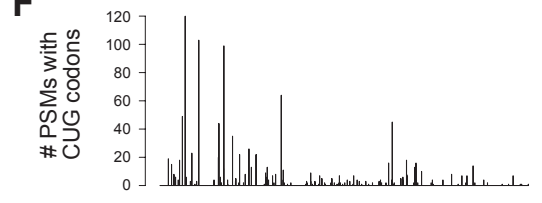
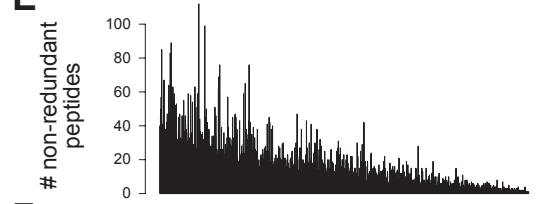
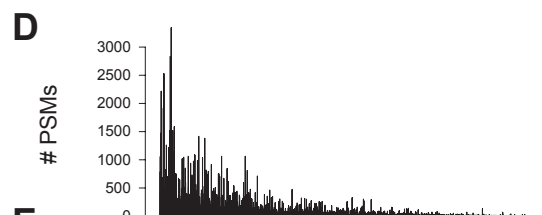
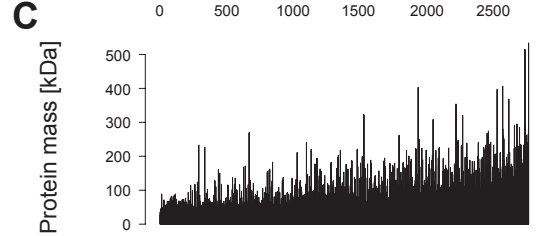
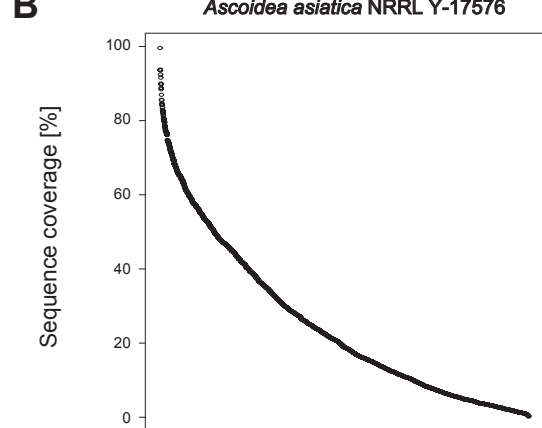
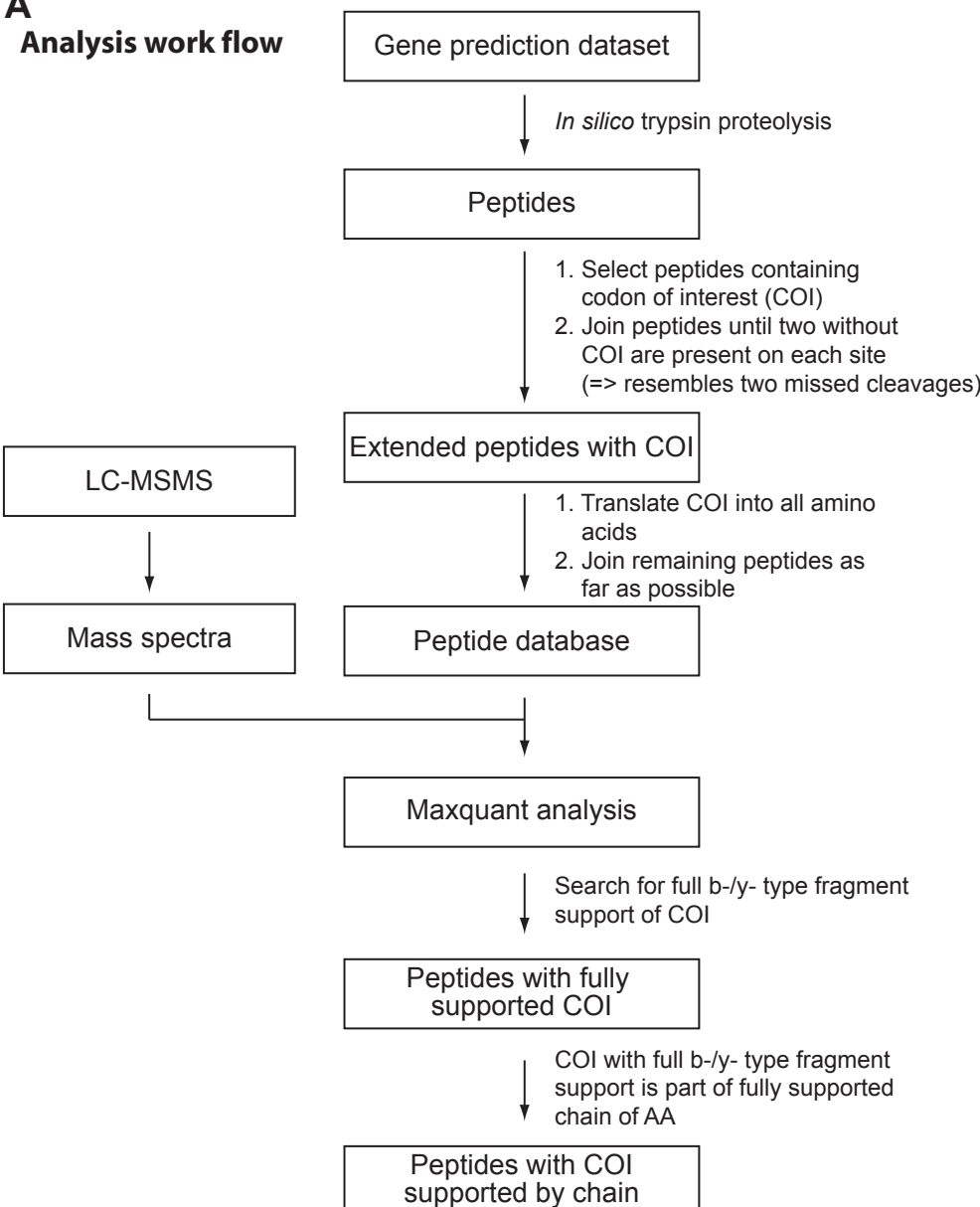


**Current Biology, Volume 28**

**Supplemental Information**

**Endogenous Stochastic Decoding  
of the CUG Codon by Competing  
Ser- and Leu-tRNAs in *Ascoidea asiatica***

**Stefanie Mühlhausen, Hans Dieter Schmitt, Kuan-Ting Pan, Uwe Plessmann, Henning Urlaub, Laurence D. Hurst, and Martin Kollmar**

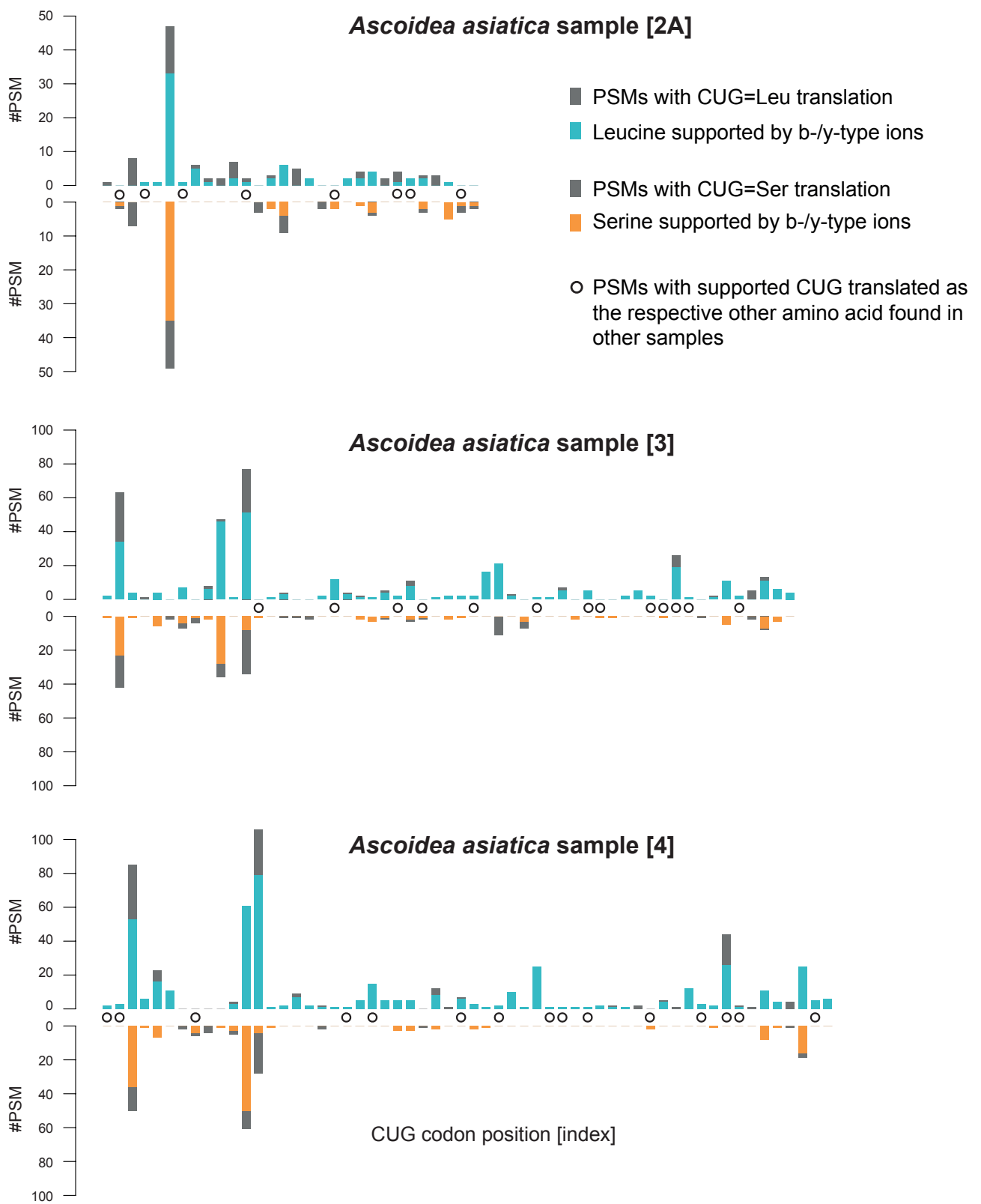


Gene index

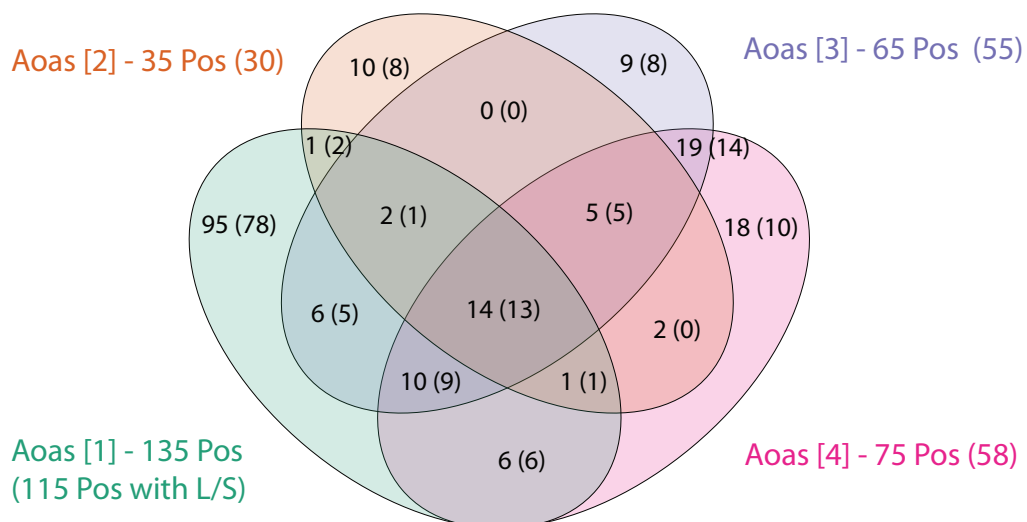
**Figure S1. Summary of the proteome analysis workflow and statistics, Related to Figure 1.**

A) For the database search we generated gene prediction datasets, in which all codons were iteratively translated into all amino acids. The peptides with differently translated codons have distinct total MW, accordingly different precursor ion masses and result in different spectra. B) Distribution of sequence coverage of identified proteins. C) Molecular weight of the respective proteins. D) Number of peptides matching to the respective protein. E) Number of non-redundant peptides matching to the respective protein. F) Number of peptide spectrum matches (PSMs) covering CUG-codon positions. G) Percentage of the CUG-codon positions per protein covered by the proteomics data. H) Representative LC-MS/MS spectra featuring CUG codons translated as serine and leucine (marked with stars).

**A**

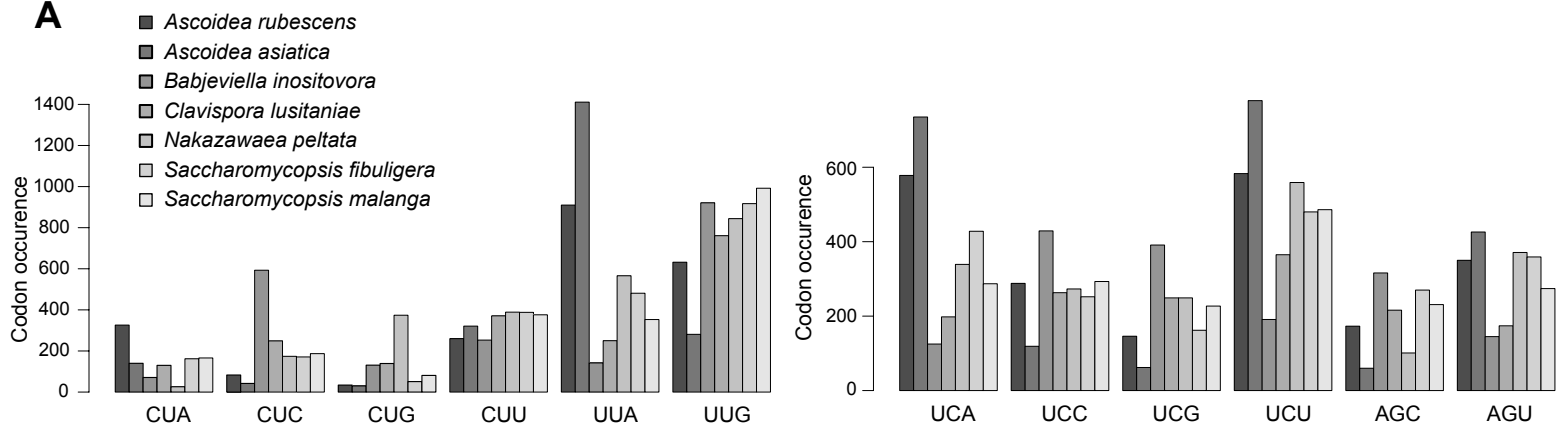
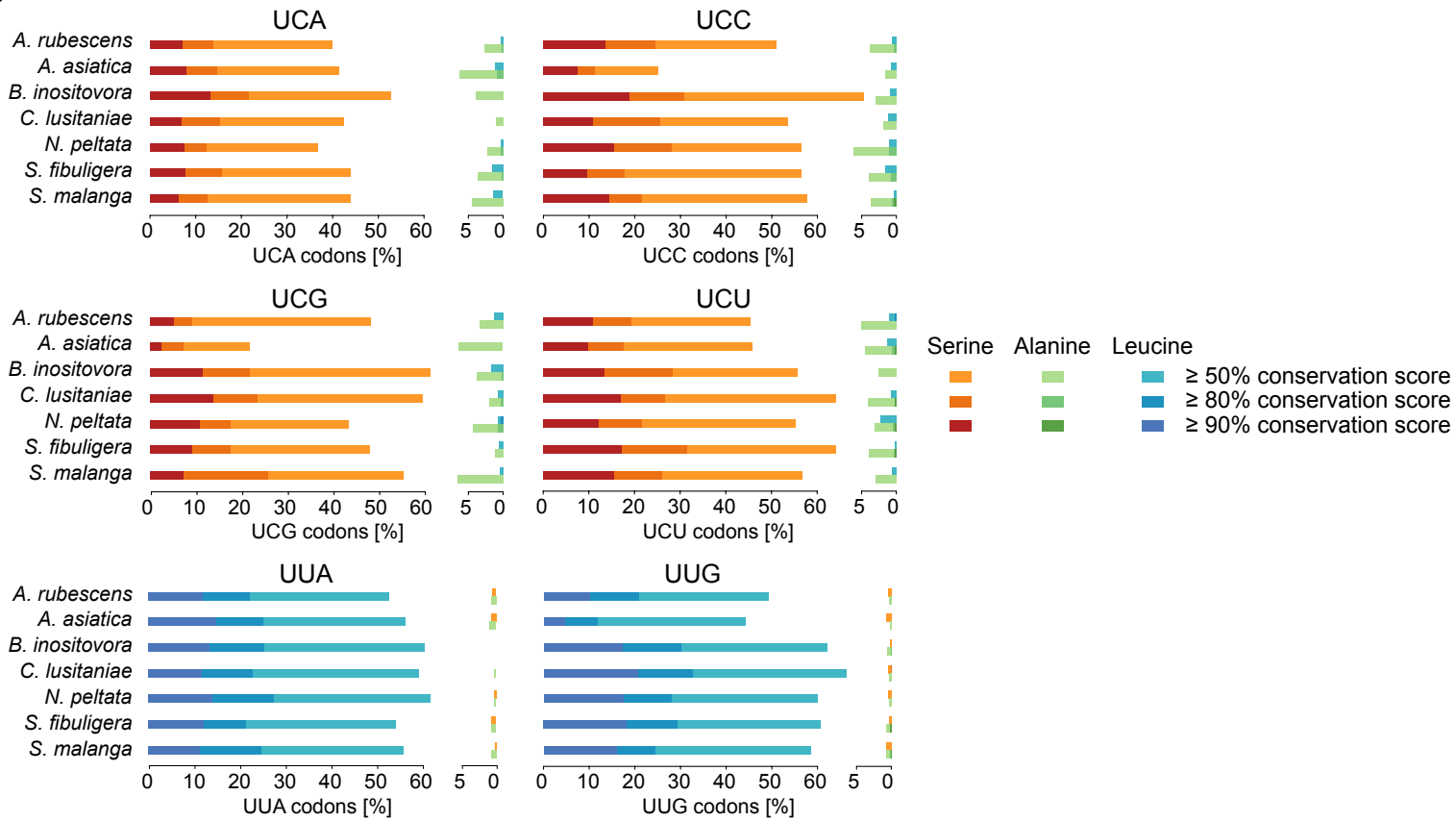
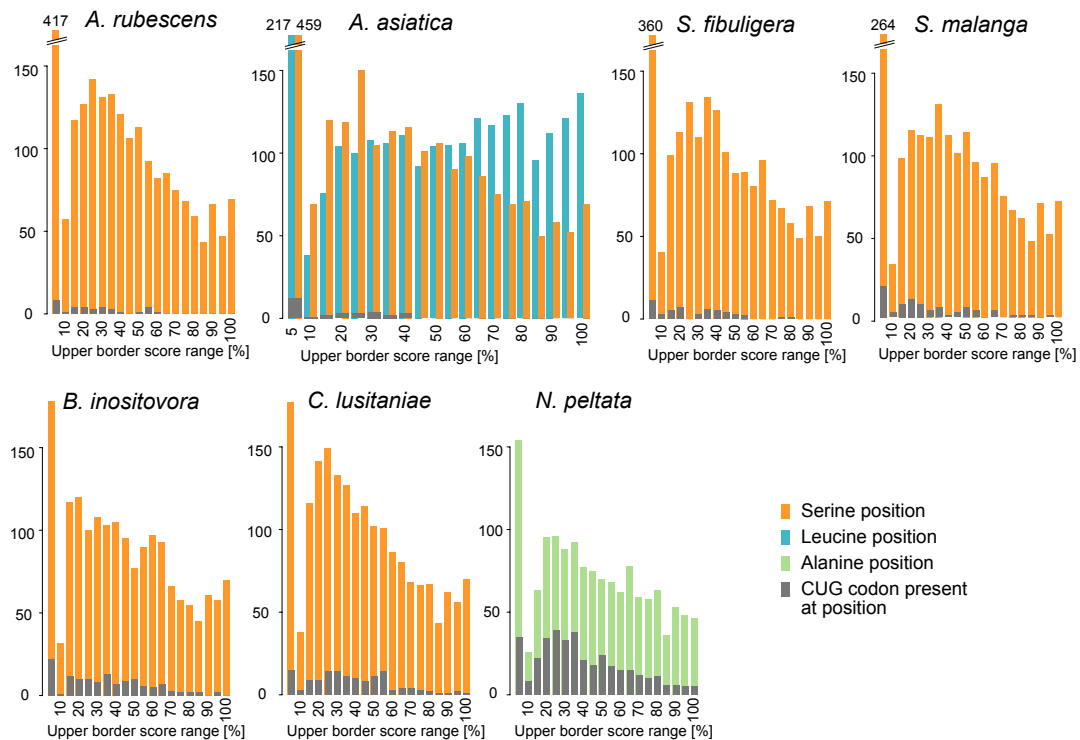


**B**



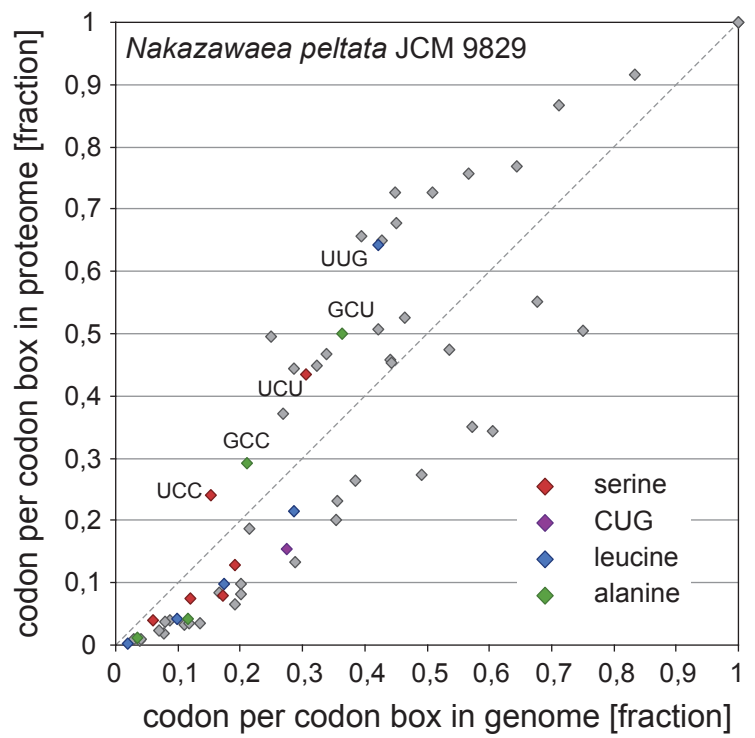
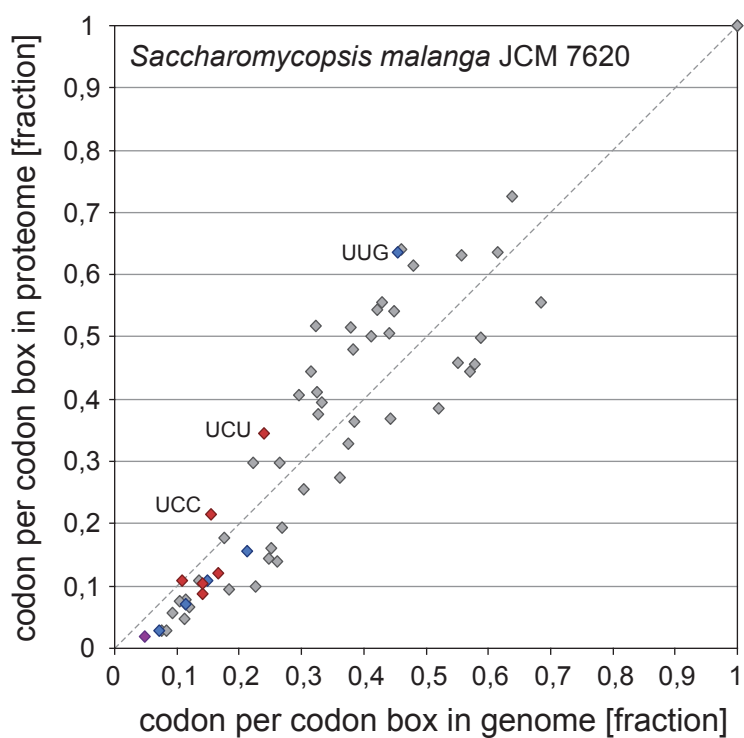
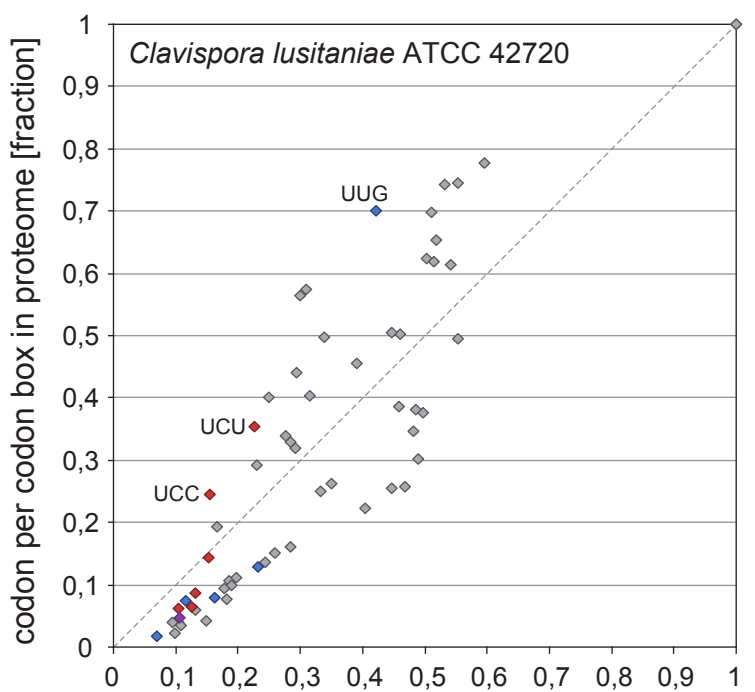
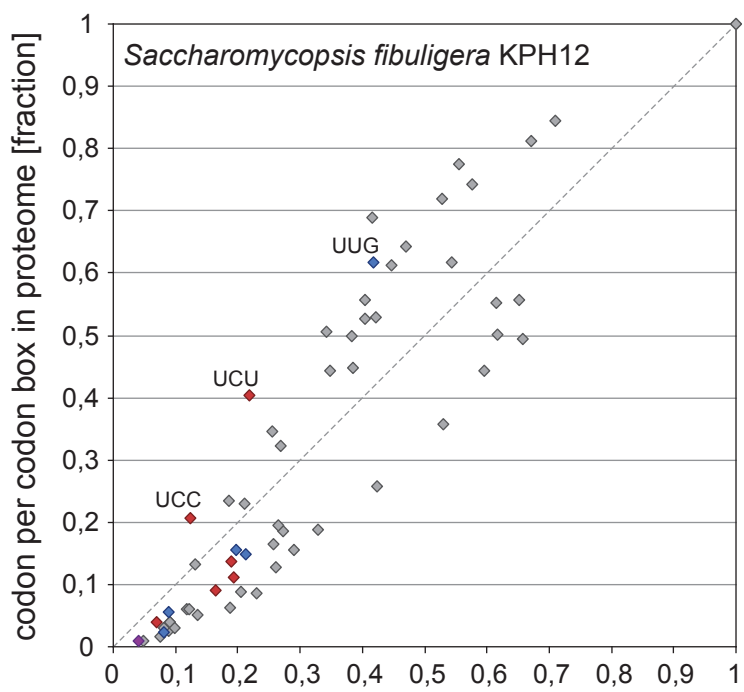
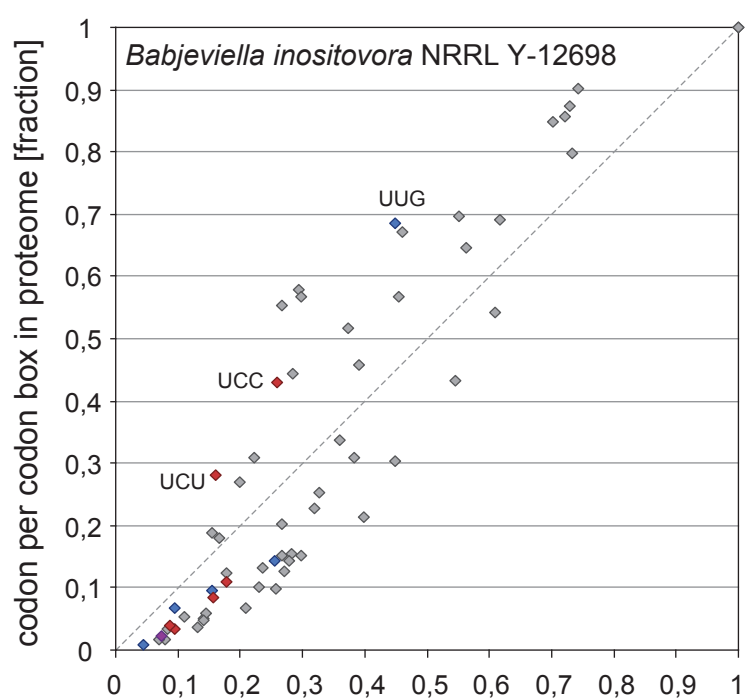
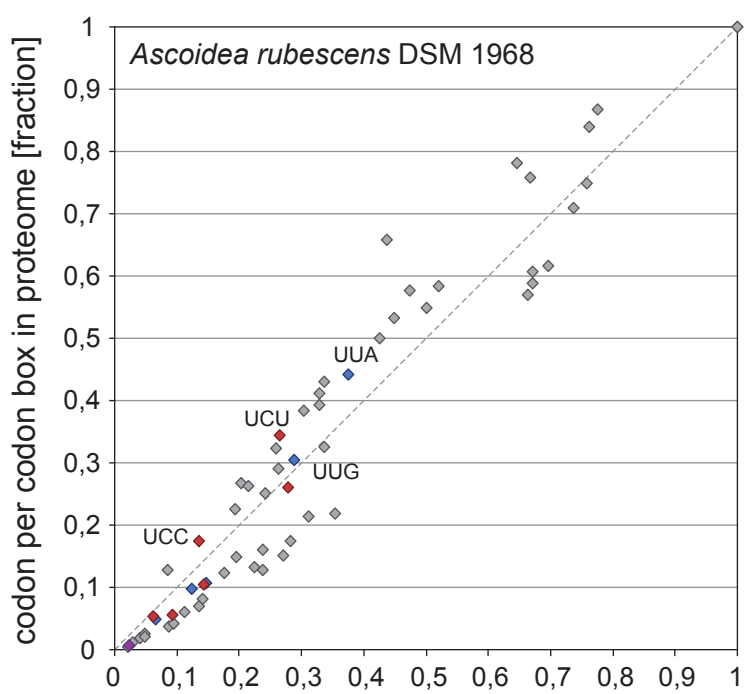
**Figure S2. Peptides with CUG codon positions found in *Ascoidea asiatica* samples [2A], [3] and [4], which were grown in different media, Related to Figure 1B.**

A) Gray bars denote the number of total PSMs covering a certain CUG position. These PSMs include those without support by b-/y-type ions. The PSMs with CUG positions supported by b-/y-type ions were colored: Blue bars represent PSMs with CUG translated as leucine and orange bars denote PSMs with CUG translated as serine. CUG positions exclusively translated with other amino acids than leucine or serine have been omitted. B) Overlap of supported CUG codon positions covered in different samples. Aoas [1] corresponds to the sample described in the main text, Aoas [2A], Aoas [3] and Aoas [4] denote further samples grown in different media. The numbers denote the total numbers of PSMs covering CUG codon positions including those not supported by b-/y-type fragment ions. For comparison, the numbers of CUG codons found with ambiguous translation (leucine or serine and, additionally, another amino acid) are given. CUG codon positions covered in multiple samples and translated by other amino acids indicate genome sequencing errors or differences between sequenced and analysed strain. This is true for one of the CUG codon positions covered by all four samples, and two positions covered by three samples. The majority of b-/y-type ion supported positions (95-99%) is, however, only translated by serine and leucine (see also Data S1).

**A****B****C**

**Figure S3. Number and conservation of leucine and serine codons in the cytoskeletal and motor protein sequence alignment, Related to Figure 3.**

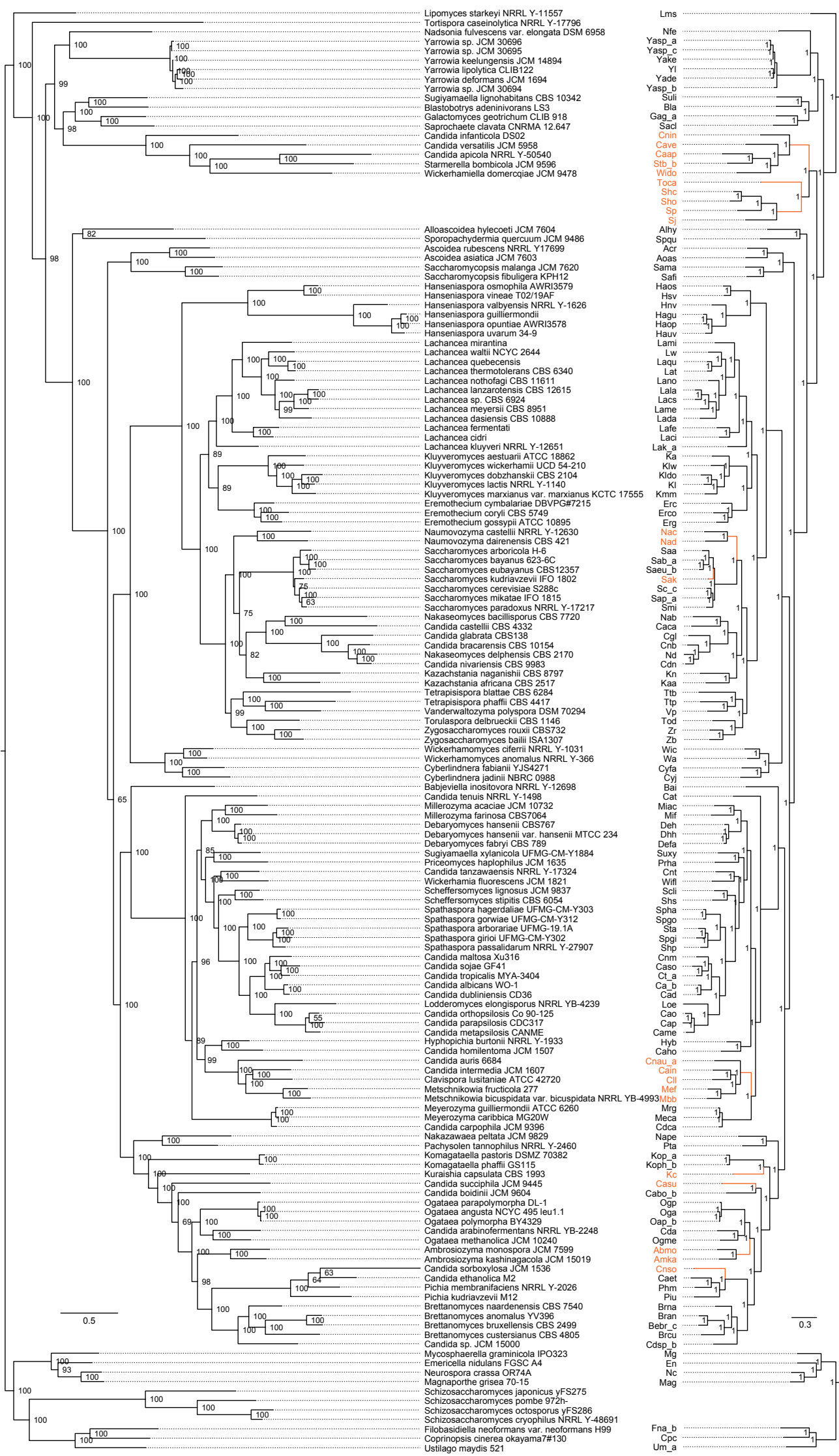
A) Occurrence of each of the leucine and serine codons in the alignment. B) Percentages of selected leucine and serine codons present at alignment positions of a certain conservation score. On the left, codons found at alignment positions enriched in the expected amino acid are shown, contrasted by codons found at alignment positions enriched in an unexpected amino acid on the right. Leucine and serine codons not shown here are part of Figure 3 of the main manuscript. C) Number of serine/ leucine/ alanine alignment positions falling into a score range. The number of those positions that contain at least one CUG codon is plotted in front. For *A. asiatica*, alignment positions with both serine and leucine have been considered.





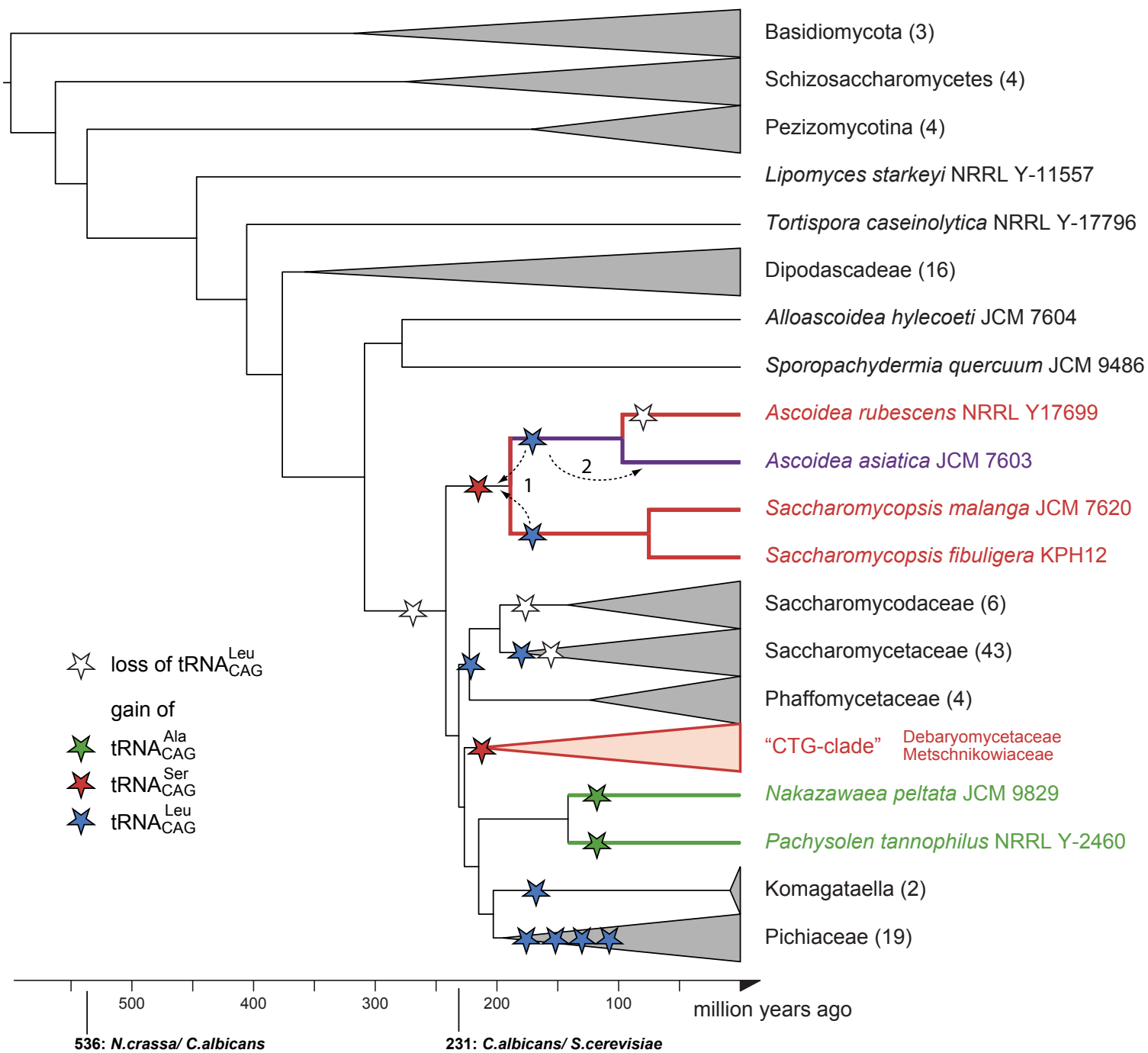
**Figure S4. Codon usage in genome versus proteome, Related to Figure 4.**

The scatter plots present the fraction of each codon per family box according to its usage in the genome, determined by analysis of the gene prediction datasets, versus its usage in the proteome, determined by analysis of the MSMS data. Serine, leucine, and alanine family box codons, and the CUG codons are highlighted by red, blue, green and purple colour, respectively. If the proteins found in the proteome were representative of the genome, all codons should be on the diagonal. However, the expressed proteins are encoded by preferred codons (upper left triangle), while other codons are considerably less used (lower right triangle). The CUG codons are all in the lower right triangle, meaning that they are mostly present in genes whose proteins were not detected in the proteomics analyses. Serine and leucine (and alanine in case of *N. peltata*) codons preferably used in the proteome are indicated for orientation.



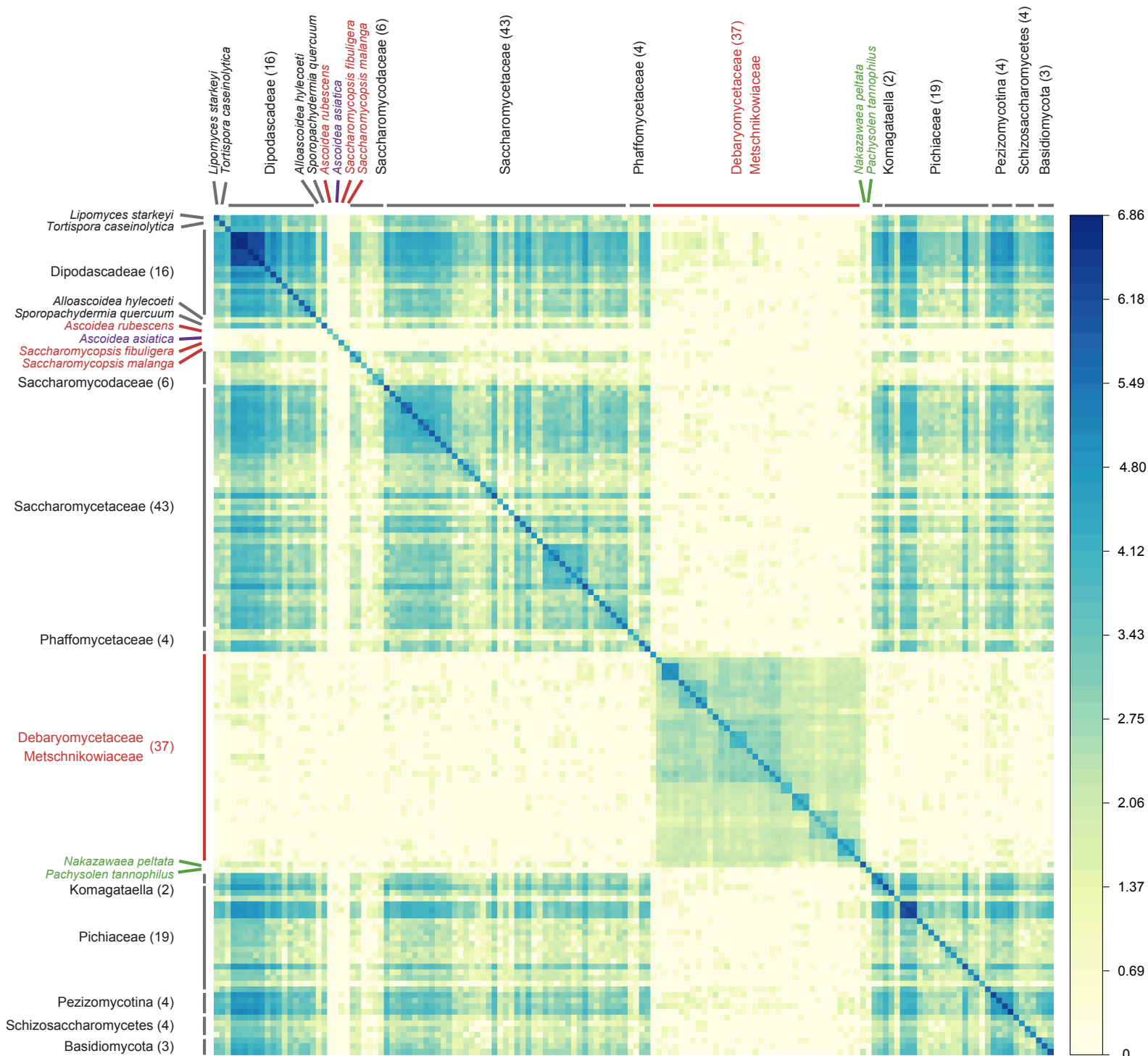
**Figure S5. Contrasting species trees generated with the Maximum-Likelihood and the Bayesian approach, Related to Figure 5.**

Left tree: RAxML generated tree with LG +G +I substitution model and 1000 bootstrap replicates. Right tree: MrBayes generated tree with mixed amino acid model and posterior probabilities given for branch support. Species and internal branchings differing between MrBayes and RAxML generated trees are indicated in red color in the MrBayes tree.



**Figure S6. Dating tRNA-loss and -gain events, Related to Figure 5.**

Dated RAxML generated tree with  $tRNA_{CAG}^{Leu}$  loss and gain events marked. The  $tRNA_{CAG}^{Leu}$  gain and loss events are placed according to [S1]. The multiple  $tRNA_{CAG}^{Leu}$  gain events in the Pichiaceae branch indicate multiple independent gains within this branch, as detailed in [S1] (see also the Leu tRNA phylogeny plot on FigShare). In an alternative scenario ("1"), the *Ascoidea* clade  $tRNA_{CAG}^{Leu}$  have a common origin. In a second alternative scenario ("2"), the  $tRNA_{CAG}^{Leu}$  could have been independently acquired by *A. asiatica* and the ancestor of the *Saccharomycopsis* yeasts, in which case *A. rubescens* never had this tRNA. Divergence times were estimated by TreePL based on constrains set on the splits between *Neurospora crassa* and *Candida. albicans* (536 million years ago) and *C. albicans* and *S. cerevisiae* (231 million years ago).



**Figure S7. Common CUG positions in 26 cytoskeletal and motor proteins of 148 fungi, Related to Figure 6.**

The diagonal denotes the total number of CUG positions. For displaying purposes, numbers have been log transformed. Some species names have been omitted and instead, the group name and number of species inside that group are given.

## Supplemental References

- S1. Mühlhausen, S., Findeisen, P., Plessmann, U., Urlaub, H., and Kollmar, M. (2016). A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. *Genome Res.* 26, 945–955.