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



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).



В

Α

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).



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.



Lipomyces starkeyi NRRL Y-11557 Tortispora caseinolytica NRRL Y-17796 Nadsonia fluvescens var. elongata DSM 6958 Yarrowia sp. JCM 30696 Yarrowia ipolytica CLB122 Yarrowia declungensis JCM 14894 Yarrowia deformans JCM 1694 Yarrowia deformans JCM 1694 Yarrowia deformans JCM 1694 Yarrowia des JCM 30694 Suglyamaella ilgnohabitans CBS 10342 Blastobotrys adeninivorans LS3 Galactomyces geotrichum CLIB 918 Saprochaete clavata CNRMA 12.647 Candida infaticola DS02 Candida versatilis JCM 5958 Candida versatilis JCM 5958 Starmerella bombicola JCM 9596 Wickerhamiella domercgiae JCM 9478

Alloascoidea hylecoeti JCM 7604 Sporopachydermia quercuum JCM 9496 Ascoidea rubescens NRRL Y17699 Ascoidea asiatica JCM 7603 Saccharomycopsis malanga JCM 7620 Saccharomycopsis malanga JCM 7620 Hanseniaspora vineae T02/19AF Hanseniaspora valtwangis UNPL Y 1567 nanseniaspora valbyensi NRRL Y-1626 Hanseniaspora valbyensis NRRL Y-1626 Hanseniaspora guilliermondii Hanseniaspora opuntiae AWRI3578 Hanseniaspora uvarum 34-9 Hainseniaspora uvarum 34-9 Lachancea mirantina Lachancea waltii NCYC 2644 Lachancea waltii NCYC 2644 Lachancea waltii NCYC 2644 Lachancea hermotolerans CBS 6340 Lachancea lanzarotensis CBS 12615 Lachancea anzarotensis CBS 12615 Lachancea meyersii CBS 8951 Lachancea dasiensis CBS 10888 Lachancea dermentati Lachancea kiuyveri NRRL Y-12651 Kluyveromyces wastuarii ATCC 18862 Kluyveromyces dobzhanskii CBS 2104 Nutyveromyces dożnanskii CBS 2104
Kluyveromyces marxianus var. marxianus KCTC 17555
Eremothecium coryli CBS 5749
Eremothecium coryli CBS 5749
Eremothecium coryli CBS 5749
Eremothecium coryli CBS 5749
Saccharomyces arbaricioa H-6
Saccharomyces eubayanus CBS12357
Saccharomyces eubayanus CBS12357
Saccharomyces cerevisiae 5288c
Saccharomyces balilisporus CBS 7720
Candida castellii CBS 432
Candida bracarensis CBS 2170
Candida albartai CBS 10154
Nakaseomyces delphensis CBS 2170
Candida anganistii CBS 803
Kazachstania africana CBS 2517
Tetrapisispora blatta CBS 6284
Tetrapisispora blatta CBS 6284
Tetrapisispora blatta CBS 6284
Torulaspora delbrueckii CBS 1146
Zygosaccharomyces salii ISA1307
Wickerhamomyces sainii (SRS 1054
Wickerhamomyces and Sata
Caygosaccharomyces sainii (SRS 10572
Zygosaccharomyces sainii (SRS 10572
Zygosaccharomyces sainii (SRS 10572
Zygosaccharomyces sainii (SRS 1054
Wickerhamomyces anomalus NRRL Y-1031
Wickerhamomyces anomalus NRRL Y-1031 Kluyveromyces dobzhanskii CBS 2104 Kluyveromyces lactis NRRL Y-1140 2ygosaccharomyces holli ISA1307 2ygosaccharomyces alomalus NRRL Y-1031 Wickerhamomyces anomalus NRRL Y-366 Cyberlindnera fabianii YJS4271 Cyberlindnera fabianii YJS4271 Cyberlindnera fabianii YJS4271 Cyberlindnera fabianii YJS4271 Cyberlindnera jadinii NBRC 0988 Babjeviella inositovora NRRL Y-12698 Candida tenuis NRRL Y-1498 Millerozyma acaciae JCM 10732 Millerozyma ecaciae JCM 10732 Debaryomyces hansenii CBS767 Debaryomyces hansenii CBS767 Debaryomyces hansenii CBS767 Debaryomyces hansenii CBS768 Sugiyamaella xylanicola UFMG-CM-Y1884 Priceomyces haplophilus JCM 1635 Candida tanzawaensis NRRL Y-17324 Wickerhamia fluorescens JCM 1821 Scheffersomyces lignosus JGM 9837 Scheffersomyces gino UFMG-CM-Y312 Spathaspora arborariae UFMG-CM-Y312 Spathaspora ginoi UFMG-CM-Y312 Spathaspora ginoi UFMG-CM-Y312 Spathaspora passaildarum NRRL Y-27907 Candida anticoa Xu316 Candida bicans WO-1 Candida aloicans WO-1 Candida aloicans KO-31 Candida aloicans KO-31 Candida aloicans KO-31 Candida aloicans WO-1 Candida aloicans KO-317 Candida aloicans KO-317 Candida antespailosis CDC317 Candida antespailosis CDC317 Candida auris 6684 Can Candida anis 6684 Candida intermedia JCM 1607 Clavispora lusitaniae ATCC 42720 Metschnikowia fructicola 277 Metschnikowia bicuspidata var. bicuspidata NRRL YB-Meyerozyma guillermondii ATCC 6260 Meyerozyma acribbica MG20W Candida caropohila JCM 9396 Nakazawaea peltata JCM 9820 Pachysolen tannophilus NRRL Y-2460 Komagataella phaffil GS115 Kuraishia capsulata CBS 1993 Candida boiciphila JCM 9445 Candida boiciphila JCM 9445 Candida boiciphila JCM 9604 Candida buccipinia JCM 9945 Candida buccipinia JCM 9945 Ogataea aprapolymorpha DL-1 Ogataea apusta NCY 495 leu1.1 Ogataea polymorpha BV4329 Candida arabinofermentans NRRL YB-2248 Ogataea methanolica JCM 10240 Ambrosiozyma kashinagacola JCM 15019 Candida sorboxylosa JCM 1536 Candida ethanolica M2 Pichia membroxylosa JCM 1536 Candida ethanolica M2 Pichia membranifaciens NRRL Y-2026 Pichia kudriavzevii M12 Brettanomyces bruxellensis CBS 7540 Brettanomyces custersianus CBS 4805 Candida sp. JCM 15000 Mycosphaerella graminicoja IPO323 Mycosphaerella graminicola IPO323 Emericella nidulans FGSC A4 Neurospora crassa OR74A Magnaporthe grisea 70-15 wagnaportne grisea 70-15 Schizosaccharomyces japonicus yFS275 Schizosaccharomyces pombe 972h-Schizosaccharomyces octosporus yFS286 Schizosaccharomyces cryophilus NRRL Y-48691 Filobasidiella neoformans var. neoformans H99 Coprinopsis cinerea okayama7#130 Ustilago maydis 521



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} 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

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.