



# Tracing Eukaryotic Ribosome Biogenesis Factors Into the Archaeal Domain Sheds Light on the Evolution of Functional Complexity

Mehmet Birikmen<sup>1</sup>, Katherine E. Bohnsack<sup>2</sup>, Vinh Tran<sup>1</sup>, Sharvari Somayaji<sup>1</sup>, Markus T. Bohnsack<sup>2,3</sup> and Ingo Ebersberger<sup>1,4,5\*</sup>

<sup>1</sup>Applied Bioinformatics Group, Institute of Cell Biology and Neuroscience, Goethe University Frankfurt, Frankfurt, Germany, <sup>2</sup>Department of Molecular Biology, University Medical Center Göttingen, Göttingen, Germany, <sup>3</sup>Göttingen Center for Molecular Biosciences, Georg-August University, Göttingen, Germany, <sup>4</sup>Senckenberg Biodiversity and Climate Research Center (S-BIK-F), Frankfurt, Germany, <sup>5</sup>LOEWE Center for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt, Germany

## OPEN ACCESS

### Edited by:

Anna La Teana,  
Polytechnic University of Marche,  
Italy

### Reviewed by:

Fabian Blombach,  
University College London,  
United Kingdom  
Arlen Johnson,  
University of Texas at Austin,  
United States

### \*Correspondence:

Ingo Ebersberger  
ebersberger@bio.uni-frankfurt.de

### Specialty section:

This article was submitted to  
Biology of Archaea,  
a section of the journal  
Frontiers in Microbiology

**Received:** 09 July 2021

**Accepted:** 17 August 2021

**Published:** 16 September 2021

### Citation:

Birikmen M, Bohnsack KE, Tran V,  
Somayaji S, Bohnsack MT and  
Ebersberger I (2021) Tracing  
Eukaryotic Ribosome Biogenesis  
Factors Into the Archaeal Domain  
Sheds Light on the Evolution of  
Functional Complexity.  
Front. Microbiol. 12:739000.  
doi: 10.3389/fmicb.2021.739000

Ribosome assembly is an essential and carefully choreographed cellular process. In eukaryotes, several 100 proteins, distributed across the nucleolus, nucleus, and cytoplasm, co-ordinate the step-wise assembly of four ribosomal RNAs (rRNAs) and approximately 80 ribosomal proteins (RPs) into the mature ribosomal subunits. Due to the inherent complexity of the assembly process, functional studies identifying ribosome biogenesis factors and, more importantly, their precise functions and interplay are confined to a few and very well-established model organisms. Although best characterized in yeast (*Saccharomyces cerevisiae*), emerging links to disease and the discovery of additional layers of regulation have recently encouraged deeper analysis of the pathway in human cells. In archaea, ribosome biogenesis is less well-understood. However, their simpler sub-cellular structure should allow a less elaborated assembly procedure, potentially providing insights into the functional essentials of ribosome biogenesis that evolved long before the diversification of archaea and eukaryotes. Here, we use a comprehensive phylogenetic profiling setup, integrating targeted ortholog searches with automated scoring of protein domain architecture similarities and an assessment of when search sensitivity becomes limiting, to trace 301 curated eukaryotic ribosome biogenesis factors across 982 taxa spanning the tree of life and including 727 archaea. We show that both factor loss and lineage-specific modifications of factor function modulate ribosome biogenesis, and we highlight that limited sensitivity of the ortholog search can confound evolutionary conclusions. Projecting into the archaeal domain, we find that only few factors are consistently present across the analyzed taxa, and lineage-specific loss is common. While members of the Asgard group are not special with respect to their inventory of ribosome biogenesis factors (RBFs), they unite the highest number of orthologs to eukaryotic RBFs

in one taxon. Using large ribosomal subunit maturation as an example, we demonstrate that archaea pursue a simplified version of the corresponding steps in eukaryotes. Much of the complexity of this process evolved on the eukaryotic lineage by the duplication of ribosomal proteins and their subsequent functional diversification into ribosome biogenesis factors. This highlights that studying ribosome biogenesis in archaea provides fundamental information also for understanding the process in eukaryotes.

**Keywords: domain architecture evolution, asgard group, phylogenetic profiles, orthology assignment, evolutionary traceability, pathway complexity, large subunit maturation**

## INTRODUCTION

Due to the essential role of ribosomes in producing all cellular proteins, their synthesis is among the few pathways that are universally necessary for organismic life. Despite substantial differences in the ways that the ribosomal proteins (RPs) and ribosomal RNAs (rRNAs) are assembled into the small and large ribosomal subunits (SSU and LSU, respectively), the fundamental basis of this process in the three domains of life (Woese et al., 1990) presumably dates back to the last universal common ancestor (LUCA). Ribosome assembly is most simple and best understood in bacteria. Bacterial ribosomes can be assembled *in vitro* without the requirement for any non-ribosomal proteins while up to approximately 30 assembly factors make the process much faster and more accurate *in vivo* (Shajani et al., 2011). In contrast, the functional network mediating the same process in eukaryotes is many-fold more extensive, and it is to date best studied in yeast (*Saccharomyces cerevisiae*; Hage and Tollervey, 2004; Henras et al., 2008; Sloan et al., 2017; Klinge and Woolford, 2019). A recent study listed 255 ribosome biogenesis factors with a confirmed or suspected direct role in this process (Ebersberger et al., 2014). Assigning these factors to different age layers revealed that precisely how eukaryotes mediate ribosome assembly is not cast into stone. Instead, an evolutionarily old set of core functions, whose emergence predates the diversification of eukaryotes, was extended and probably fine-tuned in a lineage-specific manner (Ebersberger et al., 2014). Archaea seem to assume an intermediate position between eukaryotes and bacteria with respect to the complexity of their ribosome assembly pathways (Lecompte et al., 2002; Londei and Ferreira-Cerca, 2021). Shifting focus toward the archaeal domain therefore has the potential to further disentangle the building principles of ribosomes that already evolved prior to the emergence of eukaryotes and were already established in the last eukaryotic common ancestor (LECA). Thus far, phylogenetic profiling indicated that 38 yeast RBFs have counterparts in the archaeal domain (Ebersberger et al., 2014), in parts representing functional sub-clusters that seem specifically involved in the late steps of ribosome maturation. A subset of these factors has been subsequently confirmed as RBFs by functional studies in individual archaeal models (reviewed in Londei and Ferreira-Cerca, 2021). However, we are far from fully comprehending the extent to which archaeal ribosome biogenesis resembles that of eukaryotes, where archaea have implemented alternative strategies, and the degree of diversity within the archaeal domain.

Much of the uncertainty about the common grounds of ribosome biogenesis in archaea and eukaryotes is connected to methodological issues in the large-scale profiling studies performed thus far. The objective of such studies is easily specified: “Identify the functionally equivalent archaeal protein to a eukaryotic RBF, if it is present.” Its realization, however, bears numerous pitfalls. Unidirectional searches for sequences with a significant local similarity e.g., with BLAST (Altschul et al., 2009) or Diamond (Buchfink et al., 2015) rapidly identify archaeal RBF candidates at a high sensitivity, however, the specificity is low (Chen et al., 2007). Considering only such proteins as RBF candidates that diverged no longer than the species they reside in (orthologs; e.g., Koonin et al., 2001; Ebersberger et al., 2014) reduces the false positive rate, because orthologs are the best guesses when searching for functionally equivalent proteins across species (Tatusov et al., 1997; Altenhoff et al., 2012). However, the higher specificity comes at the cost of a decrease in sensitivity (Altenhoff et al., 2016). Moreover, also orthologs may have diverged in function (Ebersberger et al., 2014). The latter problem can be ameliorated by scoring the similarity of protein domain architectures (Koestler et al., 2010), which comprise features, such as Pfam (El-Gebali et al., 2019) or Smart (Letunic et al., 2009) domains, transmembrane regions, signal peptides, or regions of biased amino acid composition. Differences in domain architectures can then indicate alterations in the respective functional spectra of the compared proteins (Jiang et al., 2020). In contrast to the ample means to cope with the spurious inference of archaeal RBF candidates based on homologous proteins, false negatives, i.e., proteins overlooked in the homolog searches because their sequences have diverged to an extent that they are no more similar than it is expected by chance, have received very little attention. Such proteins can represent missing links that are essential for concluding on the presence of functional (sub-)network based on phylogenetic profiling analyses (Jain et al., 2019). Their identification requires a targeted increase of the search sensitivity accompanied by a careful downstream curation to validate the results.

Next to the methodological issues outlined above, at least three further aspects leave the current understanding of common concepts in eukaryotic and archaeal ribosome biogenesis incomplete. In recent years, numerous yeast proteins have been discovered as RBFs (e.g., Fujiyama-Nakamura et al., 2009; van Tran et al., 2019) whose presence in the archaea has not been tested in a comprehensive screen, thus far. Moreover, structure- and mass spectrometry-based approaches have shed light on the order of selected events during the assembly process and

the identity of the participating proteins (see for example, Wu et al., 2016; Barandun et al., 2017; Kater et al., 2017; Sanghai et al., 2018; Klingauf-Nerurkar et al., 2020; Liang et al., 2020; Nieto et al., 2020). This resource, which provides an excellent basis for identifying functional sub-networks shared between eukaryotes and archaea, is largely untapped. Moreover, the focus on yeast, a highly derived model organism that lost many genes essential in other species (Peter et al., 2018), is a limiting factor itself. Knowledge on the yeast RBFs is now complemented by large-scale, RNAi-based screens in human cells, which have revealed several 100 proteins that may be either directly or indirectly required for human ribosome biogenesis (Wild et al., 2010; Tafforeau et al., 2013; Badertscher et al., 2015; Farley-Barnes et al., 2018). For a non-negligible fraction of these RBF candidates a yeast homolog is elusive (Fujiyama-Nakamura et al., 2009; Tafforeau et al., 2013; van Tran et al., 2019; Ameismeier et al., 2020). It is conceivable that at least some of these factors act as RBFs in archaea too. Lastly, both the number of archaeal taxa and their phylogenetic diversity in the public sequence databases has dramatically increased in the past few years. This data provides an excellent, yet largely unused, basis for a highly-resolved analysis of the representation of eukaryotic RBFs in the archaeal domain. Among others, it allows to test whether members of the Asgard group share more RBFs with the eukaryotes than other archaea, which would be in line with their suspected placement as the closest relative of the eukaryotes (Zaremba-Niedzwiedzka et al., 2017; Imachi et al., 2020; Liu et al., 2021).

Here, we re-address the evolutionary history of eukaryotic ribosome biogenesis and trace the deep evolutionary roots of this pathway that are shared with the archaea. We base our analysis on a set of 301 manually curated yeast and human RBFs comprising the, to date, most comprehensive collection of eukaryotic ribosome biogenesis factors. Phylogenetic profiles for each protein across more than 900 taxa, among them 727 archaea including representatives of the Asgard group, using targeted and domain architecture aware ortholog searches provide insights into the evolution of this pathway at an unprecedented resolution. The analysis is complemented by identifying RBFs that evolve too quickly to facilitate ortholog identification over longer evolutionary time scales. This helps reconciling the discrepancy between large-scale phylogenetic profiling of RBFs using ortholog/homology assignments claiming the absence of an RBF and experimental evidences showing the presence of the corresponding function. In the same line, it can direct the attention to missing functional links that may require searches at higher sensitivity to warrant their detection.

## MATERIALS AND METHODS

### Selection of Ribosome Biogenesis Factors

We compiled an initial non-redundant set of 307 yeast proteins putatively involved into ribosome biogenesis. About 255 RBF candidates were obtained from Ebersberger et al. (2014), 41 from the KEGG Brite database (KO3009; Kanehisa et al., 2016), and additionally 11 from recent publications focusing

on molecular details of yeast ribosome biogenesis. The candidate list together with the references is given in **Supplementary Table S1**. For the human RBF collection, we seeded the set with 198 proteins that are involved into human ribosome biogenesis according to KEGG (Kanehisa et al., 2016). This data were complemented with 488 proteins with at least a suspected involvement into human ribosome biogenesis according to large scale screening studies (40S/60S: Wild et al., 2010; pre-rRNA processing: Tafforeau et al., 2013; 40S: Badertscher et al., 2015; and regulators: Farley-Barnes et al., 2018). Finally, we added eight RBFs that emerged from a literature screen (Yang et al., 2006; Freed et al., 2012; Wandrey et al., 2015; van Tran et al., 2019; Ameismeier et al., 2020). The non-redundant list of 695 proteins excluding ribosomal proteins is provided in **Supplementary Table S2**. The phylogenetic profiles of all human candidates were determined and served as basis to assess their evolutionary age (see below).

### Taxon Collection

We determined the phylogenetic profiles for the RBF candidates across 982 taxa comprising 232 eukaryotes, a diverse collection of 23 bacteria representing 16 phyla, and 727 archaea. Archaeal gene sets were retrieved from the RefSeq partition of the NCBI Genome database (O'Leary et al., 2016; December 2020). The 78 Quest for Orthologs reference proteomes<sup>1</sup> complemented with 189 fungal taxa served as representatives for the bacterial and eukaryotic taxa. The taxon list is provided in **Supplementary Table S3**.

### Pathway Analysis

Information about pathway organization and complex composition was obtained from KEGG (Kanehisa et al., 2021) and Reactome v76 (Fabregat et al., 2018).

### Domain-Aware Phylogenetic Profiling and Gene Age Estimation

RBF orthologs were identified with the targeted ortholog search tool fDog (Jiang et al., 2020)<sup>2</sup> using the following parameter settings: `-checkCoorthologsRef`, `--countercheck`, `--minDist=family`, and `--maxDist=kingdom`. Protein domain architectures were compared pair-wise between each seed protein and its orthologs, and an architecture similarity score was computed with FAS<sup>3</sup> implemented into fDOG. In brief, FAS compares the domain architectures of two proteins, using one architecture as the reference and the second architecture as the query. The FAS score ranges between 0 (architectures are completely dissimilar) and 1 (the architecture of the reference is at least a sub-architecture of the query; Koestler et al., 2010). Because the score is not symmetric, we computed FAS scores once using the seed protein as reference (FAS forward), and once using the ortholog as reference (FAS backward). The Domain architecture-aware phylogenetic profiles generated by

<sup>1</sup>[https://www.ebi.ac.uk/reference\\_proteomes/](https://www.ebi.ac.uk/reference_proteomes/)

<sup>2</sup><https://github.com/BIONF/fDog/>

<sup>3</sup><https://github.com/BIONF/FAS>

fDog were visualized and analyzed with PhyloProfile (Tran et al., 2018). The evolutionary emergence of individual RBF candidates was dated using an LCA algorithm implemented into PhyloProfile. In brief, the last common ancestor (LCA) of the two most distantly related species harboring an ortholog to an RBF candidate was used as a minimal age estimate for the corresponding gene. Comparisons of FAS score distributions and domain architecture comparisons between two taxonomic groups were performed with the group comparison function implemented into PhyloProfile.

## Computation of Evolutionary Traceability

Orthologs of quickly evolving proteins may lose a sufficiently high sequence similarity to warrant their detection already over short evolutionary distances. The evolutionary traceability index (Ti) is a simulation-based score on the interval [0,1] that captures, for a protein and a given evolutionary distance, whether an ortholog likely still shares a sufficiently high sequence similarity that allows its detection, or whether it likely has diverged beyond recognition (Jain et al., 2019). We extracted, for each yeast protein, the corresponding traceability indices across 273 taxa distributed across the tree of life from Jain et al. (2019). Traceability indices on a kingdom level were computed as the mean Ti across all members in the kingdom. An RBF was considered as low traceable in case the mean traceability index fell below 0.75 (Jain et al., 2019). Note, we limited this computation to the yeast proteins, since the human RBFs were pre-selected based on the presence of orthologs already in evolutionary distantly related taxa.

## Phylogenetic Analyses

Multiple Sequence alignments were generated with Muscle v3.8.1551 (Edgar, 2004) using default settings. Sequence logos were generated with WebLogo (Crooks et al., 2004) provided as a web service available at <https://weblogo.berkeley.edu>. For maximum likelihood phylogenetic tree reconstruction with IQ-TREE using the LG+G+I model of sequence evolution (Nguyen et al., 2015), alignments were post-processed by removing alignment columns with more than 50% gaps. Phylogenetic trees were visualized with iTol (Letunic and Bork, 2021).

## RESULTS

### Ribosome Biogenesis From Two Perspectives

Production of ribosomes is essential for organismic life. Here, we set out to identify the common basis of ribosome biogenesis in eukaryotes and archaea using a manually curated list of eukaryotic RBFs (Figure 1). The most comprehensive inventory of protein factors involved in eukaryotic ribosome biogenesis is that of the yeast *S. cerevisiae* S288C (Woolford and Baserga, 2013; Bassler and Hurt, 2019). An initial screen for yeast RBFs retrieved 307 candidates (see Figure 1 for numbers and references). To consider also eukaryotic RBFs that are either absent in yeast or have a function other than RBF, we used an initial set of 686 potential human RBFs obtained largely from comprehensive

RNAi-based screens (Wild et al., 2010; Tafforeau et al., 2013; Badertscher et al., 2015; Farley-Barnes et al., 2018) as a second starting point of the analysis.

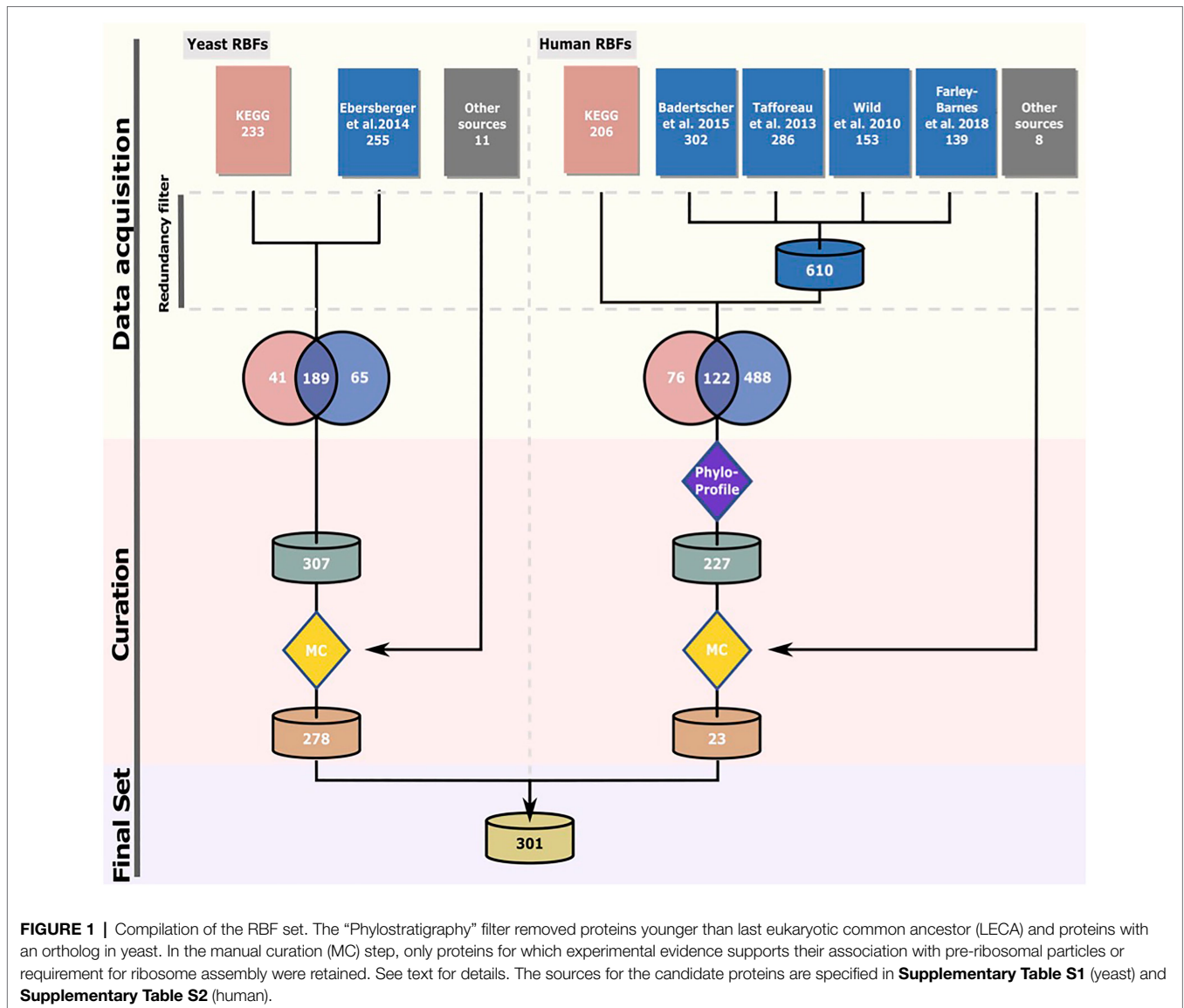
The phylogenetic profile for each RBF candidate was subsequently determined by searching for orthologs in 232 eukaryotes, 727 archaea, and 23 bacteria. For each detected ortholog, we then computed the domain architecture similarity to that of the corresponding yeast or human seed protein. On this basis, a two-stage curation procedure was devised to extract the final set of RBF candidates for further analysis. In the evolutionary-motivated first stage (Supplementary Figure S1; Supplementary Table S4), the subset of RBFs that can be traced back to the LCA of all eukaryotes (LECA) was identified. We then removed all proteins with a yeast ortholog that is considered already in the yeast RBF collection leaving 227 human proteins. We implemented this first stage for the human RBF candidates to reduce the number of proteins that enter the second stage of manual curation. In the subsequent curation step, we retained only proteins that fulfill at least one of the following criteria: (i) have a known function in ribosome biogenesis; (ii) have been identified associated with pre-ribosomal particles; or (iii) their depletion or deletion causes a defect in ribosome assembly. This retained 278 yeast proteins (RBF<sub>yeast</sub>) and 17 human proteins (RBF<sub>human</sub>). We complemented this data with six known human RBFs, ILF2, ILF3, TMA16, NOL11, NKRF, and ZCCHC4, most likely younger than the LECA. The final set, RBF<sub>euk</sub>, comprised 301 factors. We added two subunits of the RNA polymerase II, Rpb5, and Rpo21, as positive controls for our profiling approach. The two proteins are evolutionarily highly conserved and orthologs should be identifiable in the majority of the species analyzed here. Thus, the total number of analyzed proteins sums up to 303.

### PhyloRBF: Interactive Access to the Data

The phylogenetic profiles for proteins in the RBF<sub>euk</sub> set across 982 taxa provide the first unifying resource for tracing eukaryotic ribosome biogenesis factors across the organismal diversity. To facilitate easy and interactive exploration and analysis of these data, we provide two options. An online instance of PhyloProfile (Tran et al., 2018), PhyloRBF, provides web-access to these data via the URL: <https://aplbio.biologie.uni-frankfurt.de/phyloRBF>. Users can display the full data set (Figure 2A), customized subsets of RBFs and taxa (Figure 2B), zoom in on individual ortholog pairs (Figure 2C), and ultimately display the domain architectures of the yeast or human RBF and of its ortholog (Figure 2D). Interactive links connect the information about taxon, protein sequence and Pfam (El-Gebali et al., 2019), or SMART domains (Letunic et al., 2009) with the corresponding public databases. For an offline analysis of the RBF<sub>euk</sub> phylogenetic profiles with a local installation of PhyloProfile, the input data are provided as Supplementary Data 2.

### The Evolutionary Trajectory of Eukaryotic Ribosome Biogenesis Factors

Figure 3A provides an overview of the phylogenetic profiles for the RBF<sub>euk</sub> set and the two positive controls. Within

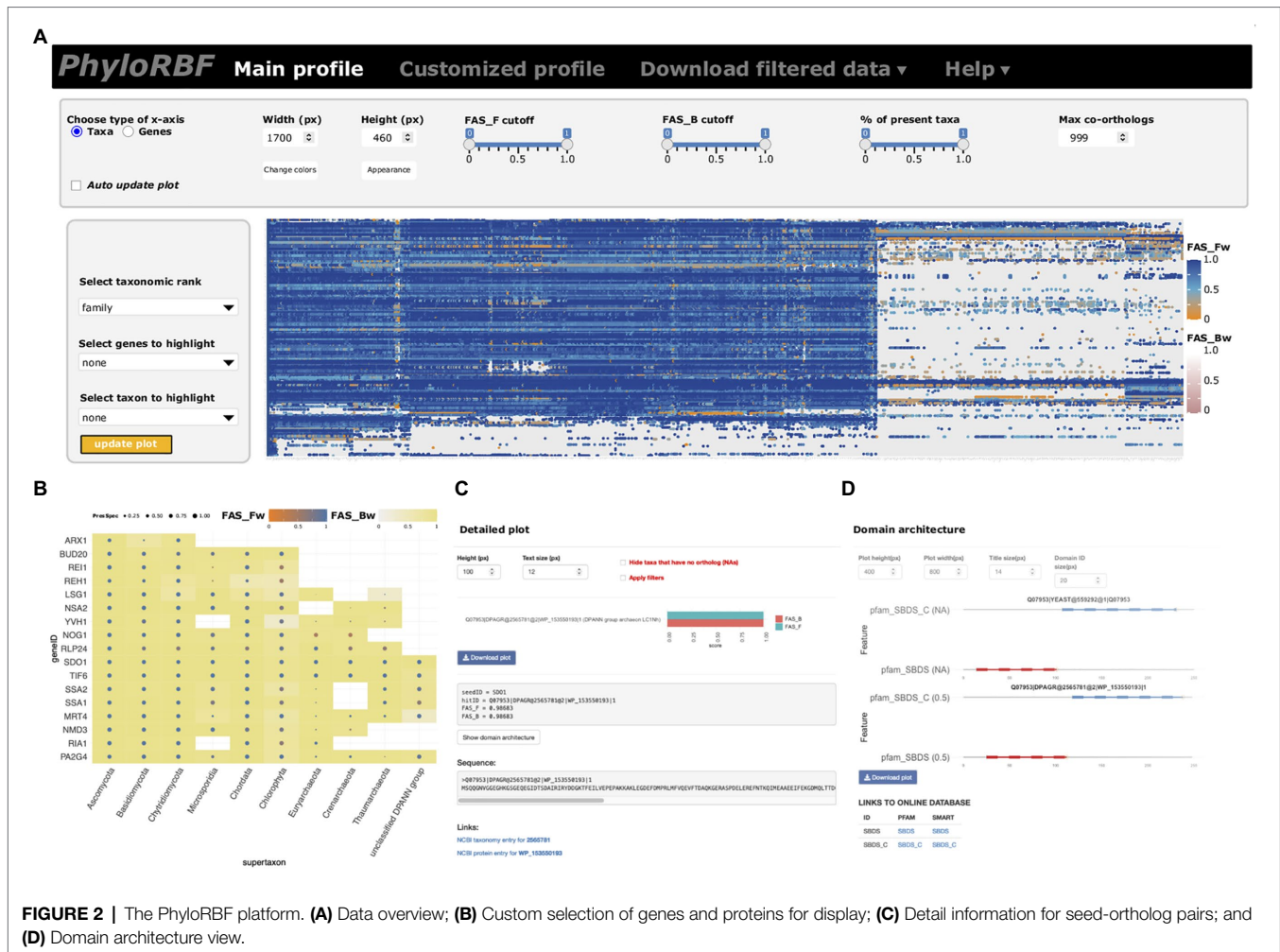


eukaryotes, orthologs of most RBFs are represented in all investigated lineages. Losses of RBFs are rare, but nevertheless they seem to occur. Examples exist of yeast RBFs lacking orthologs in either all or at least most of the Pezizomycotina, the sister clade of the Saccharomycotina, although orthologs are found in more distantly related fungi (**Supplementary Figure S2**). Substantially more pronounced is the apparent loss of RBFs in the microsporidia, obligate intracellular fungal parasites, which lack orthologs to 56 yeast RBFs that can be traced back to LECA (**Supplementary Table S4**). This reveals that the overall trend of microsporidia to evolve toward a highly simplified variant of a eukaryotic organism (Keeling et al., 2010) extends also to ribosome biogenesis. Likewise, 17 human RBFs are represented in a diverse set of eukaryotes, but orthologs are missing either in all fungi, or specifically on the lineage toward *S. cerevisiae* (see below).

In contrast to the overall conservation of RBFs in the eukaryotic domain, the representation of orthologs in archaea

and bacteria is substantially sparser. Bacteria lack orthologs for most the eukaryotic RBFs, and only individual factors are detected consistently across the sampled taxa. The picture in archaea is more differentiated; archaea possess orthologs of a considerable number of RBFs. For individual factors, orthologs are consistently detected across the individual archaeal groups, whereas for others, they are confined to individual lineages or they occur only sporadically. Overall, very few factors are ubiquitously present across all three domains of life or are consistently present in the bacterial representatives but not in the archaea.

To provide a more quantitative view of these observations, we stratified the proteins into different age layers. In the most permissive setting (**Figures 3B–D**), we assessed the minimal evolutionary age of a protein by assigning it to the LCA that the seed species shares with the most distantly related species in which an ortholog was detected. However, such unfiltered data may suffer from overestimates due to spurious orthology



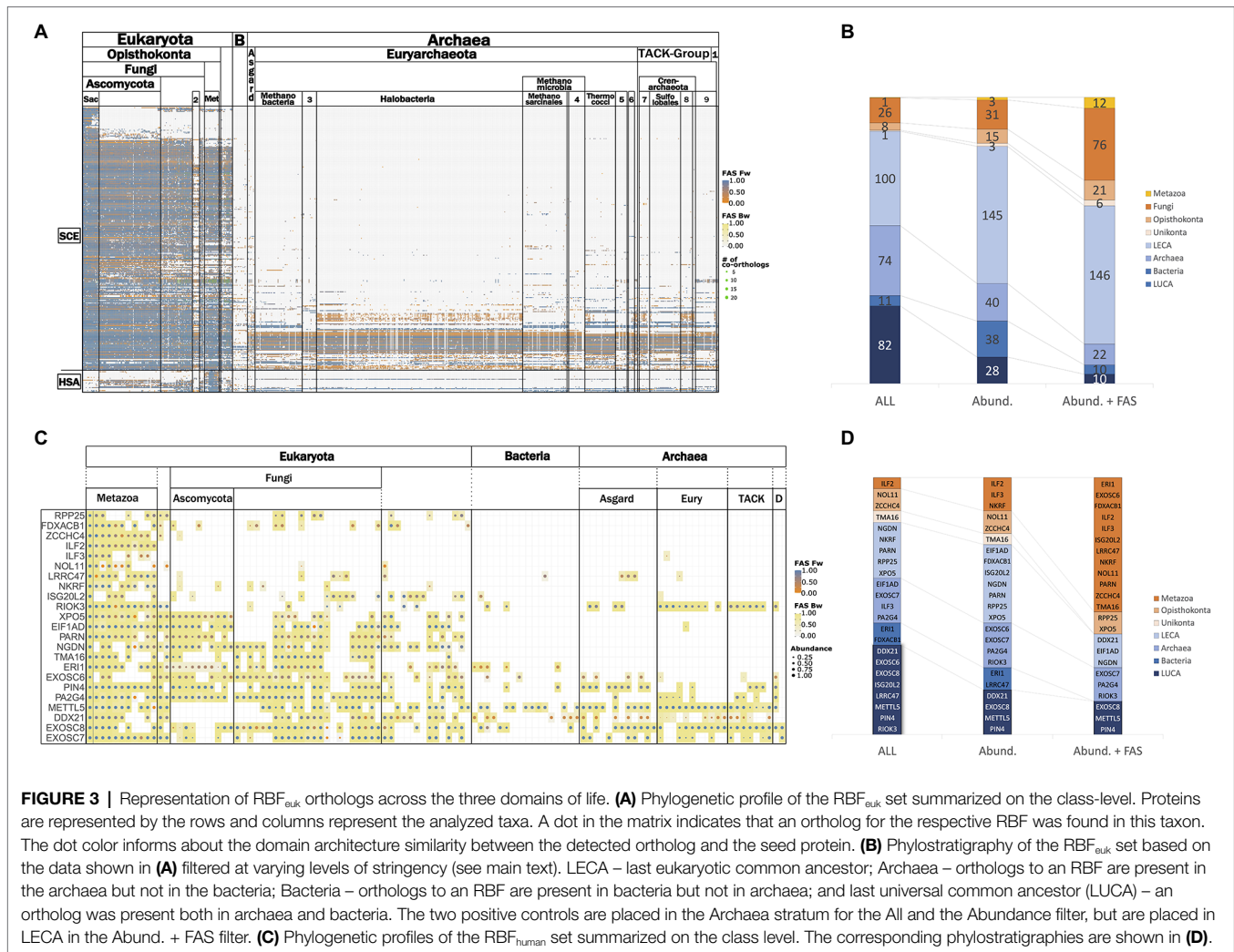
**FIGURE 2 |** The PhyloRBF platform. **(A)** Data overview; **(B)** Custom selection of genes and proteins for display; **(C)** Detail information for seed-ortholog pairs; and **(D)** Domain architecture view.

assignments and gene sets contaminated with sequences from other species (Steinegger and Salzberg, 2020). We therefore introduced two filters to reconstruct the phylostrata with two increasing levels of confidence (see **Supplementary Figure S3** for a visualization of the filtering effect). An abundance filter was established, building on the observation that secondary losses of RBFs are rare. It is, thus, expected that an RBF predicted to be present in the LCA of a systematic group, should be detected in all, or at least the majority, of its descendants. In turn, orthologs that show low prevalence in a group are likely spurious or represent contaminations that can be disregarded. For the second filter, we additionally propose that functionally equivalent orthologs participating in ribosome biogenesis in different organisms should have similar domain architectures (Koestler et al., 2010). We, thus, applied the abundance filter on the phylogenetic profiles where we kept only orthologs with a domain architecture that is similar to that of the seed protein ( $\Delta\text{FAS} < 0.25$ ). The resulting phylostratigraphies after application of the filters are shown in **Figure 3B**, and the corresponding assignment of the RBFs to the individual phylostrata is provided as **Supplementary Table S5**. As expected, the proportion of RBFs

assigned to evolutionarily younger phylostrata increases with filtering strength, and even the two positive controls were assigned only to LECA in the very stringent combined abundance and FAS filter. The results for the human RBFs are shown in **Figures 3C,D**. We note, however, that an *ad hoc* specification of the appropriate filtering criteria is difficult. The phylostratigraphies shown in **Figure 3** should be interpreted such that they allow dynamic analysis strategies. The most stringent filtering serves to reconstruct the evolutionary scaffold of ribosome biogenesis across the tree of life at high confidence. This scaffold can then be successively extended, where appropriate, with results obtained only with the more permissive filters. This clearly indicates where a focus on subsequent data curation and/or experimental validation should be placed.

### Absence of Human RBFs in the Fungal Lineage – Functional Plasticity or Methodological Artefact?

**Figure 3A** reveals that evidence of lineage-specific losses of RBFs within the eukaryotes is rare, with the notable exception of the microsporidia. This makes the existence of 17 human RBFs that can be traced back at least to LECA, but either



lack an ortholog in yeast, or their ortholog has diverged in function, an intriguing observation. Analyses of these proteins over, in the scope of this study, the moderate evolutionary distances between human and yeast serves two important needs. They can highlight that even evolutionarily old mechanisms of eukaryotic ribosome biogenesis can change on individual lineages. Probably even more important, they can shed light on methodological issues that become relevant when extending the analysis to the substantially more distantly related archaea.

METTL5 and PIN4 (Parvulin 14) represent two particularly prominent examples of gene loss on the evolutionary lineage leading to *S. cerevisiae*. Both proteins are assigned to the evolutionarily oldest RBF stratum under all filtering conditions (cf. Figures 3C,D). Within fungi, METTL5 is confined to the early branching lineages with only three putative orthologs in the Basidiomycota. METTL5 is an RNA methyltransferase that mediates N<sup>6</sup>-methylation of adenosine 1832 in the 18S rRNA in humans (van Tran et al., 2019), and a similar function has been described for METTL5 in *Haloflex vulcanii*, where it installs m<sup>6</sup>A1432 in the 16S rRNA (Kowalak et al., 2000; Grosjean et al., 2008). Yeast, in turn, has no m<sup>6</sup>A present at

the corresponding position of the 18S rRNA (Taoka et al., 2016; Sergiev et al., 2018). The loss of METTL5 in the course of fungal diversification is, thus, an evolutionarily unique event that changed a long-standing event in ribosome biogenesis. PIN4, a peptidyl-prolyl cis-trans isomerase required for pre-rRNA processing (Fujiyama-Nakamura et al., 2009), displays a different timing of gene loss. Like METTL5, PIN4 is also present in the archaea (Figures 3C,D), which is in line with previous findings (Jaremko et al., 2011; Hoppstock et al., 2016). In contrast to METTL5, PIN4 is prevalent in fungi. A sequence comparison between fungal and animal PIN4 orthologs indicates the presence of the sequence motif relevant for the association of this protein with pre-ribosomal complexes (Supplementary Figure S4). The loss of PIN4 immediately predates the diversification of the Saccharomycetales. The functional consequences of this loss are, to our knowledge, not yet explored. It will be interesting to see whether yeast and its close relatives utilize a non-homologous protein to functionally replace PIN4, whether its action is not necessary in the context of yeast ribosome biogenesis, or whether an alternative mechanism exists to compensate this loss of function.

## Domain Architecture Changes Indicate Lineage-Specific Loss of RBF Function

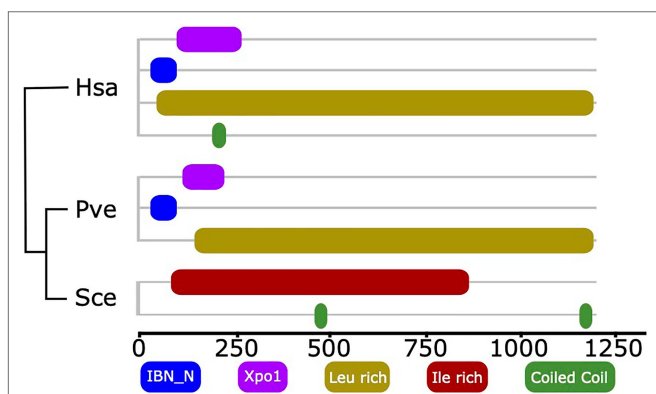
The example of XPO5 reveals that events subtler than gene loss also need to be considered when tracing factors involved in ribosome biogenesis across the organismic diversity. XPO5 is a human export factor that shuttles a diverse set of cargos, most prominently non-coding RNAs, such as tRNAs and pre-miRNA, but also ribosomal subunit precursors out of the nucleus (Bohnsack et al., 2002; Calado et al., 2002; Yi et al., 2003; Bohnsack et al., 2004; Lund et al., 2004; Wild et al., 2010; Leisegang et al., 2012). The corresponding ortholog in yeast serves also as an export factor, but with a more restricted set of cargos that does not include ribosomal subunits (Moy and Silver, 1999; Stage-Zimmermann et al., 2000; Sloan et al., 2016). Notably, the domain architectures of human XPO5 and its ortholog in yeast, Msn5, are substantially different (Figure 4). Two N-terminal Pfam domains, Xpo1 (Pfam ID: PF08389) and IBN\_N (Pfam ID: PF03810), that are characteristic for human XPO5 and also other proteins involved in nuclear import and export, e.g., yeast Xpo1 (P30822), are missing in yeast Msn5. Notably, we find the human domain architecture for XPO5 reflected in orthologs from early branching fungal lineages, suggesting that here the contribution of this protein to ribosome biogenesis could still be maintained. Although experimental proof that the apparent loss of the two domains causes the shift in the functional spectrum of Msn5 has to be delivered, this result suggests that changes in domain architecture can highlight RBF orthologs that may have altered their function.

## Evolutionary Traceability of RBFs

During integration of the RBF<sub>yeast</sub> and RBF<sub>human</sub> sets, no human protein was considered that has a yeast ortholog also annotated as an RBF. Interestingly, we noted that individual human factors, in particular, six of the 10 protein components of the RNase MRP complex, were initially retained because no yeast ortholog

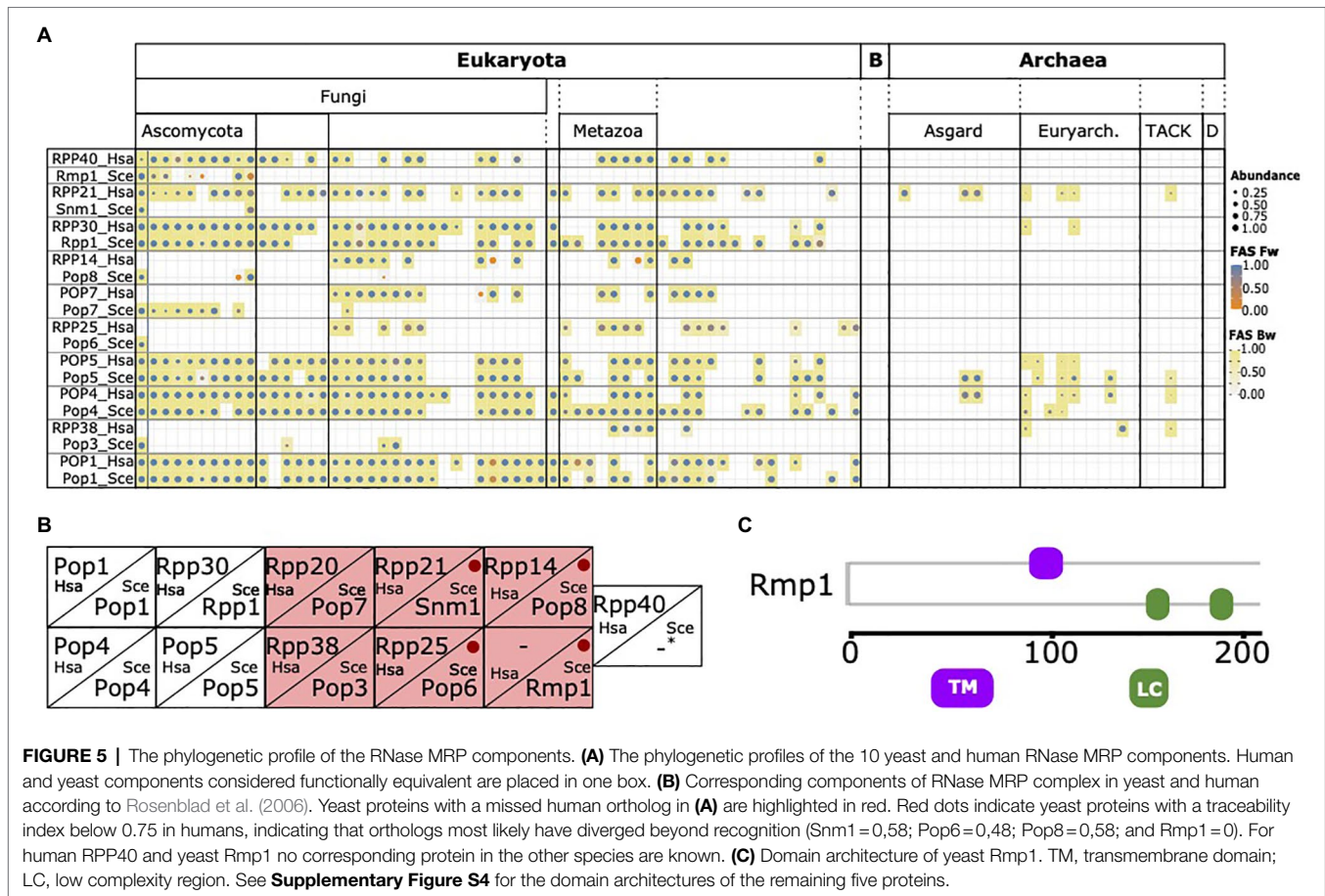
could be detected (Figure 5A). While this is consistent with the notion that yeast has substantially modified its RNase MRP complex because orthologs of many of its components are confined to yeast and its close relatives (Ebersberger et al., 2014), it is at odds with experimental evidence that functional equivalents exist for these factors in yeast and the function of the RNase MRP complex in pre-rRNA processing is conserved between yeast and humans (Chu et al., 1994; Lygerou et al., 1996; Rosenblad et al., 2006; Goldfarb and Cech, 2017; Figure 5B). To resolve this contradiction, we first confirmed that the absence of a yeast ortholog is independently supported by the InParanoid database (Ostlund et al., 2010), one of the most sensitive and specific public databases for ortholog assignments between pairs of species (Altenhoff et al., 2016). Thus, the missing of orthologs for these proteins in the respective species seems inherent to the ortholog search itself, and is not dependent on the ortholog search tool. Investigating the six proteins in greater detail revealed that all are short with lengths between 150 and 200 amino acids, and Rmp1 and Pop8 are additionally devoid of any Pfam domain (Figure 5C; Supplementary Figure S5). Both characteristics in combination are indicative for proteins that lose a significant sequence similarity to their orthologs already over small evolutionary distances (Jain et al., 2019). Indeed, the evolutionary traceability indices for these proteins in humans (see Materials and Methods) indicate that human orthologs of four proteins likely have diverged beyond recognition (cf. Figure 3B), and thus escape detection in large-scale phylogenetic profiling studies. In conclusion, the discrepancy between the *in silico* approaches to assess phylogenetic distribution and evolutionary age of the yeast RNase MRP components, and the experimental evidence, can be explained in at least four out of six cases by the limited sensitivity of the ortholog search.

To see whether other proteins in the RBF<sub>euk</sub> set may suffer from the same limitations, in particular, when extending the scope of the ortholog search to the archaeal domain, we subsequently computed the traceability indices across the full range of taxa considered here (Figure 6A). This identified 42 proteins that evolve at rates likely to hinder detection of orthologs in evolutionarily more distant lineages (Figure 6B). For these proteins, it is important to consider that the phylogenetic profiles may underestimate the taxonomic range in which orthologs are present, and hence will result in underestimated evolutionary ages. Among them, we find a further factor, Arx1, which associates with late LSU particles in the nucleus and facilitates their export (Bradatsch et al., 2007; Hung et al., 2008). PA2G4, which is a member of the RBF<sub>human</sub> set, was initially suggested to be its functional equivalent in humans, but we failed to establish orthology relationships between Arx1 and PA2G4 (Supplementary Figure S6). Meanwhile, evidences accumulated that PA2G4 differs in regions that are implicated in nucleoporin interaction in Arx1 (Squatrito et al., 2004; Bradatsch et al., 2007; Wild et al., 2010; Bhaskar et al., 2021). While this suggests a functional diversification between the two proteins, their precise evolutionary relationships remain unclear. A phylogenetic analysis revealed gene duplication paired with a substantial acceleration of the evolutionary rate on the



**FIGURE 4 |** Domain architecture comparison of Xpo5 between human and fungal orthologs. The two Pfam domains linked to import/export activity of human Xpo5, IBN\_N (Pfam ID: PF03810) and Xpo1 (Pfam ID: PF08389) are still present in an early branching fungus and were secondarily lost in yeast. The phylogenetic relationships of the three taxa are indicated by the tree. Hsa, *Homo sapiens*; Pve, *Podila verticillata* (Mucoromycota); and Sce, *Saccharomyces cerevisiae*.





**FIGURE 5 |** The phylogenetic profile of the RNase MRP components. **(A)** The phylogenetic profiles of the 10 yeast and human RNase MRP components. Human and yeast components considered functionally equivalent are placed in one box. **(B)** Corresponding components of RNase MRP complex in yeast and human according to Rosenblad et al. (2006). Yeast proteins with a missed human ortholog in **(A)** are highlighted in red. Red dots indicate yeast proteins with a traceability index below 0.75 in humans, indicating that orthologs most likely have diverged beyond recognition (Snm1 = 0,58; Pop6 = 0,48; Pop8 = 0,58; and Rmp1 = 0). For human RPP40 and yeast Rmp1 no corresponding protein in the other species are known. **(C)** Domain architecture of yeast Rmp1. TM, transmembrane domain; LC, low complexity region. See **Supplementary Figure S4** for the domain architectures of the remaining five proteins.

Arx1 lineage of Saccharomycotina and Taphrinomycotina (**Supplementary Figure S7**). This lends support to the hypothesis that similar to the components of the RNase MRP complex, the orthology between PA2G4 and Arx1 was overlooked thus far.

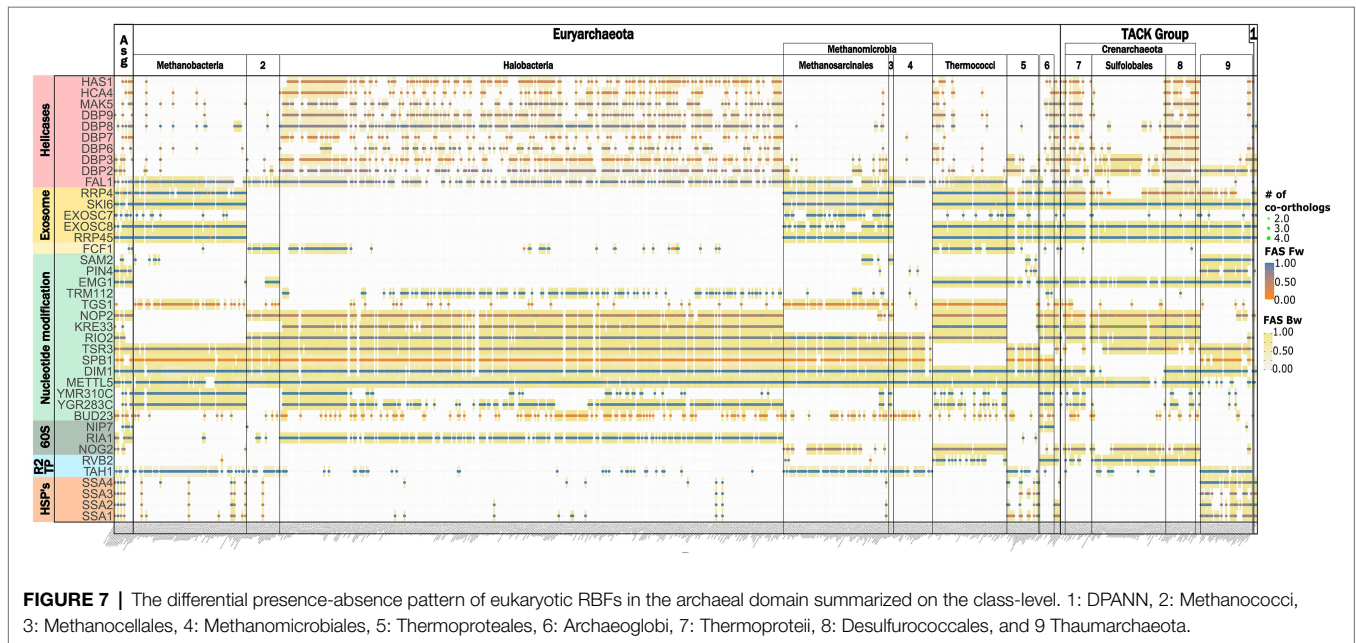
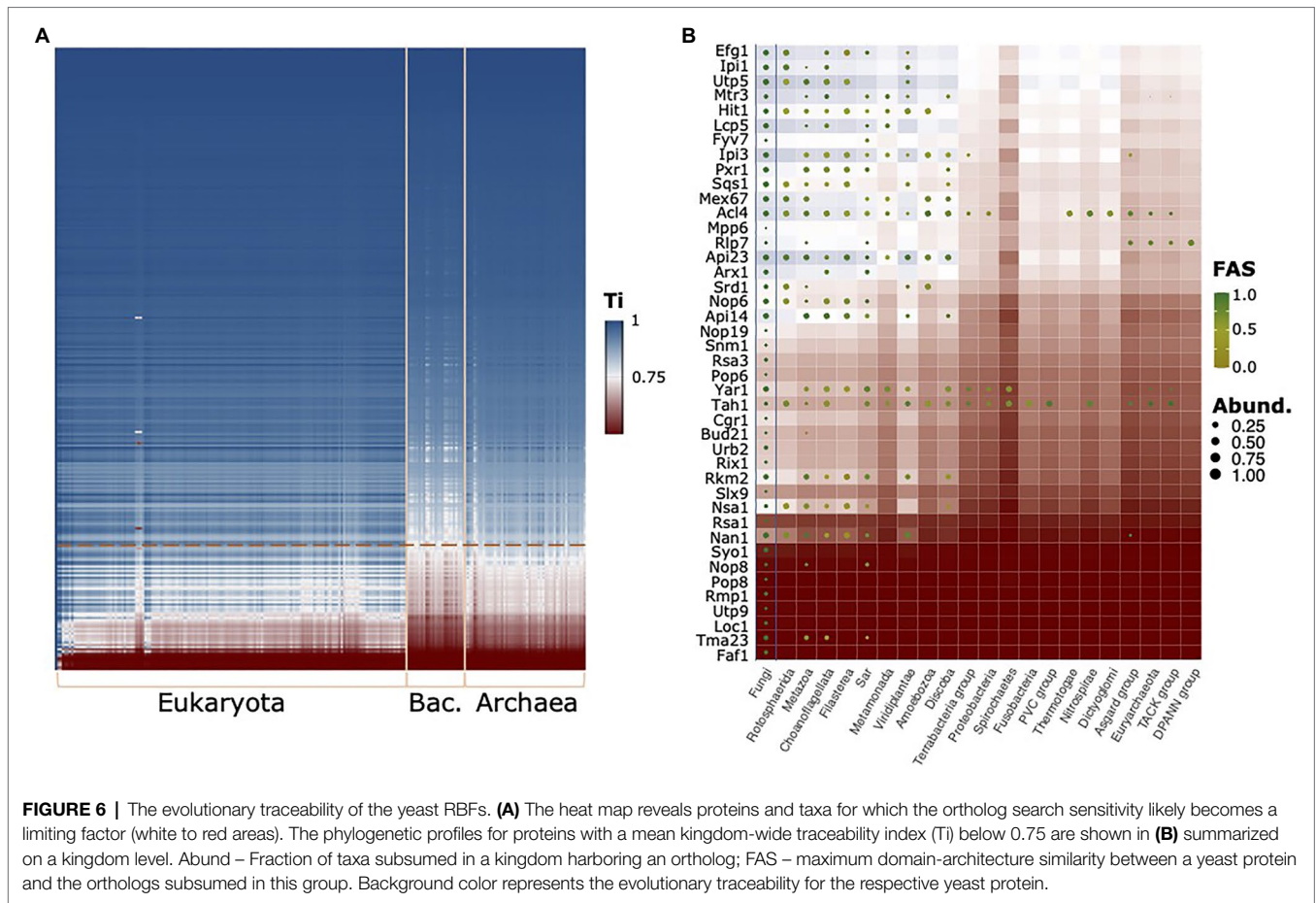
## Eukaryotic RBFs in the Archaeal Domain

The analysis thus far has set the methodological stage for projecting concepts of eukaryotic ribosome biogenesis into the archaeal domain. Subsequently, we focused on the representation of the RBF<sub>euk</sub> in archaea. Sampling more than 700 taxa covering the full known archaeal diversity, including the Asgard group that are proposed as the closest relatives to the eukaryotes (Zaremba-Niedzwiedzka et al., 2017), constitutes an unprecedented basis for this analysis. Most importantly, it provides the first comprehensive overview of which factors are represented by orthologs in the archaeal domain, together with information about their lineage-specific prevalence (**Supplementary Figure S8**). For 156 factors in the RBF<sub>euk</sub> set, we found in the unfiltered data at least one ortholog in the archaeal domain (cf. **Figure 3B**; **Supplementary Figure S8**). The number of orthologs per species varies considerably (**Supplementary Table S5**), and *Candidatus Prometheoarchaeum syntrophicum* (Asgard group) harbors the highest number of RBF orthologs (67) across the sampled taxa. Applying an abundance filter (see above) to select only proteins consistently

seen in the archaea reduces the number to 17 factors that are represented by an ortholog in more than 85% of the investigated taxa (**Supplementary Table S7**), among the two positive controls. This number reduces to only four when setting the inclusion threshold at 90%. These nearly ubiquitously represented proteins comprise two RNA methyltransferases, METTL5, and Nop1/Fibrillarin, the pseudouridine synthase Pus7 that catalyzes pseudouridylation within the 5S rRNA, and one of our two positive controls, the largest subunit of the RNA polymerase II, Rpo21. The presence of many of these proteins, and particular of these latter factors, in archaea is already well acknowledged (e.g., Jaremko et al., 2011; Ebersberger et al., 2014; Londei and Ferreira-Cerca, 2021). Their prevalence in our data set suggests that insights gained from detailed experimental analyses in individual taxa can, in some cases, be projected to the entire archaeal domain. Factors that are present only in a subset of the taxa, in turn, can provide insights into lineage-specific differences of presumably related to ribosome biogenesis within the archaea, an area that is still considerably uncharted.

## Lineage-Specific Presence of Eukaryotic RBFs in Archaea

From the archaeal section of **Figure 3A**, we selected the subset of RBF<sub>euk</sub> that is confined to individual archaeal clades and grouped these proteins into seven functional



categories (Figure 7). This revealed a loss of several components of the RNA exosome, a major RNA decay machinery (Sloan et al., 2012), once in Halobacteria and once in

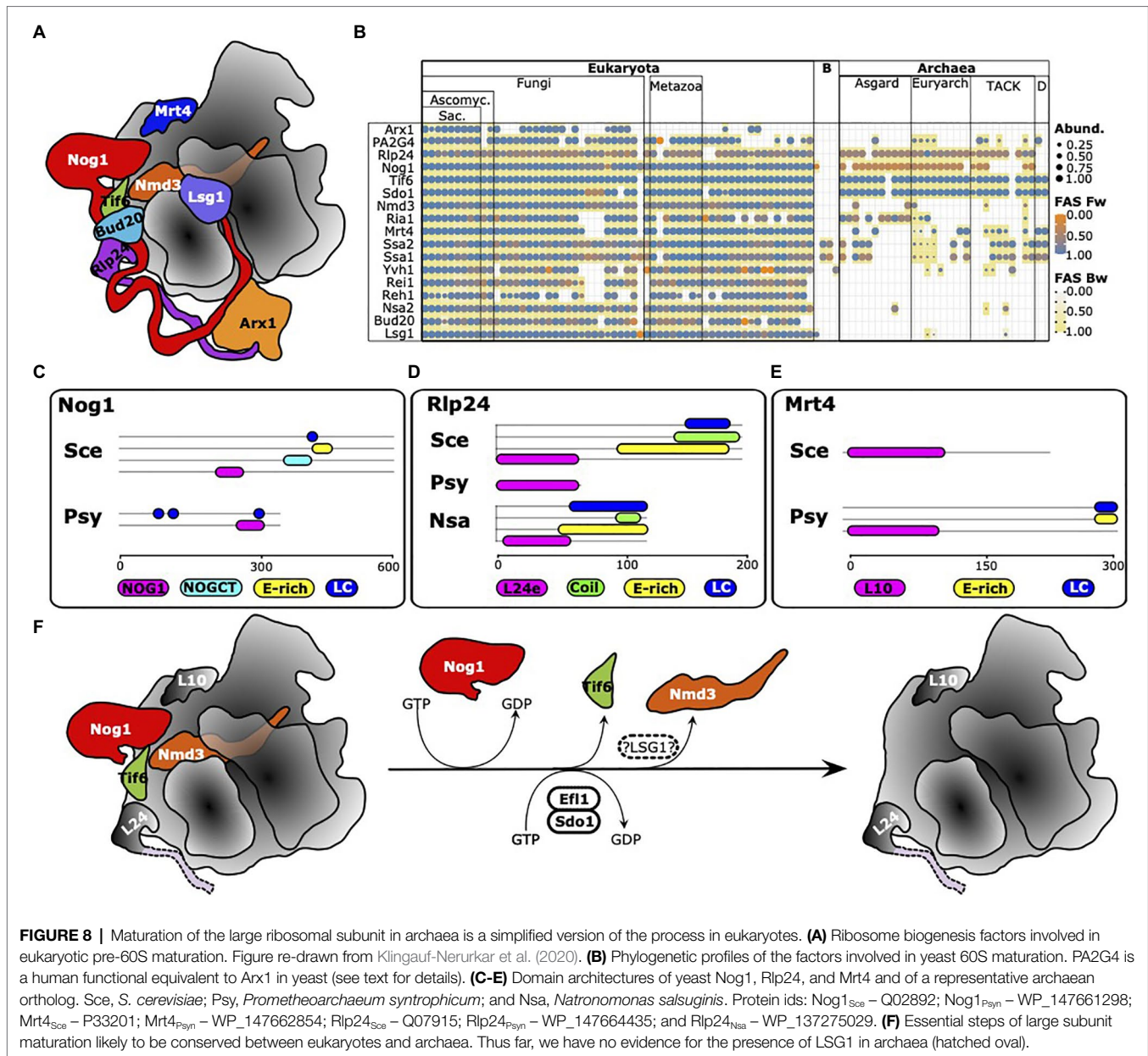
Methanomicrobiales. Two small-scale analyses hinted toward the possible absence of the exosome complex in these clades (Koonin et al., 2001; Phung et al., 2020). The concerted absence

of the exosome components across all members of the two clades seen here, provides substantial evidence that the RNA exosome has indeed been lost on these lineages. A similarly prominent signal is seen for the RNA helicases. However, we note that the domain architecture similarities between the RBFs and their archaeal orthologs are typically small for these proteins (Figure 7; Supplementary Figure S9). As RNA helicases are involved in a plethora of different processes in a cell, it is thus too early to conclude on the existence of these RBFs in the archaeal domain. The other functional categories do not share a consistent phylogenetic profile. The HSP70-family chaperones SSA1-4 are no exception here, as they all share the same archaeal proteins as orthologs, indicating that gene duplications on the eukaryote lineage gave rise to these four distinct RBFs. This directs further attention to the presence/absence pattern of individual proteins, most of them involved in nucleotide modification. Here, the missing of orthologs to Dim1, a dimethyladenosine transferase involved in rRNA modification in the Sulfolobales is, at first glance, an intriguing finding. Dim1 is a protein that is otherwise almost fully conserved across taxon collection. We detected an ortholog in 99% of the eukaryotic taxa and it is present in all but one (*Aquifex aeolicus*) of the bacterial species. The domain architectures of both eukaryotic and bacterial/archaeal orthologs are virtually identical to that of yeast Dim1 (Supplementary Figure S10), and the traceability of Dim1 is one across all taxa. Despite this convincing indication for a gene loss, the finding is at odds with the detection of an N<sup>6</sup>-dimethyladenosine in the 16S rRNA (Noon et al., 1998) and a recent characterization of a Dim1-like enzyme in *Sulfolobus acidocaldarius* (Knuppel et al., 2021). In the first step, we confirmed that a BlastP search with the alleged Dim1 from *S. acidocaldarius* revealed no hit in yeast (not shown). To increase the sensitivity of the ortholog search, we used a “stepping-stone” approach. In brief, we performed a second search for Dim1 orthologs, this time seeding the search with the Dim1 of *S. acidocaldarius*. This served to reduce the evolutionary distance between seed protein and possible orthologs in the archaea (Martin-Duran et al., 2017). This search identified orthologs throughout the archaeal domain and sporadically also in the eukaryotes (Supplementary Figure S11). In all investigated cases, these orthologs were the same as those obtained with yeast Dim1 as a seed sequence. Based on their overlapping phylogenetic profiles, we conclude that yeast Dim1 and the protein in *S. acidocaldarius* are indeed orthologs, although they share no significant sequence similarity. A further comparison revealed that Dim1 in Sulfolobales is only about 200 aa in length, while Dim1 in other taxa has a length of around 300 amino acids (Supplementary Figure S12). In summary, one of the evolutionarily most conserved proteins known to date has been modified in one archaeal lineage comprising largely extremophiles, such that the orthologs have diverged, from a eukaryotic viewpoint, beyond recognition. It will be interesting to see the precise functional consequences that are likely accompanied with this change. In contrast to the spurious absence of Dim1, there is strong evidence for a lineage specific loss of the kinase/ATPase Rio2, the RNA helicase/acetyltransferase Kre33, and

the RNA methyltransferase Emg1 within the archaea. For these proteins, the stepping-stone approach provided no indication of overlooked orthologs (Supplementary Figure S13), and we conclude that the gaps in their phylogenetic profiles represent the genuine absence of these proteins in the respective taxa.

## Large Subunit Processing in Archaea

The analyses thus far resulted in a high-resolution overview of the prevalence and distribution of eukaryotic RBFs in the archaeal domain. However, they also revealed that a face-value interpretation of phylogenetic profiles can be misleading with the erroneous assumption of factor (and function) absence and presence. Careful curation of the profiles *via* a comparison of domain architectures, the consideration of the evolutionary traceability and case-by-case, more sensitive analyses *via* the stepping-stone approach, or even *via* the search for proteins harboring only a characteristic Pfam domain (see Jain et al., 2019), can substantially increase the reliability of the conclusions drawn. Such comprehensive analyses are challenging to apply on large data sets and are performed best in a context of functional sub-clusters. Here, we focused on late steps during maturation of the 60S subunit regulated by the GTPase Nog1 (Klingauf-Nerurkar et al., 2020) as a showcase example (Figure 8A). The phylogenetic profiles of 16 RBFs involved are shown in Figure 8B and reveal an intriguing pattern. Only two factors are consistently present in the archaea with domain architecture similarities that leave little doubt of functional equivalence to the yeast proteins, Tif6 (Benelli et al., 2009) and Sdo1. These proteins functionally interact, where Sdo1, in combination with Ria1 (see below), promotes removal of Tif6 from pre-60S particles. Orthologs to Nog1 are found in all major archaeal lineages, except in the two representatives of the DPANN group. However, their domain architecture similarity scores are low (Supplementary Figure S14), which indicates either a spurious ortholog assignment or a change in protein function during the evolution of Nog1. In strong support of the latter hypothesis, we first confirmed that in fungi, animals, and archaea, typically only a single gene in the gene set of an organism encodes a protein carrying a Nog1 Pfam domain (Pfam Id: PF06858; Supplementary Figure S15). A subsequent comparison of the domain architectures revealed that archaeal Nog1 is only half the size of its eukaryotic counterpart (Figure 8C). While the N-terminal part harboring the GTPase activity is present, archaeal Nog1 lacks the C-terminal portion that mediates the interaction with Arx1. The situation is similar for Rlp24, which was suggested to act as placeholder for the ribosomal protein eL24 during the maturation process (Warner, 2015; Figures 8B,D), but with two key differences. Almost all archaeal orthologs of Rlp24 are approximately 70 amino acids in length and harbor only the L24e Pfam domain (PF01246). The C-terminal tail that is essential for recruiting Drg1 is missing. But surprisingly, we find individual taxa within the Halobacteria and the Thermoplasmata whose Rlp24 ortholog is slightly longer. Their domain architectures match with that of yeast Rlp24, and they also appear in possession of the C-terminal tail of the protein found in yeast despite the absence of Drg1 orthologs in these clades. The second difference is the functional annotation of the archaeal orthologs. In all cases, we identified the ribosomal protein L24 as the ortholog



to Rlp24, and this is the only protein encoded in archaeal genomes that harbors the L24e Pfam domain, again leaving no scope for overlooked orthologs (Supplementary Figure S16). This is perfectly in line with a scenario where eukaryotic Rlp24 and eL24 emerged by a gene duplication of an ancient ribosomal protein. One copy retained the ancestral function as a ribosomal protein, whereas the other copy evolved into an RBF. Interestingly, the same scenario applies to Mrt4 (Figures 3A,E), which is known to function as a placeholder for the acidic 60S ribosomal protein uL10 (Warner, 2015). The archaeal orthologs are annotated as Rpl10, and again each of the investigated archaeal taxa harbors exactly one protein with the L10e domain (Supplementary Figure S17). Notably, the archaeal orthologs can be longer than yeast Mrt4. They often harbor an acidic (E-rich) stretch and a low complexity region at the C-terminus

(Supplementary Figure S18), similar to what is seen for yeast Rlp24 and Nog1 (Figures 8C,D). Judging by the domain architecture, they seem to resemble the eukaryotic ribosomal protein P0 rather than Mrt4, indicating that the architecture of Mrt4 is the evolutionarily-derived form (Supplementary Figure S19). Note, the considerably sparse representation of Mrt4 orthologs must be attributed to methodological issues (see Dim1 above), as all archaeal gene sets we checked contained exactly one protein with the L10e Pfam domain (*cf.* Supplementary Figure S17). Validating the presence of the other proteins involved in late 60S maturation in archaea provided unambiguous evidence only for Nmd3 (Supplementary Figure S20), a protein that, in eukaryotes, contributes to pre-60S nuclear export by acting as an adaptor for the nuclear transport receptor Crm1 (Ho et al., 2000).

Efl1/Ria1, which interacts with Sdo1 in the release of Tif6, identifies the elongation factor EF2 in archaea as an ortholog. Similar to Mrt4 and Rlp24, also eukaryotic EF2 and Efl1/Ria1 likely emerged by a gene duplication in LECA (see **Supplementary Figures S21, S22** for details). All other proteins appear absent in the archaea. In case an ortholog was identified in our search, it turned out to resemble a remote paralog of the RBF with an evolutionarily highly conserved function, e.g., MAP2 in the case of PA2G4 (*cf.* **Supplementary Figure S6**). A single factor, however, remains elusive, Lsg1. Its GTPase activity is required for the release of Nmd3 as one of the last steps of 60S maturation (Hedges et al., 2005). Searches at higher sensitivity identified potential homologs, but neither sequence similarity nor domain architecture conservation sufficed to conclude on the presence of this protein in the archaea. Taken together, these data suggest that many late steps of large subunit maturation in archaea follow the same principles as in eukaryotes. However, instead of using the placeholders Mrt4 and Rlp24, archaea most likely directly install the ribosomal proteins RPP0 and L24. As a consequence, the downstream machinery required for the successive release of Rlp24 and Mrt4 by Drg1 (Pertschy et al., 2007) and Yvh1 (Kemmler et al., 2009; Lo et al., 2009), and the incorporation of corresponding ribosomal proteins, is not necessary. Likewise, the lack of an ortholog to Arx1 in the archaea is in line with the findings that the C-terminus of Nog1, which is essential for interaction with Arx1, is absent from the archaeal proteins. Interestingly, Mrt4 also represents one of the factors that is implicated in pre-ribosome surveillance in eukaryotes (reviewed in Karbstein, 2013; Klinge and Woolford, 2019), as it was reported to block ribosomal stalk assembly until it is released from pre-ribosomal complexes (Kemmler et al., 2009; Lo et al., 2009; Rodriguez-Mateos et al., 2009). Its absence in archaea might therefore indicate a simplification of the process in this clade. In essence, it appears that the late steps of large subunit maturation in contemporary archaea largely resemble the primordial pathway in the LCA of archaea and eukaryotes. This basic functionality was later extended in the eukaryotic lineage to facilitate nuclear export and to implement various quality control steps.

## DISCUSSION

Eukaryotic ribosome biogenesis is currently best understood in the yeast model system, *S. cerevisiae*, but knowledge on the human pathway is growing (Bohnsack and Bohnsack, 2019). However, the increasing availability of genome sequences of species from the remotest corners of cellular life now enables investigation of the conservation and plasticity of this pathway across the eukaryotic and the archaeal domains. Integrating these diverse efforts into a comprehensive view of how ribosome biogenesis is accomplished across the organismal diversity strongly benefits from a unifying data basis. This allows an interpretation of the outcomes of functional studies on individual RBFs or complexes contributing to ribosome biogenesis in the context of their evolutionary trajectory across the organismic diversity. Various public databases partially fulfil this requirement. For example, KEGG (Kanehisa et al., 2016) or REACTOME

(Fabregat et al., 2018) provide access to pathways and interaction networks of proteins involved in ribosome biogenesis together with a representation of pathway components in other taxa. Dedicated ortholog databases, such as InParanoid (Ostlund et al., 2010), OMA (Altenhoff et al., 2018), and orthoDB (Zdobnov et al., 2017) provide comprehensive collections of orthologous groups across the full proteomes of hundreds to thousands of species. Despite the wealth of information contained in these resources, they also have limitations. Neither KEGG nor REACTOME consider the full set of factors currently considered as RBFs. Furthermore, the ortholog databases have no focus on a dedicated pathway, they provide no direct access to phylogenetic profiles for individual or groups of proteins, and they do not facilitate a comparison of domain architectures or a scoring of architecture similarities. Last but not least, it is almost impossible to extend the analysis to custom factors or species. Here, we have combined data from various resources to compile a manually curated, non-redundant set of 301 eukaryotic RBFs (*cf.* **Figure 1**). Their domain architecture-aware phylogenetic profiles across more than 900 taxa are a first step to close these gaps in the currently available resources. The data can be interactively explored and analyzed *via* a light-weight web-portal, or for more in-depth analyses, downloaded and analyzed offline. They can be visualized as a whole to explore the concerted evolutionary behavior of groups of proteins (*cf.* **Figure 7**), or subsets of taxa and proteins can be extracted for in depth analyses down to exploring the lineage-specific fate of individual protein domains (Roustan et al., 2016). This allows to discern factors whose functions likely have changed between orthologs from those where no such indication exists (Jiang et al., 2020). Eventually, the data basis is flexible. Our approach makes it straightforward to extend these profiles with further taxa or factors in custom fashion. We therefore hope that the broad community interested in ribosome biogenesis will benefit from this data and it could serve as template for further pathway-specific analyses.

## Secondary Modification of Ribosome Biogenesis in Fungi

Large-scale screens in humans paired with targeted identification of proteinaceous factors *via* the characterization of ribosomal complexes have the potential to complement the yeast perspective on eukaryotic ribosome biogenesis (Wild et al., 2010; Tafforeau et al., 2013; Badertscher et al., 2015; Farley-Barnes et al., 2018). Interestingly, the integration of this data revealed only 23 human factors that lack an ortholog in yeast. We note that this number is most likely an underestimate. Manual curation determining whether a protein was already shown to associate with pre-ribosomal particles/factors or has a known direct function in the pathway was reserved to human proteins at least as old as LECA. Thus, many candidates in the screens have simply not been analyzed in enough detail yet to know if they fulfil these criteria. Still, 17 evolutionarily old RBFs are missing in yeast and in part also in other fungal taxa. This indicates that traditional concepts of ribosome biogenesis have been modified during fungal diversification, despite the overall evolutionary conservation of this process (Wild et al., 2010;

Tafforeau et al., 2013; Badertscher et al., 2015; Farley-Barnes et al., 2018; Bohnsack and Bohnsack, 2019). With METTL5 and XPO5, we have highlighted one prominent example each for loss-of-function and alteration of function, respectively. Specifically, we could date the loss of METTL5 most likely to the LCA<sub>Dikarya</sub>. Thus, the absence of m<sup>6</sup>A that has been described for yeast most likely applies to all dikarya. We could correlate a change in the domain architecture of Msn5, the yeast ortholog of XPO5, with an expansion of its functional repertoire. On closer inspection, we observed a recurrent change in domain architecture within the ascomycetes for this protein, and it will be interesting to see to what extent this affects the cargo(s) Msn5/XPO5 can transport. Few other human RBFs are absent in fungi, and the effect of their alleged loss in fungi should be investigated in greater detail and related with their functions in other eukaryotes.

### Sensitivity of the Ortholog Search as a Limiting Factor

The integration of the human data has opened up a further aspect that has so far received little attention in the interpretation of presence/absence patterns in phylogenetic profiles. A number of human factors for which functionally equivalent proteins exist in yeast, appear to lack yeast orthologs. Taken at face value, this must be interpreted as evidence for non-orthologous functional replacements (Koonin et al., 1996). However, for quickly evolving proteins, the grey zone where orthologs are no more similar than it is expected by chance (Rost, 1999; Jain et al., 2019) is within reach, even for the evolutionary distances between humans and yeast, and even more so when extending analyses into the archaeal domain. Here, we have provided evidence that for 42 yeast RBFs, the sensitivity of the ortholog search can become a limiting factor. Among these we found many proteins whose functional equivalent in humans is not identified as an ortholog, eventually reconciling the findings from targeted functional and the large-scale evolutionary approaches. Increasing the search sensitivity, e.g., by switching to unidirectional profile-based searches scanning for proteins sharing the same Pfam domains as the target protein, is an obvious solution. This, however, comes at the costs of a substantially reduced specificity making careful downstream validation of the findings necessary (Jain et al., 2019). However, the example of Dim1 shows that traceability indices can also be positively misleading. The underlying simulation approach is based on the assumption that evolutionary rates and constraints for a protein do not change over time. Our results indicate that this does not apply to Dim1 in the Sulfolobales. The lineage-specific change of both rate and mode of sequence evolution for this otherwise ubiquitously conserved protein, which causes the orthologs to be missed, is intriguing. Whether this is a consequence of an altered function of Dim1 and/or a change of the selective constraints resting on the corresponding gene remains to be determined. While experimental evidence exists for the presence of Dim1 in the Sulfolobales (Knuppel et al., 2021); this is not necessarily the case for other proteins. In such cases, the stepping-stone approach that serves to reduce the evolutionary distance over which orthologs need to be identified can be used to identify even evolutionarily diverged orthologs.

### Eukaryotic Ribosome Biogenesis Factors in the Archaeal Domain

Catalogues of eukaryotic ribosome biogenesis factors that are also present in the archaeal domain have been compiled before (e.g., Ebersberger et al., 2014). These earlier studies sufficed to identify the presence/absence of factors in individual archaeal species. Yet, the limited taxon sampling did not allow to differentiate between signal and noise with respect to varying abundance patterns across the analyzed taxa. Along the same line, the co-occurrence of eukaryotic RBFs in individual taxa, an obvious pre-requisite for but also an indication of their functional interaction, could not be exhaustively tested. Here, we could show that, in contrast to the situation in the eukaryotes, very few factors are consistently found throughout the archaeal domain. This can be an effect of limited search sensitivity (see above). However, even the application of the stepping-stone approach indicated the lineage-specific absence of otherwise essential factors, such as Kre33, Rio2, and Emg1. These findings can serve as starting points to elucidate whether ribosome biogenesis in the archaea is more plastic than is the case for the eukaryotes, whether individual factors are not essential (e.g., Knuppel et al., 2021), or whether they convey their activity to a different pathway.

Within our taxon sampling, we considered also all sequenced representatives of the Asgard group, which have been proposed to be the closest relatives to eukaryotes (Zaremba-Niedzwiedzka et al., 2017). This suggested close relationship is not reflected by the presence of any eukaryotic RBF exclusively in members of this clade. However, we note that among all analyzed archaeal taxa, the recently sequenced *Candidatus Prometheoarchaeum syntrophicum* (Asgard group) harbors the largest number of orthologs to eukaryotic RBFs (**Supplementary Table S9**). Whether this is co-incidence or if this indicates that ribosome biogenesis in this species is indeed more similar to that of eukaryotes remains to be determined.

### Large Subunit Maturation in Archaea

Over the past years, the structural characterization of pre-ribosomal complexes together with studies on the precise functions of the proteins involved have elucidated the roles and interplay of many RBFs in the stepwise formation and maturation of the ribosomal subunits (e.g., Saveanu et al., 2003; Woolford and Baserga, 2013; Henras et al., 2015; Wu et al., 2016; Bassler and Hurt, 2019; Bohnsack and Bohnsack, 2019; Kargas et al., 2019; Zhou et al., 2019; Klingauf-Nerurkar et al., 2020; Liang et al., 2020). Clusters of RBFs provide an unprecedented opportunity to *in silico* “assemble” archaeal RBF orthologs into a comprehensive picture. This allows well-informed and testable predictions about differences and similarities of the corresponding processes in archaea. Projecting the steps in late maturation of the 60S subunit (Kargas et al., 2019; Zhou et al., 2019; Klingauf-Nerurkar et al., 2020) into the archaeal domain revealed that archaea follow, by and large, similar principles compared to eukaryotes. Yet this process is significantly more simplistic. The absence of C-terminal extensions in both archaeal Nog1, and L24, the archaeal ortholog

of both the eukaryotic RBF Rlp24 and the ribosomal protein eL24, indicates that the parts of the maturation pathway that require protein-protein interaction mediated *via* these extensions (Klingauf-Nerurkar et al., 2020) are missing in the archaea. Moreover, the ribosomal proteins uL10 and eL24 appear associated with the 50S ribosomal complex directly, without the need for previous recruitment of placeholder proteins. The apparent simplicity of this process probably contributes to the fact that the 50S subunit of various archaeal species can be spontaneously reconstituted *in vitro* with RBFs (Londei et al., 1986; Sanchez et al., 1990). The eukaryotic RBFs Mrt4 and Rlp24 that serve as placeholders for the RPPs uL10 and eL24, respectively, emerged by a duplication of the genes encoding the corresponding ribosomal proteins only in the LECA. Consequently, some factors that are required for steps leading up to the installation of these ribosomal proteins in eukaryotes were not identified in archaea and it is probably safe to consider them absent. In essence, archaea seem to bypass various assembly and surveillance steps during large subunit maturation that are characteristic for eukaryotes. Overall, this represents a showcase example of the evolution of complexity *via* the duplication and subsequent functional diversification of ancient genes (Ohno, 1970; Taylor and Raes, 2004) paired with a lineage-specific modification of evolutionary old genes.

Two aspects however remain puzzling. Firstly, the mechanism of release of the pre-60S biogenesis and export factor Nmd3, which is ubiquitously found in archaea, is elusive, as we did not find any evidence for an archaeal counterpart to the GTPase Lsg1, which releases Nmd3 from the 60S complex in yeast (Malyutin et al., 2017). While its eukaryotic function in pre-60S export will not be present in archaea due to the absence of a nucleus, archaeal Nmd3 might share its role in uL16 loading onto pre-ribosomes with its eukaryotic counterpart (Hedges et al., 2005; Hofer et al., 2007; Kargas et al., 2019; Zhou et al., 2019). Future experiments will have to elucidate the precise role of Nmd3 orthologs in the archaea and whether a different protein may serve as the functional equivalent to Lsg1. Secondly, our observation that the ribosomal protein L24 in archaea comes in two flavors is intriguing. The most widespread form is a short protein harboring exclusively the Ribosome\_L24e Pfam domain. Representatives of Halobacteria and of Thermoplasmata, however, possess a protein that resembles a miniaturized version of the yeast RBF Rlp24. Their L24 includes a C-terminal extension with exactly the same features as in the yeast protein, a coiled-coil domain, an E-rich stretch and a low-complexity region. These proteins have a second interesting characteristic. Eukaryotic Rlp24 share with archaeal but not with eukaryotic L24e the presence of four conserved cysteine residues at the N-terminus (Saveanu et al., 2003). Three of the four Cysteine residues have been mutated in the L24e with the eukaryote-like domain architecture (**Supplementary Figure S23**). So far, we have no evolutionary explanation for these observations and it will be interesting to investigate the functional role of these changes.

In summary, we established a common data base for integrating research on ribosome biogenesis in eukaryotes and archaea. We have highlighted the potential, as well as possible pitfalls,

in the interpretation of these data and have highlighted the cascade of *in silico* analyses that is necessary to faithfully reconstruct the evolutionary trajectory of individual or groups of RBFs from these data. Tracing eukaryotic RBFs into the archaeal domain revealed a surprisingly fragmented abundance pattern. We typically found highly conserved factors missing in individual lineages without any evidence for a methodological artefact. This passes the baton to the experimentalists to determine if, and to what extent, these factors are indeed involved in ribosome biogenesis. Using the late steps in large subunit maturation as an example, we provide evidence that archaea follow fundamentally similar but, at least in parts, highly simplified strategies to assemble their ribosomes. This underlines the role of archaea as potential model systems to elucidate general concepts of ribosome biogenesis and highlights evolutionary strategies to increase the fidelity of this process on the eukaryotic lineage.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: <https://applbio.biologie.uni-frankfurt.de/download/RibosomeBiogenesis>.

## AUTHOR CONTRIBUTIONS

IE conceived and planned the study and drafted the manuscript with help from MB. MTB and KB curated the data. MB, VT, and SS performed the analyses. IE, MB, MTB, and KB interpreted the results. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Research Funding Program Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz (LOEWE) of the State of Hessen, Research Center for Translational Biodiversity Genomics (TBG) to IE, the Deutsche Forschungsgemeinschaft (SFB1190 to MTB, SFB860 to KB, and EB-285-2/2 to IE), and the University Medical Centre Göttingen (to MTB and KB).

## ACKNOWLEDGMENTS

The authors wish to thank all researchers for making annotated genome sequences available to the public domain.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.739000/full#supplementary-material>

## REFERENCES

- Altenhoff, A. M., Boeckmann, B., Capella-Gutierrez, S., Dalquen, D. A., DeLuca, T., Forslund, K., et al. (2016). Standardized benchmarking in the quest for orthologs. *Nat. Methods* 13, 425–430. doi: 10.1038/nmeth.3830
- Altenhoff, A. M., Glover, N. M., Train, C. M., Kaleb, K., Warwick Vesztrocy, A., Dylus, D., et al. (2018). The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* 46, D477–D485. doi: 10.1093/nar/gkx1019
- Altenhoff, A. M., Studer, R. A., Robinson-Rechavi, M., and Dessimoz, C. (2012). Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Comput. Biol.* 8:e1002514. doi: 10.1371/journal.pcbi.1002514
- Altschul, S. F., Gertz, E. M., Agarwala, R., Schäffer, A. A., and Yu, Y.-K. (2009). PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Res.* 37, 815–824. doi: 10.1093/nar/gkn981
- Ameismeier, M., Zemp, I., van den Heuvel, J., Thoms, M., Berninghausen, O., Kutay, U., et al. (2020). Structural basis for the final steps of human 40S ribosome maturation. *Nature* 587, 683–687. doi: 10.1038/s41586-020-2929-x
- Badertscher, L., Wild, T., Montellese, C., Alexander, L. T., Bammert, L., Sarazova, M., et al. (2015). Genome-wide RNAi screening identifies protein modules required for 40S subunit synthesis in human cells. *Cell Rep.* 13, 2879–2891. doi: 10.1016/j.celrep.2015.11.061
- Barandun, J., Chaker-Margot, M., Hunziker, M., Molloy, K. R., Chait, B. T., and Klinge, S. (2017). The complete structure of the small-subunit processome. *Nat. Struct. Mol. Biol.* 24, 944–953. doi: 10.1038/nsmb.3472
- Bassler, J., and Hurt, E. (2019). Eukaryotic ribosome assembly. *Annu. Rev. Biochem.* 88, 281–306. doi: 10.1146/annurev-biochem-013118-110817
- Benelli, D., Marzi, S., Mancone, C., Alonzi, T., la Teana, A., and Londei, P. (2009). Function and ribosomal localization of aIF6, a translational regulator shared by archaea and eukarya. *Nucleic Acids Res.* 37, 256–267. doi: 10.1093/nar/gkn959
- Bhaskar, V., Desogus, J., Graff-Meyer, A., Schenk, A. D., Cavadini, S., and Chao, J. A. (2021). Dynamic association of human Ebp1 with the ribosome. *RNA* 27, 411–419. doi: 10.1261/rna.077602.120
- Bohnsack, K. E., and Bohnsack, M. T. (2019). Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J.* 38:e100278. doi: 10.15252/embj.2018100278
- Bohnsack, M. T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 10, 185–191. doi: 10.1261/rna.5167604
- Bohnsack, M. T., Regener, K., Schwappach, B., Saffrich, R., Paraskeva, E., Hartmann, E., et al. (2002). Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. *EMBO J.* 21, 6205–6215. doi: 10.1093/emboj/cdf613
- Bradatsch, B., Katahira, B., Kowalinski, E., Bange, G., Yao, W., Sekimoto, T., et al. (2007). Arx1 functions as an unorthodox nuclear export receptor for the 60S preribosomal subunit. *Mol. Cell* 27, 767–779. doi: 10.1016/j.molcel.2007.06.034
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Calado, A., Treichel, N., Muller, E. C., Otto, A., and Kutay, U. (2002). Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. *EMBO J.* 21, 6216–6224. doi: 10.1093/emboj/cdf620
- Chen, F., Mackey, A. J., Vermunt, J. K., and Roos, D. S. (2007). Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS One* 2:e383. doi: 10.1371/journal.pone.0000383
- Chu, S., Archer, R. H., Zengel, J. M., and Lindahl, L. (1994). The RNA of RNase MRP is required for normal processing of ribosomal RNA. *Proc. Natl. Acad. Sci. U. S. A.* 91, 659–663. doi: 10.1073/pnas.91.2.659
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Ebersberger, I., Simm, S., Leisegang, M. S., Schmitzberger, P., Mirus, O., von Haeseler, A., et al. (2014). The evolution of the ribosome biogenesis pathway from a yeast perspective. *Nucleic Acids Res.* 42, 1509–1523. doi: 10.1093/nar/gkt1137
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995
- Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., Ping, P., et al. (2018). Reactome graph database: efficient access to complex pathway data. *PLoS Comput. Biol.* 14:e1005968. doi: 10.1371/journal.pcbi.1005968
- Farley-Barnes, K. I., McCann, K. L., Ogawa, L. M., Merkel, J., Surovtseva, Y. V., and Baserga, S. J. (2018). Diverse regulators of human ribosome biogenesis discovered by changes in nucleolar number. *Cell Rep.* 22, 1923–1934. doi: 10.1016/j.celrep.2018.01.056
- Freed, E. F., Prieto, J. L., McCann, K. L., McStay, B., and Baserga, S. J. (2012). NOL11, implicated in the pathogenesis of north American Indian childhood cirrhosis, is required for pre-rRNA transcription and processing. *PLoS Genet.* 8:e1002892. doi: 10.1371/journal.pgen.1002892
- Fujiyama-Nakamura, S., Yoshikawa, H., Homma, K., Hayano, T., Tsujimura-Takahashi, T., Izumikawa, K., et al. (2009). Parvulin (Par14), a peptidyl-prolyl cis-trans isomerase, is a novel rRNA processing factor that evolved in the metazoan lineage. *Mol. Cell. Proteomics* 8, 1552–1565. doi: 10.1074/mcp.M900147-MCP200
- Goldfarb, K. C., and Cech, T. R. (2017). Targeted CRISPR disruption reveals a role for RNase MRP RNA in human preribosomal RNA processing. *Genes Dev.* 31, 59–71. doi: 10.1101/gad.286963.116
- Grosjean, H., Gaspin, C., Marck, C., Decatur, W. A., and de Crecy-Lagard, V. (2008). RNomics and Modomics in the halophilic archaea *Haloferax volcanii*: identification of RNA modification genes. *BMC Genomics* 9:470. doi: 10.1186/1471-2164-9-470
- Hage, A. E., and Tollervy, D. (2004). A surfeit of factors: why is ribosome assembly so much more complicated in eukaryotes than bacteria? *RNA Biol.* 1, 10–15. doi: 10.4161/rna.1.1.932
- Hedges, J., West, M., and Johnson, A. W. (2005). Release of the export adapter, Nmd3p, from the 60S ribosomal subunit requires Rpl10p and the cytoplasmic GTPase Lsg1p. *EMBO J.* 24, 567–579. doi: 10.1038/sj.emboj.7600547
- Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A., and Gleizes, P. E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* 6, 225–242. doi: 10.1002/wrna.1269
- Henras, A. K., Soudet, J., Gêrus, M., Lebaron, S., Caizergues-Ferrer, M., Mougins, A., et al. (2008). The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell. Mol. Life Sci.* 65, 2334–2359. doi: 10.1007/s00018-008-8027-0
- Ho, J. H., Kallstrom, G., and Johnson, A. W. (2000). Nmd3p is a Crm1p-dependent adapter protein for nuclear export of the large ribosomal subunit. *J. Cell Biol.* 151, 1057–1066. doi: 10.1083/jcb.151.5.1057
- Hofer, A., Bussiere, C., and Johnson, A. W. (2007). Mutational analysis of the ribosomal protein Rpl10 from yeast. *J. Biol. Chem.* 282, 32630–32639. doi: 10.1074/jbc.M705057200
- Hoppstock, L., Trusch, F., Lederer, C., van West, P., Koenneke, M., and Bayer, P. (2016). Nmpin from the marine thaumarchaeote *Nitrosopumilus maritimus* is an active membrane associated prolyl isomerase. *BMC Biol.* 14:53. doi: 10.1186/s12915-016-0274-1
- Hung, N. J., Lo, K. Y., Patel, S. S., Helmke, K., and Johnson, A. W. (2008). Arx1 is a nuclear export receptor for the 60S ribosomal subunit in yeast. *Mol. Biol. Cell* 19, 735–744. doi: 10.1091/mbc.e07-09-0968
- Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., et al. (2020). Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577, 519–525. doi: 10.1038/s41586-019-1916-6
- Jain, A., Perisa, D., Fliedner, F., von Haeseler, A., and Ebersberger, I. (2019). The evolutionary traceability of a protein. *Genome Biol. Evol.* 11, 531–545. doi: 10.1093/gbe/evz008
- Jaremko, L., Jaremko, M., Elfaki, I., Mueller, J. W., Ejchart, A., Bayer, P., et al. (2011). Structure and dynamics of the first archaeal parvulin reveal a new functionally important loop in parvulin-type prolyl isomerases. *J. Biol. Chem.* 286, 6554–6565. doi: 10.1074/jbc.M110.160713
- Jiang, Z., Carlantoni, C., Allanki, S., Ebersberger, I., and Stainier, D. Y. R. (2020). Tek (Tie2) is not required for cardiovascular development in zebrafish. *Development* 147:dev193029. doi: 10.1242/dev.193029
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551. doi: 10.1093/nar/gkaa970
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070



- Karbstein, K. (2013). Quality control mechanisms during ribosome maturation. *Trends Cell Biol.* 23, 242–250. doi: 10.1016/j.tcb.2013.01.004
- Kargas, V., Castro-Hartmann, P., Escudero-Urquijo, N., Dent, K., Hilcenko, C., Sailer, C., et al. (2019). Mechanism of completion of peptidyltransferase Centre assembly in eukaryotes. *eLife* 8:e44904. doi: 10.7554/eLife.44904
- Kater, L., Thoms, M., Barrio-Garcia, C., Cheng, J., Ismail, S., Ahmed, Y. L., et al. (2017). Visualizing the assembly pathway of nucleolar pre-60S ribosomes. *Cell* 171:e1514. doi: 10.1016/j.cell.2017.11.039
- Keeling, P. J., Corradi, N., Morrison, H. G., Haag, K. L., Ebert, D., Weiss, L. M., et al. (2010). The reduced genome of the parasitic microsporidian *Enterocytozoon bienersi* lacks genes for core carbon metabolism. *Genome Biol. Evol.* 2, 304–309. doi: 10.1093/gbe/evq022
- Kemmler, S., Occhipinti, L., Veisu, M., and Panse, V. G. (2009). Yvh1 is required for a late maturation step in the 60S biogenesis pathway. *J. Cell Biol.* 186, 863–880. doi: 10.1083/jcb.200904111
- Klingauf-Nerurkar, P., Gillet, L. C., Portugal-Calisto, D., Oborska-Oplova, M., Jager, M., Schubert, O. T., et al. (2020). The GTPase Nog1 co-ordinates the assembly, maturation and quality control of distant ribosomal functional centers. *eLife* 9:e52474. doi: 10.7554/eLife.52474
- Klinge, S., and Woolford, J. L. Jr. (2019). Ribosome assembly coming into focus. *Nat. Rev. Mol. Cell Biol.* 20, 116–131. doi: 10.1038/s41580-018-0078-y
- Knuppel, R., Trahan, C., Kern, M., Wagner, A., Grunberger, F., Hausner, W., et al. (2021). Insights into synthesis and function of KsgA/Dim1-dependent rRNA modifications in archaea. *Nucleic Acids Res.* 49, 1662–1687. doi: 10.1093/nar/gkaa1268
- Koestler, T., Av, H., and Ebersberger, I. (2010). FACT: functional annotation transfer between proteins with similar feature architectures. *BMC Bioinformatics* 11:417. doi: 10.1186/1471-2105-11-417
- Koonin, E. V., Mushegian, A. R., and Bork, P. (1996). Non-orthologous gene displacement. *Trends Genet.* 12, 334–336. doi: 10.1016/0168-9525(96)20010-1
- Koonin, E. V., Wolf, Y. I., and Aravind, L. (2001). Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.* 11, 240–252. doi: 10.1101/gr.162001
- Kowalak, J. A., Bruenger, E., Crain, P. F., and McCloskey, J. A. (2000). Identities and phylogenetic comparisons of posttranscriptional modifications in 16 S ribosomal RNA from *Haloflex volcanii*. *J. Biol. Chem.* 275, 24484–24489. doi: 10.1074/jbc.M002153200
- Lecompte, O., Ripp, R., Thierry, J. C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* 30, 5382–5390. doi: 10.1093/nar/gkf693
- Leisegang, M. S., Martin, R., Ramirez, A. S., and Bohnsack, M. T. (2012). Exportin T and Exportin 5: tRNA and miRNA biogenesis—and beyond. *Biol. Chem.* 393, 599–604. doi: 10.1515/hsz-2012-0146
- Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Res.* 37, D229–D232. doi: 10.1093/nar/gkn808
- Liang, X., Zuo, M. Q., Zhang, Y., Li, N., Ma, C., Dong, M. Q., et al. (2020). Structural snapshots of human pre-60S ribosomal particles before and after nuclear export. *Nat. Commun.* 11:3542. doi: 10.1038/s41467-020-17237-x
- Liu, Y., Makarova, K. S., Huang, W. C., Wolf, Y. I., Nikolskaya, A. N., Zhang, X., et al. (2021). Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* 593, 553–557. doi: 10.1038/s41586-021-03494-3
- Lo, K. Y., Li, Z., Wang, F., Marcotte, E. M., and Johnson, A. W. (2009). Ribosome stalk assembly requires the dual-specificity phosphatase Yvh1 for the exchange of Mrt4 with P0. *J. Cell Biol.* 186, 849–862. doi: 10.1083/jcb.200904110
- Londei, P., and Ferreira-Cerca, S. (2021). Ribosome biogenesis in archaea. *Front. Microbiol.* 12:686977. doi: 10.3389/fmicb.2021.686977
- Londei, P., Teixeira, J., Acca, M., Cammarano, P., and Amils, R. (1986). Total reconstitution of active large ribosomal subunits of the thermoacidophilic archaeobacterium *Sulfolobus solfataricus*. *Nucleic Acids Res.* 14, 2269–2285. doi: 10.1093/nar/14.5.2269
- Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science* 303, 95–98. doi: 10.1126/science.1090599
- Lygerou, Z., Allmang, C., Tollervey, D., and Seraphin, B. (1996). Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science* 272, 268–270. doi: 10.1126/science.272.5259.268
- Malyutin, A. G., Musalgaonkar, S., Patchett, S., Frank, J., and Johnson, A. W. (2017). Nmd3 is a structural mimic of eIF5A, and activates the cpGTPase Lsg1 during 60S ribosome biogenesis. *EMBO J.* 36, 854–868. doi: 10.15252/embj.201696012
- Martin-Duran, J. M., Ryan, J. F., Vellutini, B. C., Pang, K., and Hejnl, A. (2017). Increased taxon sampling reveals thousands of hidden orthologs in flatworms. *Genome Res.* 27, 1263–1272. doi: 10.1101/gr.216226.116
- Moy, T. I., and Silver, P. A. (1999). Nuclear export of the small ribosomal subunit requires the ran-GTPase cycle and certain nucleoporins. *Genes Dev.* 13, 2118–2133. doi: 10.1101/gad.13.16.2118
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nieto, B., Gaspar, S. G., Moriggi, G., Pestov, D. G., Bustelo, X. R., and Dosil, M. (2020). Identification of distinct maturation steps involved in human 40S ribosomal subunit biosynthesis. *Nat. Commun.* 11:156. doi: 10.1038/s41467-019-13990-w
- Noon, K. R., Bruenger, E., and McCloskey, J. A. (1998). Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*. *J. Bacteriol.* 180, 2883–2888. doi: 10.1128/JB.180.11.2883-2888.1998
- Ohno, S. (1970). *Evolution by Gene Duplication*. New York: Springer.
- O’Leary, N. A., Wright, M. W., Brister, J. R., Ciuffo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189
- Ostlund, G., Schmitt, T., Forslund, K., Kostler, T., Messina, D. N., Roopra, S., et al. (2010). InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* 38, D196–D203. doi: 10.1093/nar/gkp931
- Pertschy, B., Saveanu, C., Zisser, G., Lebreton, A., Teng, M., Jacquier, A., et al. (2007). Cytoplasmic recycling of 60S preribosomal factors depends on the AAA protein Drg1. *Mol. Cell. Biol.* 27, 6581–6592. doi: 10.1128/MCB.00668-07
- Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergstrom, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5
- Phung, D. K., Etienne, C., Batista, M., Langendijk-Genevaux, P., Moalic, Y., Laurent, S., et al. (2020). RNA processing machineries in archaea: the 5’-3’ exoribonuclease rNase J of the beta-CASP family is engaged specifically with the helicase ASH-Ski2 and the 3’-5’ exoribonucleolytic RNA exosome machinery. *Nucleic Acids Res.* 48, 3832–3847. doi: 10.1093/nar/gkaa052
- Rodriguez-Mateos, M., Garcia-Gomez, J. J., Francisco-Velilla, R., Remacha, M., de la Cruz, J., and Ballesta, J. P. (2009). Role and dynamics of the ribosomal protein P0 and its related trans-acting factor Mrt4 during ribosome assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 37, 7519–7532. doi: 10.1093/nar/gkp806
- Rosenblad, M. A., Lopez, M. D., Piccinelli, P., and Samuelsson, T. (2006). Inventory and analysis of the protein subunits of the ribonucleases P and MRP provides further evidence of homology between the yeast and human enzymes. *Nucleic Acids Res.* 34, 5145–5156. doi: 10.1093/nar/gkl626
- Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng.* 12, 85–94.
- Roustan, V., Jain, A., Teige, M., Ebersberger, I., and Weckwerth, W. (2016). An evolutionary perspective of AMPK-TOR signaling in the three domains of life. *J. Exp. Bot.* 67, 3897–3907. doi: 10.1093/jxb/erw211
- Sanchez, M. E., Urena, D., Amils, R., and Londei, P. (1990). In vitro reassembly of active large ribosomal subunits of the halophilic archaeobacterium *Haloflex volcanii*. *Biochemistry* 29, 9256–9261. doi: 10.1021/bi00491a021
- Sanghai, Z. A., Miller, L., Molloy, K. R., Barandun, J., Hunziker, M., Chaker-Margot, M., et al. (2018). Modular assembly of the nucleolar pre-60S ribosomal subunit. *Nature* 556, 126–129. doi: 10.1038/nature26156
- Saveanu, C., Namane, A., Gleizes, P.-E., Lebreton, A., Rousselle, J.-C., Noaillac-Depeyre, J., et al. (2003). Sequential protein association with nascent 60S ribosomal particles. *Mol. Cell. Biol.* 23, 4449–4460. doi: 10.1128/MCB.23.13.4449-4460.2003

- Sergiev, P. V., Aleksashin, N. A., Chugunova, A. A., Polikanov, Y. S., and Dontsova, O. A. (2018). Structural and evolutionary insights into ribosomal RNA methylation. *Nat. Chem. Biol.* 14, 226–235. doi: 10.1038/nchembio.2569
- Shajani, Z., Sykes, M. T., and Williamson, J. R. (2011). Assembly of bacterial ribosomes. *Annu. Rev. Biochem.* 80, 501–526. doi: 10.1146/annurev-biochem-062608-160432
- Sloan, K. E., Gleizes, P. E., and Bohnsack, M. T. (2016). Nucleocytoplasmic transport of RNAs and RNA-protein complexes. *J. Mol. Biol.* 428, 2040–2059. doi: 10.1016/j.jmb.2015.09.023
- Sloan, K. E., Schneider, C., and Watkins, N. J. (2012). Comparison of the yeast and human nuclear exosome complexes. *Biochem. Soc. Trans.* 40, 850–855. doi: 10.1042/BST20120061
- Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L. J., and Bohnsack, M. T. (2017). Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. doi: 10.1080/15476286.2016.1259781
- Squarrito, M., Mancino, M., Donzelli, M., Areces, L. B., and Draetta, G. F. (2004). EBP1 is a nucleolar growth-regulating protein that is part of pre-ribosomal ribonucleoprotein complexes. *Oncogene* 23, 4454–4465. doi: 10.1038/sj.onc.1207579
- Stage-Zimmermann, T., Schmidt, U., and Silver, P. A. (2000). Factors affecting nuclear export of the 60S ribosomal subunit in vivo. *Mol. Biol. Cell* 11, 3777–3789. doi: 10.1091/mbc.11.11.3777
- Steinberger, M., and Salzberg, S. L. (2020). Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biol.* 21:115. doi: 10.1186/s13059-020-02023-1
- Tafforeau, L., Zorbas, C., Langhendries, J. L., Mullineux, S. T., Stamatopoulou, V., Mullier, R., et al. (2013). The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. *Mol. Cell* 51, 539–551. doi: 10.1016/j.molcel.2013.08.011
- Taoka, M., Nobe, Y., Yamaki, Y., Yamauchi, Y., Ishikawa, H., Takahashi, N., et al. (2016). The complete chemical structure of *Saccharomyces cerevisiae* rRNA: partial pseudouridylation of U2345 in 25S rRNA by snRNA snR9. *Nucleic Acids Res.* 44, 8951–8961. doi: 10.1093/nar/gkw564
- Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997). A genomic perspective on protein families. *Science* 278, 631–637. doi: 10.1126/science.278.5338.631
- Taylor, J. S., and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 38, 615–643. doi: 10.1146/annurev.genet.38.072902.092831
- Tran, N. V., Greshake Tzavaras, B., and Ebersberger, I. (2018). PhyloProfile: dynamic visualization and exploration of multi-layered phylogenetic profiles. *Bioinformatics* 34, 3041–3043. doi: 10.1093/bioinformatics/bty225
- van Tran, N., Ernst, F. G. M., Hawley, B. R., Zorbas, C., Ulryck, N., Hackert, P., et al. (2019). The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Res.* 47, 7719–7733. doi: 10.1093/nar/gkz619
- Wandrey, F., Montellese, C., Koos, K., Badertscher, L., Bammert, L., Cook, A. G., et al. (2015). The NF45/NF90 heterodimer contributes to the biogenesis of 60S ribosomal subunits and influences nucleolar morphology. *Mol. Cell Biol.* 35, 3491–3503. doi: 10.1128/MCB.00306-15
- Warner, J. R. (2015). Twenty years of ribosome assembly and ribosomopathies. *RNA* 21, 758–759. doi: 10.1261/rna.050435.115
- Wild, T., Horvath, P., Wyler, E., Widmann, B., Badertscher, L., Zemp, I., et al. (2010). A protein inventory of human ribosome biogenesis reveals an essential function of exportin 5 in 60S subunit export. *PLoS Biol.* 8:e1000522. doi: 10.1371/journal.pbio.1000522
- Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains archaea, bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* 87, 4576–4579.
- Woolford, J. L. Jr., and Baserga, S. J. (2013). Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* 195, 643–681. doi: 10.1534/genetics.113.153197
- Wu, S., Tutuncuoglu, B., Yan, K., Brown, H., Zhang, Y., Tan, D., et al. (2016). Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. *Nature* 534, 133–137. doi: 10.1038/nature17942
- Yang, X. C., Purdy, M., Marzluff, W. F., and Dominski, Z. (2006). Characterization of 3hExo, a 3' exonuclease specifically interacting with the 3' end of histone mRNA. *J. Biol. Chem.* 281, 30447–30454. doi: 10.1074/jbc.M602947200
- Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* 17, 3011–3016. doi: 10.1101/gad.1158803
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Backstrom, D., Juzokaite, L., Vancaester, E., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358. doi: 10.1038/nature21031
- Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simao, F. A., Ioannidis, P., et al. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* 45, D744–D749. doi: 10.1093/nar/gkw1119
- Zhou, Y., Musalgaonkar, S., Johnson, A. W., and Taylor, D. W. (2019). Tightly-orchestrated rearrangements govern catalytic center assembly of the ribosome. *Nat. Commun.* 10:958. doi: 10.1038/s41467-019-08880-0

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Birikmen, Bohnsack, Tran, Somayaji, Bohnsack and Ebersberger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.