Zoological Systematics, 41(1): 1–47 (January 2016), DOI: 10.11865/zs.201601

ORIGINAL ARTICLE

Natural selection on various sites of ribosomal proteins: a cladistic view

Haoyang Wu[1 †], Yang Liu[1 †], Yanhui Wang[1 †], Jinzhong Lin[2], Qiang Xie[1 *], Wenjun Bu[1]

[1]*Department of Zoology and Developmental Biology, College of Life Sciences, Nankai University, Tianjin 300071, China*
[2]*National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China*
[†]*These authors contributed equally to this work*
[*]*Corresponding authors, E-mail: qiangxie@nankai.edu.cn*

Abstract Deducing the function of certain sites within a protein necessitates *a priori* recognition of the strength of selective pressure. Currently, statistical method is the only option to evaluate the degree of conservation. In the statistical framework, the types of selective pressure can be divided into classifications of negative, nearly neutral and positive. However, such quantitative methods may omit some crucial amino acid sites among the nearly neutral results. In this study, we propose that the cladistic information can be also important to evaluate the functional importance of various amino acid sites. The ribosomal proteins of 62 eukaryotic species were chosen as the case for statistical and cladistic analysis. The evolutionary changes of each site in the aligned sequences were matched on a currently well-accepted cladogram of eukaryotes. Hundreds of synapomorphic sites were discovered in various clades, in which only part of them were suggested to be potentially significant in the statistical framework. Notably, the mutation on His213 of RPL10 in human beings, which are synapomorphic in vertebrates but only be identified as being under neutral selection, is account for the disease Autism. Therefore, the cladistic information can be complementary to the statistical framework in understanding lineage-specific selection event. Additionally, the bias in the accumulation of apomorphic amino acids is significant when going from the Chordata to the Mammalia lineages. This study emphasizes the value of analyzing transcriptomic and proteomic data in a cladistic way to recognize the presence of group-specific selection on various sites in proteins.

Key words Autapomorphy, synapomorphy, functional prediction.

1 Introduction

Adaptive changes in genes and genomes are the essential factors which contribute to evolutionary innovation and species differences, while strongly negative selections on sites maintain the key function of biomacromolecule and ultimately keep the heredity along the lineages. Therefore, in recent years, detect the signals of natural selection via sequence comparison has become a powerful approach in predicting the function of new genes (Yang, 2005). To investigate the function of a certain site, it is critical to evaluate its conservation via comparison with homologous sites in other organisms. In the statistical framework, the selection pressures on amino acids sites are divided into three groups: positive, negative and nearly neutral. The analyses are carried by inferring the selection power, which contrasts the rates of synonymous (silent) and nonsynonymous (replacement) mutations fixed in the population. After grouping sites according to the selection power, sites under positive and negative selection will be seen as having a high likelihood of functionality

urn:lsid:zoobank.org:pub:71F44A42-B117-4BA5-9921-3132F0E3DB7C
Received 3 April 2015, accepted 9 September 2015
Executive editor: Fuqiang Chen

as such region are possibly linked to the ability of species to survive and reproduce (Freudenberg-Hua *et al.*, 2005). The statistical methods have been shown to be effective in many cases (Yang, 2005; Pang *et al.*, 2009), nonetheless, they are sometimes lack of power. As adaptive evolution may occur in only an episodic fashion but not through the whole evolutionary history, the signals of positive selection may be greatly overwhelmed. As a result, some functional sites may be omitted due to the judgement of nearly neutral selection on them, especially in the case of highly conserved genes, in which the statistical framework could provide little evolutionary background for the further functional prediction.

An improvement in detection is the introduction of lineages-specific statistical analysis. However, due to the computational intensiveness, such methods are restricted when dealing with complicated evolutionary scenarios. We propose that the introduction of cladistics into the detection of lineages-specific selection can be helpful in settling these scenarios. In cladistics, a synapomorphy or synapomorphic character is a trait shared by two or more taxa and their most recent common ancestor, whose own ancestor does not possess the trait (Steven, 1983). A unique trait of a given set of terminal groups is often treated as true synapomorphy, i.e., autapomorphy. Cladistics collectively terms these two types of group-specific conservation as “apomorphy”. From the cladistic view, stable apomorphic sites in a diversified group are thought to be under “group-specific” negative selective pressure and may be related to evolutionarily adaptive innovations of their common ancestor. For this reason, analysis through a cladistic view may offer more detailed information about the evolutionary history of sites.

In this study, a cladistic view was introduced to differentiate the levels of selection or conservation, and many more evolutionary hints could thus be discovered. Ribosomal proteins (RP) play a critical role in the assembly and function of ribosomes (Brodersen & Nissen, 2005). In addition, several ribosomal proteins perform important extra-ribosomal functions, including roles in DNA repair, transcriptional regulation and apoptosis (Wool, 1996; O'Brien, 2003). Several clinical syndromes have been demonstrated to be related to ribosomal protein deficiency, such as diamond-blackfan anemia (DBA), isolated congenital asplenia (ICA), and autism (Zhou *et al.*, 2015). It is expected that more related ribosomal protein genes will be discovered and that gene therapy will find a way to target these pathogenic genes appropriately (Flygare *et al.*, 2008). Due to their key roles in life activities, rapid progress in structural biology (Klinge *et al.*, 2011; Rabl *et al.*, 2011), abundance of sequences deposited in public databases and utility in phylogenetic reconstruction (Bleidorn *et al.*, 2009), eukaryotic ribosomal proteins were selected as the test case for a bioinformatics analysis of the conservation of different sites. The analysis of 79 ribosomal protein genes was performed using both statistical and cladistic methods, which includes 62 eukaryotic species representing 23 phyla and 41 classes. According to the cladistic point of view, the aligned sequences of ribosomal proteins were matched on a currently well accepted cladogram of eukaryotes to reconstruct the ancestral state of each amino acid site. A considerable number of apomorphic amino acid sites that are considered to be under nearly neutral selection were found in the alignment. The results also show a significant bias in the accumulation of apomorphic amino acids when going from the Chordata to the Mammalia lineages. In the cladistic view of selection and conservation, it is possible to provide a more granular subsection within the concept of conservation. In addition, the evaluation of the functional importance of specific sites will be more precise in medical and structural biology studies.

2 Materials and methods

2.1 Ribosomal protein gene sequence datasets generation

The sequence data of the RP genes for most species used in this study were downloaded from GenBank (www.ncbi.nlm.nih.gov/genbank/), KEGG (www.genome.jp/kegg/), and Ribosomal Protein Gene (RPG) Database (http://ribosome.med.miyazaki-u.ac.jp/). The EST data from *Locusta migratoria*, *Reticulitermes flavipes*, *Homalodisca vitripennis*, *Nilaparvata lugens* and *Rhodnius prolixus* were downloaded from GenBank using the proseqco script (Peters *et al.*, 2011). Sequence assembling was completed by CAP3 (Huang & Madan, 1999). We set the overlap-length cutoff to greater than 100 and the overlap-percent-identity cutoff to greater than 90 to balance the number of contigs and the quality of the assembly. The sequence data from *Pantala flavescens* was obtained from an unpublished transcriptomic dataset assembled by Velvet (Zerbino & Birney, 2008). We utilized Blast2GO (Conesa *et al.*, 2005; Conesa & Götz, 2008; Götz *et al.*, 2008; Götz *et al.*, 2011) to automatically annotate the contigs (E-valve threshold = 10-6) temporarily. The contigs that are annotated as ribosomal protein genes were rechecked by a Blast Search of the RPG database. The entire non-coding regions were manually excluded from the sequences. Newly generated RP sequence of *Pantala flavescens* have been submitted to GenBank with the accession numbers KC507298 to KC507367. The sequences were sorted by homologous

relationship and aligned using ClustalW (Larkin *et al.*, 2007) with a default parameter. Obvious misalignments were manually corrected.

2.2 Estimation of conservation level

A consensus tree of eukaryote diversification was summarized from recently published phylogenomic results (Burki *et al.*, 2007; Burki *et al.*, 2008; Hampl *et al.*, 2009; Minge *et al.*, 2009). In the cladistic framework, the cladistics information was first obtained by a parsimony method using PAUP 4.0b (Swofford, 2002). After importing the dataset of each ribosomal protein and corresponding treefile, state optimization of parsimony was set to DELTRAN (delayed transformation). Consequently, the log-file option was activated in order to obtain the labeled tree with a complete list of apomorphies (describetrees/plot=phylogram labelnode=yes apolist=yes). The output revealed all of the possible apomorphies of the corresponding dataset. Changes that were potentially homoplasious were abandoned according to the filter criterion (consistency index lower than 0.3) in the study of Marin *et al.* (2005). We further conduct another ancestral state reconstruction using Mesquite version 2.75 build 564 (Maddison *et al.*, 2011) to recheck and classify the apomorphies according to the conservation level of each homologous site. Three types of evolutionary restraint were summarized: amino acids with equivalent biochemical properties serving as synapomorphy *sensu lato* (*s. lat.*), amino acids serving as synapomorphy *sensu stricto* (*s. str.*), and amino acids serving as unique synapomorphy, i.e., potential autapomorphy. The classification of functionally equivalent amino acid residues was based on the default biochemistry table of amino acid residues included in MEGA 5.0 (Tamura *et al.*, 2011).

Meanwhile, the estimation of the conservation level in the statistical framework was performed using PAML 4.5 (Yang, 1997, 2007) using the branch-site model. The treefiles were the same as the ones used in the cladistic analysis. Codon frequency model was set to F3X4. For rates constancy or variation among lineages, no clock model was applied, which assumes the rates are entirely free to vary from branch to branch. All parameters, i.e., kappa, omega, gamma, and rho, were set to be estimated. The Bayes empirical Bayes (BEB) (Yang *et al.*, 2005) for each site was calculated, which is used to show the posterior probabilities for site classes, i.e., negative selection, neutral selection, and positive selection.

2.3 Clinical significance detecting and labeling of sites in the structures of RP

To reassess the functionality of the synapomorphies detected in the dataset of RP, we conducted a comparative analysis between the sets of synapomorphies and the data of disease-related single nucleotide polymorphisms (SNP) in human beings. Site-specific clinical significance data were obtained from the OMIM (http://omim.org/), dbSNP (www.ncbi.nlm.nih.gov/projects/SNP/) and DBA Mutation (http://www.dbagenes.unito.it) databases. For each synapomorphy of RP, we checked its statistical information and scanned the clinical data to find whether it could be related to a disease. Synapomorphic sites and disease-related sites were labeled onto the 3D structures of RP. The labeling was conducted by the UCSF Chimera package (Pettersen *et al.*, 2004) using the 3.0 Å crystal structure of Saccharomyces cerevisiae 80S ribosome (Ben-Shem *et al.*, 2011). As the structural information of RPL29, RPP1, RPP2 at 3.0 Å was absent in the current data, their crystal structures at a resolution of 6.1 Å were utilized (Armache *et al.*, 2010).

3 Results

3.1 Detection of conservation in cladistic and statistical frameworks

Amino acids serving as the synapomorphy *s. str.* and *s. lat.* of Eukaryota are listed in Tables S1 and S2, while the unique synapomorphic amino acids are listed in Tables S3 and S4. Most of the apomorphic sites were found to be under pure selection (for more information, see Tables S5 and S6) with moderate and significant posterior possibilities. Meanwhile, a relatively small number of apomorphic sites were discovered to be nearly neutrally selected (Table S7). The number of synapomorphic sites with uniqueness and synapomorphic sites *s. str.* of each clade within the eukaryotic cladogram was summarized (Fig. 1). It may not be surprising to find that the number of apomorphic sites varied among the different clades considering the origin time of these groups. However, it is still astonishing to find that there exist significant biases in the number of apomorphic sites between the Chordata group and the Ecdysozoa group. The lineage from Chordata to Mammalia has accumulated significantly more synapomorphic sites with uniqueness and synapomorphic sites *s. str.* compared to the lineage from Ecdysozoa to Neoptera, especially in the Vertebrata clade (28/115 and 50/209 correspond to the large and small subunits, respectively). Comparatively, the number of apomorphic sites in the

Ecdysozoa-Neoptera lineage, which contains the groups of insects that are shown to exhibit great diversity (Benton, 1997; Carroll, 2001; Chapman, 2006), is significantly lower than the other groups such as plants and Dikarya fungi.

3.2 Structural evaluation of ribosomal protein

For example, in the labeled sites RPL11 and RPS19 (Fig. 2, for more information, see Fig. S1 and Fig. S2), the synapomorphic sites *s. str.* and synapomorphic sites *s. lat.* of Eukaryota were primarily located within the structural core of certain proteins that play a vital role in the process of protein folding. However, the apomorphic sites according to the lower category levels than Eukaryota are scattered within the overall structure of ribosomal proteins and show no tendency of concentration. Although it has been suggested that RP closely assembled to rRNA have more important functions (Karbstein, 2011), the scattered distribution of conserved sites implies that every single ribosomal protein plays its own vital roles in the function of a ribosome.

3.3 Mutation analysis of RP-related syndrome

Conserved sites may potentially play vital roles in the studies of certain diseases; therefore, we compared the apomorphy dataset with the disease-related mutation information and ribosomal protein coding SNPs (cSNPs). The intersections between the two datasets are listed in Table 1 and Table S8. Although most sites of the cSNPs related to disease are shown to be under negative selection in the statistical framework, a cSNP (His213Gln on RPL10) was shown to be under nearly neutral selection (BEB 0.99991), which has been reported to be related to autism (Klauck *et al.*, 2006). This site is significantly important in the cladistic framework due to the clear synapomorphic state “Histidine” shared by all vertebrates, which indicates the possibility of pseudo-negativity for some crucial sites obtained through statistical methods. Thus, it is necessary to use the method of ancestral state reconstruction and the view of cladistics in future structural and functional studies.

Table 1. Disease-related substitution on RP and their conservation based on statistical and cladistic views.

Gene	Position in *Homo sapiens*	Disease	Statistical Conservation	Cladistic Conservation
RPSA	Arg180Gly/Trp	ICA	Negative	Synapomorphy of Eukaryota*
RPSA	Thr54Asn	ICA	Negative	Synapomorphy of Eukaryota*
RPSA	Leu58Phe	ICA	Negative	Synapomorphy of Eukaryota**
RPS19	Val15Phe	DBA	Negative	Synapomorphy of Eukaryota**
	Leu28Arg	DBA	Negative	Synapomorphy of Eukaryota**
	Leu45Gln	DBA	Negative	Synapomorphy of Eukaryota*
	Pro47Leu	DBA	Negative	Synapomorphy of Eukaryota*
	Thr55Met	DBA	Negative	Synapomorphy of Bilateria
	Arg62Trp/Gln	DBA	Negative	Synapomorphy of Eukaryota**
	Leu64Pro	DBA	Negative	Synapomorphy of Eukaryota**
	Thr76Pro	DBA	Negative	Synapomorphy of Bilateria
	Arg94Leu	DBA	Negative	Synapomorphy of Bilateria
	Arg101Cys/His	DBA	Negative	Synapomorphy of Eukaryota*
	Arg102Pro	DBA	Negative	Synapomorphy of Tetrapoda
	Leu107Arg	DBA	Negative	Synapomorphy of Eukaryota**
	Gly120Ser	DBA	Negative	Synapomorphy of Eukaryota*
	Gly127Glu	DBA	Negative	Synapomorphy of Eukaryota*
	Leu131Pro/Arg	DBA	Negative	Synapomorphy of Eukaryota**
	Ala135Thr	DBA	Negative	Synapomorphy of Eukaryota*
RPS26	Asp33Asn	DBA	Negative	Synapomorphy of Eukaryota*
RPL5	GLY140SER	DBA	Negative	Synapomorphy of Bilateria
RPL10	Leu206Met	Autism	Negative	Synapomorphy of Eukaryota**
	His213Gln	**Autism**	**Neutral**	**Synapomorphy of Vertebrata**
RPL35A	Val33Ile	DBA	Negative	Synapomorphy of Eukaryota*

* Certain amino acids serve as synapomorphy *s. str.* of Eukaryota.

** Certain amino acids serve as synapomorphy *s. lat.* of Eukaryota.

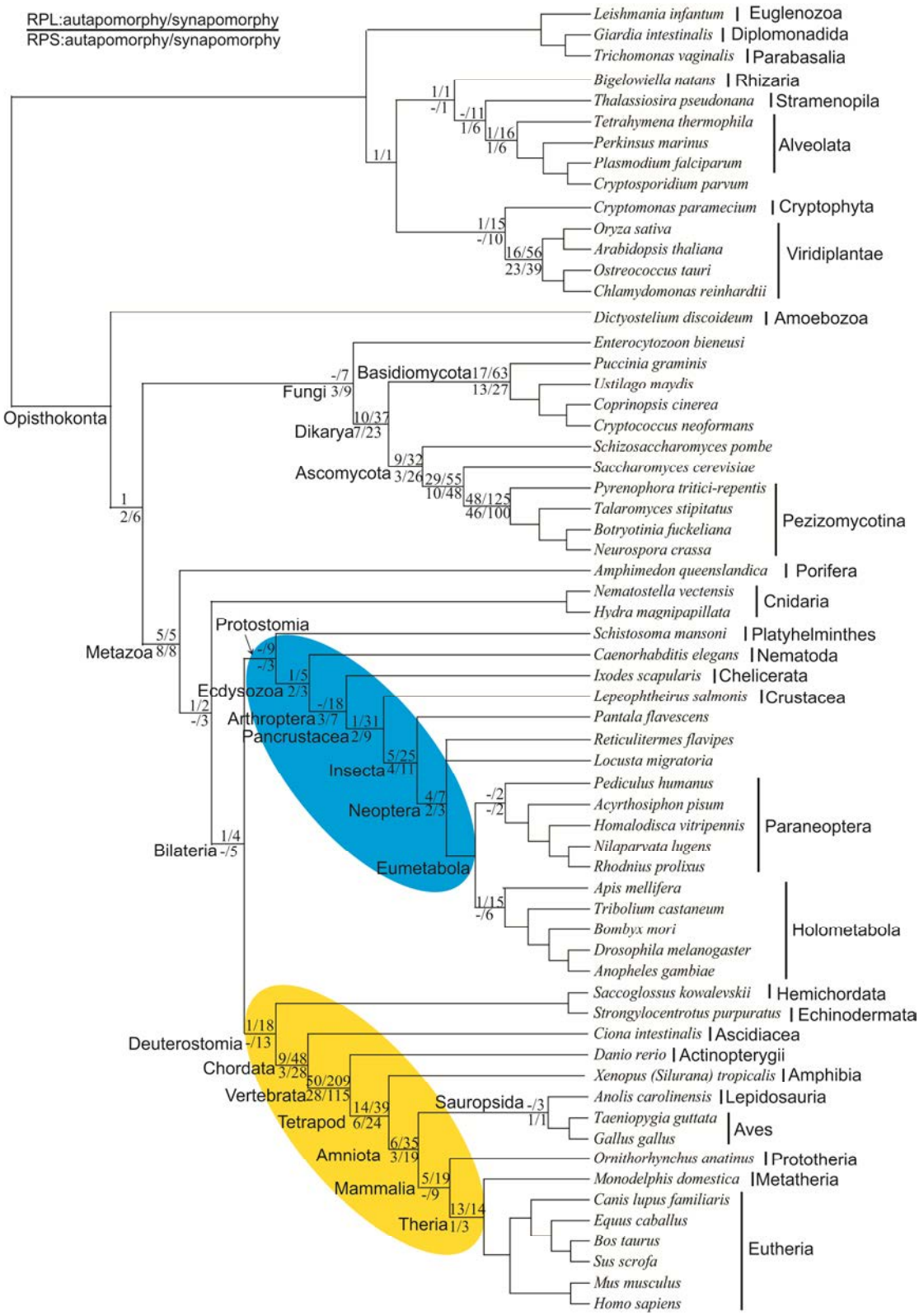

Figure 1. Summarized distribution of group-specific sites in the cladogram of eukaryotic diversification. The numerals shown above the line correspond to RPL, and the numerals shown below correspond to RPS. The Ecdysozoa-Neoptera and Chordata-Mammalia lineages are highlighted with blue and yellow, respectively. The amino acids with equivalent biochemical properties were not taken into account.

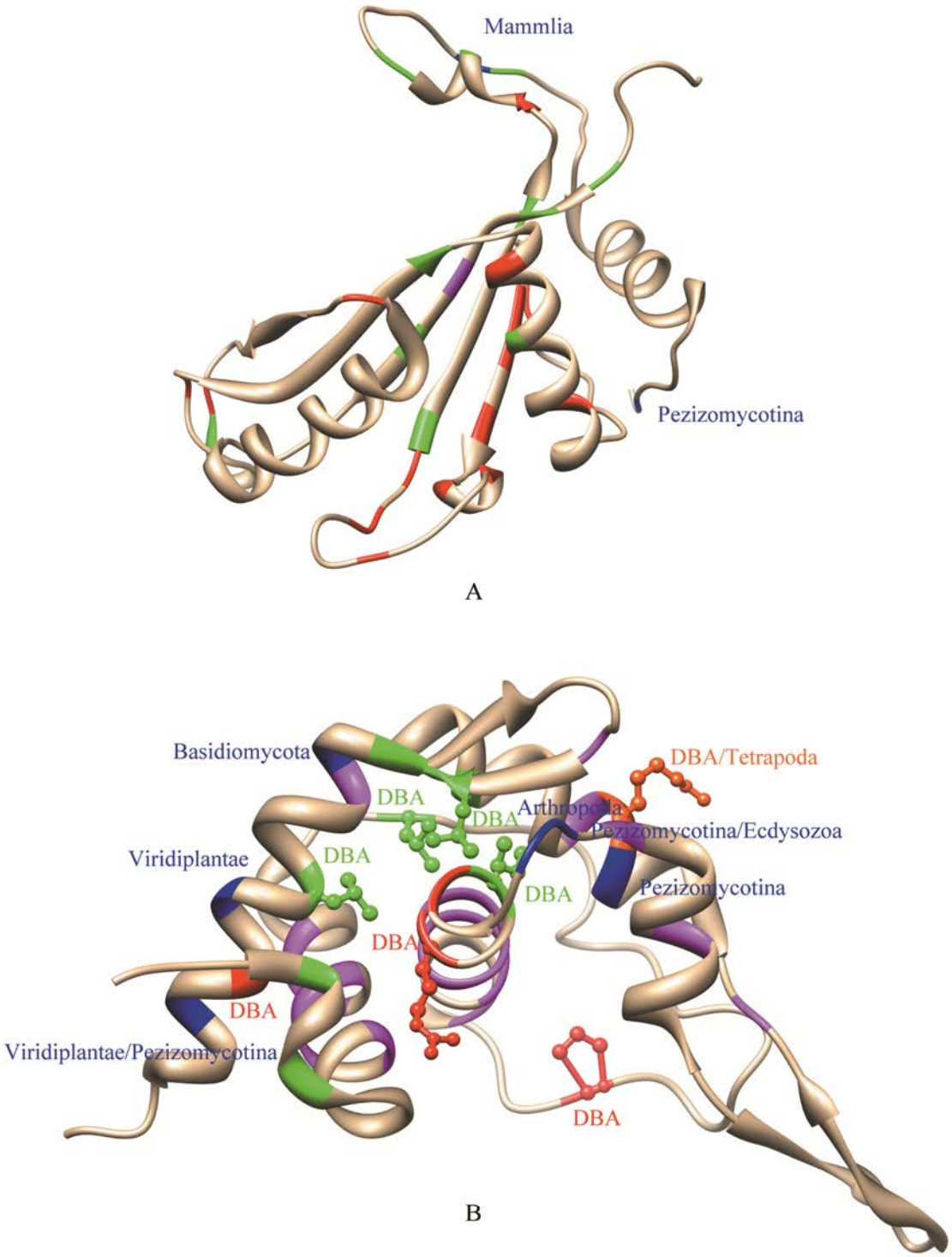

Figure 2. Conserved sites in the tertiary structures of RPL11 (A) and RPS19 (B). The red ribbons represent the synapomorphic *s. str.* sites of Eukaryota, the green ones represent the synapomorphic sites *s. lat.* of Eukaryota, and the blue ones represent the potential autapomorphic sites of clades lower than Eukaryota. The ribbons with sticks and balls represent sites related to DBA. The synapomorphic sites related to DBA are shown in orange. The magenta ones represent the sites for which it is difficult to deduce the synapomorphy but appear to be related to DBA.

4 Discussion

4.1 Inference from the obvious bias between the two major metazoan lineages

According to fossil records, the oldest chordates were found to be existed in the Cambrian age, almost 525 million years ago (Chen *et al.*, 1995), while the oldest ecdysozoans emerged nearly 540 million years ago (Lieberman, 1995). The comparisons of the ages inferred from the earliest fossil records between inner clades, such as Vertebrata vs. Arthropoda (Shu *et al.*, 1999; Yuan *et al.*, 2002), Tetrapoda vs. Pancrustacea (Briggs, 1978; Clack, 2002), Amniota vs. Insecta (Engel & Grimaldi, 2004), and Mammalia vs. Neoptera (Luo *et al.*, 2011), also show similar results. Therefore, the bias of accumulation of apomorphies cannot simply be attributed to the earlier appearance of Chordata. As a terminal taxon in the lineage from Chordata to Mammalia, mammals have accumulated the most apomorphic sites in RP in all of the organisms at the class level. We propose that there may be two possible reasons that contribute to such bias. The genetic effective population size (Ne) has been shown to have great influence on the genetic architecture and molecular evolution (Lynch, 2007; Charlesworth, 2009). A reduced Ne will result in decreased amounts of genetic polymorphism in a panmictic Wright-Fisher population (Kimura, 1983). Meanwhile, as the Ne is inversely proportional to the rate of genetic drift, a reduced Ne will also result in an increased power of genetic drift that counteracting the effect of natural selection (Kimura, 1983; Ohta, 1987; Lynch, 2007). Therefore, a group of organisms with smaller Ne, such as the vertebrates, may tend to have mutations spread and fixed easier due to the relaxed constraints from natural selection, which results in the bias of accumulation of apomorphies in the lineage from Chordata to Mammalia. On the other hand, considering the correlation between some of these sites and known disease-related cSNPs, these apomorphies accumulated in the lineage from Chordata to Mammalia may also indicate a gradually functional specialization of RP, which benefits its members by increasing the fitness of them. As there may be interactions among certain amino acid residues of proteins, it is possible that an accidental fixation of mutation in the ancestors of Chordata opened a narrow window of distant positive selection. In this scenario, the bias of accumulation of apomorphies in the lineage from Chordata to Mammalia can be driven by the power of natural selection. Further studies on the function of apomorphic sites may enhance the understanding of the mechanisms underlying these constraints.

4.2 Necessity of the cladistic view in structural and functional studies

The popularization of next-generation sequencing technology has resulted in a large amount of sequence data, which could be utilized in several different areas, such as structural and function studies of proteins. The evolution of protein under selection directly affects the molecular phenotype of amino acid residues, which is characterized by properties that affect protein function such as protein structure, protein stability, and protein binding specificity (Liberles *et al.*, 2012). Long-term evolutionary histories of the sites within genomes offer a diversified background for comparing SNPs within human genomes and homologous sites in other organisms. This comparison may aid the reassessment of the importance of the sites that are thought to play key roles in certain diseases (Dudley *et al.*, 2012). Phylomedicine has emerged as a discipline in recent years (Kumar *et al.*, 2011), and several studies have been provided from the view of evolution (Dudley *et al.*, 2012) and comparative study (Xie *et al.*, 2012). Diversified sites in proteins are vital during diversification and adaptation, whereas conservative sites have essential impacts on housekeeping functions. Thus, comparative information can be used for the a priori assessing of the unknown function of certain sites within a protein.

However, the statistical methods may not be sufficiently robust in the case of proteins in which the signals of positive selection are overwhelmed. As aforementioned, in the long-term evolutionary history, the adaptive evolutions on certain sites may be intermittent and only in a narrow window of evolutionary time (Gillespie, 1991). Therefore, it is possible that a certain site with functionality may be detected as under neutral selection and would be omitted in the statistical analysis. Lineage-specific statistical methods, which are hopeful to make improvements in such cases, are sometimes restricted due to the computational intensiveness. As a result, deducing the function of certain sites within a highly conserved protein remains a challenging task. From the cladistic view, the detection of stable apomorphic sites in a diversified group may help solve this problem because of their potential relationship with the adaptive innovations of their common ancestor. In this study, some of the group-specific conservation discovered by ancestral reconstruction cannot be effectively detected using a statistical method (Table 1, S8). The intersection of apomorphic sites with some famous disease-related cSNPs indicates the potential importance of group-specific conservation. For this reason, the sites with apomorphies listed in Table S7, which are all under nearly neutral selection, may be also with potential functionality and is noteworthy for further genomic disease analysis.

The amazing development of high-throughput sequencing technology and the great decrease in its cost make the

investigation of large-scale genomic data from different organisms both possible and indispensable. To date, extensive research based on EST, genomic and transcriptomic data have shown promising problem-solving potential for such issues as evolutionary history reconstruction, lineage evolution illumination and tree of life clarification. Based on the results of RP, apomorphic sites comprise approximately 3.5% of all aligned sites. Assuming that the fraction of apomorphic sites is similar in the transcriptome of all eukaryotes and that the yeast transcriptome is the smallest eukaryotic transcriptome, approximately 30 000 apomorphic sites exist in the eukaryotic transcriptomes. These sites may contribute to most of the molecular mechanisms responsible for the diversification of eukaryotes. Compared with statistical methods, the view of cladistics can offer a more intuitionistic qualitative identification of important sites, which would enable researchers to take the evolutionary background and cladistic information into account. Therefore, we propose that the cladistic view should be taken into account in structural and functional studies.

Nonetheless, we do not claim that the cladistic method can completely replace the statistical methods in the study concerning molecular evolution and function prediction. The information of cladistics provides an intuitive way for understanding the evolutionary history. As a result, the cladistic methods could effectively detect the evolutionary restrictions that related to distant adaptations at different category levels even the sites is judged to be under neutral selection. On the other hands, the statistical methods within phylogenetic context can provide a framework of detection with quantitative analysis and likelihoods, which is imperative in analyzing the positive selection in recent or series of fast adaptation event. Therefore, the two methodologies can be complementary to each other in structural and functional studies.

Funding This work was supported by the National Natural Science Foundation of China (31222051, J1210005).

References

Armache, J.P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., *et al.* 2010. Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 107: 19 754–19 759.

Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., Yusupov, M. 2011. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, 334: 1 524–1 529.

Benton, M.J. 1997. Models for the diversification of life. *Trends in Ecology & Evolution*, 12: 490–495.

Bleidorn, C., Podsiadlowski, L., Zhong, M., Eeckhaut, I., Hartmann, S., Halanych, K.M., Tiedemann, R. 2009. On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evolutionary Biology*, 9: 150.

Briggs, D.E.G. 1978. The morphology, mode of life, and affinities of Canadaspis perfecta (Crustacea: Phyllocarida), Middle Cambrian, Burgess Shale, British Columbia. *Philosophical Transactions of the Royal Society B*, 281: 439–487.

Brodersen, D.E., Nissen, P. 2005. The social life of ribosomal proteins. *FEBS*, 272: 2 098–2 108.

Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjæveland, A., Nikolaev, S.I., Jakobsen, K.S., Pawlowski, J. 2007. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One*, 2: e790.

Burki, F., Shalchian-Tabrizi, K., Pawlowski, J. 2008. Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. *Biology Letters*, 4: 366-369.

Carroll, S.B. 2001. Chance and necessity: the evolution of morphological complexity and diversity. *Nature*, 409: 1 102–1 109.

Chapman, A.D. 2006. *Numbers of living species in Australia and the World*. Australian Biological Resources Study, Canberra.

Charlesworth, B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10: 195–205.

Chen, J.Y., Dzik, J., Edgecombe, G.D., Rasmskold, L., Zhou, G.Q. 1995. A possible Early Cambrian chordate. *Nature*, 377: 720–722.

Clack, J.A. 2002. *Gaining ground: the origin and evolution of tetrapods*. Indiana University Press, Bloomington, Indiana, USA.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21: 3 674–3 676.

Conesa, A., Götz, S. 2008. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Genomics*, 2008: 1–13.

Dudley, J.T., Chen, R., Sanderford, M., Butte, A.J., Kumar, S. 2012. Evolutionary meta-analysis of association studies reveals ancient constraints affecting disease marker discovery. *Molecular Biology and Evolution*, 29: 2 087–2 094.

Engel, M.S., Grimaldi, D.A. 2004. New light shed on the oldest insect. *Nature*, 427: 627–630.

Flygare, J., Olsson, K., Richter, J., Karlsson, S. 2008. Gene therapy of Diamond Blackfan anemia CD34 (+) cells leads to improved erythroid development and engraftment following transplantation. *Experimental Hematology*, 36: 1 428–1 435.

Freudenberg-Hua, Y., Freudenberg, J., Winantea, J., Kluck, N., Cichon, S., Brüss, M., Propping, P., Nöthen, M.M. 2005. Systematic investigation of genetic variability in 111 human genes-implications for studying variable drug response. *Pharmacogenomics*, 5: 183–192.

Gillespie, J. H. 1991. *The Causes of Molecular Evolution*. Oxford Univ. Press, Oxford.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36: 3420–3435.

Götz, S., Arnold, R., Sebastián-León, P., Martín-Rodríguez, S., Tischler, P., Jehl, M.A., Dopazo, J., Rattei, T., Conesa, A. 2011. B2G-FAR, a species centered GO annotation repository. *Bioinformatics*, 27: 919–924.

Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G., Roger, A.J. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proceedings of the National Academy of Sciences of the United States of America*, 106: 3859–3864.

Huang, X., Madan, A. 1999. CAP3: A DNA sequence assembly program. *Genome Research*, 9: 868–877.

Karbstein, K. 2011. Inside the 40S ribosome assembly machinery. *Current Opinion in Chemical Biology*, 15: 657–663.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Klauck, S.M., Felder, B., Kolb-Kokocinski, A., Schuster, C., Chiocchetti, A., Schupp, I., Wellenreuther, R., Schmötzer, G., Poustka, F., Breitenbach-Koller, L., Poustka, A. 2006. Mutations in the ribosomal protein gene RPL10 suggest a novel modulating disease mechanism for autism. *Molecular Psychiatry*, 11: 1073–1084.

Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., Ban, N. 2011. Crystal structure of the Eukaryotic 60S ribosomal subunit in complex with Initiation Factor 6. *Science*, 334: 941–948.

Kumar, S., Dudley, J.T., Filipski, A., Liu, L. 2011. Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends in Genetics*, 27: 377–386.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* 2007. ClustalW and ClustalX version 2. *Bioinformatics*, 23: 2947–2948.

Liberles, D.A., Teichmann, S.A., Bahar, I., Bastolla, U,, Bloom, J., Bornberg-Bauer, E., Colwell, L.J., de Koning, A.P., Dokholyan, N.V., Echave, J., Elofsson, A., Gerloff, D.L., Goldstein, R.A., Grahnen, J.A., Holder, M.T., Lakner, C., Lartillot, N., Lovell, S.C., Naylor, G., Perica, T., Pollock, D.D., Pupko, T., Regan, L., Roger, A., Rubinstein, N., Shakhnovich, E., Sjölander, K., Sunyaev, S., Teufel, A.I., Thorne, J.L., Thornton, J.W., Weinreich, D.M., Whelan, S. 2012. The interface of protein structure, protein biophysics, and molecular evolution. *Protein Science*, 21: 769–785.

Lieberman, B.S. 1995. Testing the Darwinian legacy of the Cambrian radiation using trilobite phylogeny and biogeography. *Journal of Paleontology*, 73: 176.

Luo, Z.X., Yuan, C.X., Meng, Q.J., Ji, Q. 2011. A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature*, 476: 442–445.

Lynch, M. 2007. *The Origins of Genome Architecture*. Sinauer Associates Inc, Sunderland.

Maddison, W.P., Maddison, D.R. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75. Available from http://mesquiteproject. org.

Marin, B., Nowack, E.C., Melkonian, M. 2005. A Plastid in the Making: Evidence for a Second Primary Endosymbiosis. *Protist.*, 156: 425–432.

Minge, M.A., Silberman, J.D., Orr, R.J., Cavalier-Smith, T., Shalchian-Tabrizi, K., Burki, F., Skjaeveland, A., Jakobsen, K.S. 2009. Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proceedings of the Royal Society B: Biological Sciences*, 276: 597–604.

O'Brien, T.W. 2003. Properties of human mitochondrial ribosomes. *IUBMB Life*, 55: 505–513.

Ohta, T. 1987. Very slightly deleterious mutations and the molecular clock. *Journal of Molecular Evolution*, 26: 1–6.

Pang, G.S., Wang, J., Wang, Z., Lee, C.G. 2009. Predicting potentially functional SNPs in drug-response genes. *Pharmacogenomics*, 10: 639–653.

Peters, R.S., Meyer, B., Krogmann, L., Borner, J., Meusemann, K., Schütte, K., Niehuis, O., Misof, B. 2011. The taming of an impossible child: a standardized all-in approach to the phylogeny of Hymenoptera using public database sequences. *BMC Biology*, 9: 55.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25: 1605–1612.

Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., Ban, N. 2011. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science*, 331: 730–736.

Shu, D.G., Conway-Morris, S., Zhang, X.L. 1999. Lower Cambrian vertebrates from south China. *Nature*, 402: 42–46.

Steven, J.D. 1983. *Hen's Teeth and Horse's Toes.* Norton Press, New York. pp 358.

Swofford, D.L. 2002. *PAUP* Phylogenetic Analysis using Parsimony (* and other Methods). Version 4.* Sinauer Associates, Sunderland, Massachusetts.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, 28: 2731–2739.

Wool, I.G. 1996. Extraribosomal functions of ribosomal proteins. *Trends in Biochemical Sciences*, 21: 164–165.

Xie, Q., Wang, Y., Lin, J., Qin, Y., Wang, Y., Bu, W.J. 2012. Potential key bases of ribosomal RNA to kingdom-specific spectra of antibiotic susceptibility and the possible Archaeal origin of Eukaryotes. *PLoS One*, 7: e29468.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS*, 13: 555–556.

Yang, Z. 2005. The power of phylogenetic comparison in revealing protein function. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 3179–3180.

Yang, Z., Wong, W. S. W., Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, 22: 1107–1118.

Yang, Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24: 1586–1591.

Yuan, X., Xiao, S., Parsley, R.L., Zhou, C., Chen, Z., Hu, J. 2002. Towering sponges in an Early Cambrian Lagerstätte: Disparity between nonbilaterian and bilaterian epifaunal tierers at the Neoproterozoic-Cambrian transition. *Geology*, 30: 363–366.

Zerbino, D.R., Birney, E. 2008. "Velvet: Algorithms for de novo short read assembly using de Bruijn graphs". *Genome Research*, 18: 821–829.

Zhou, X., Liao, W.J., Liao, J.M., Liao, P., Lu, H. 2015. Ribosomal proteins: functions beyond the ribosome. *Journal of Molecular Cell Biology*, 7: 92–104.

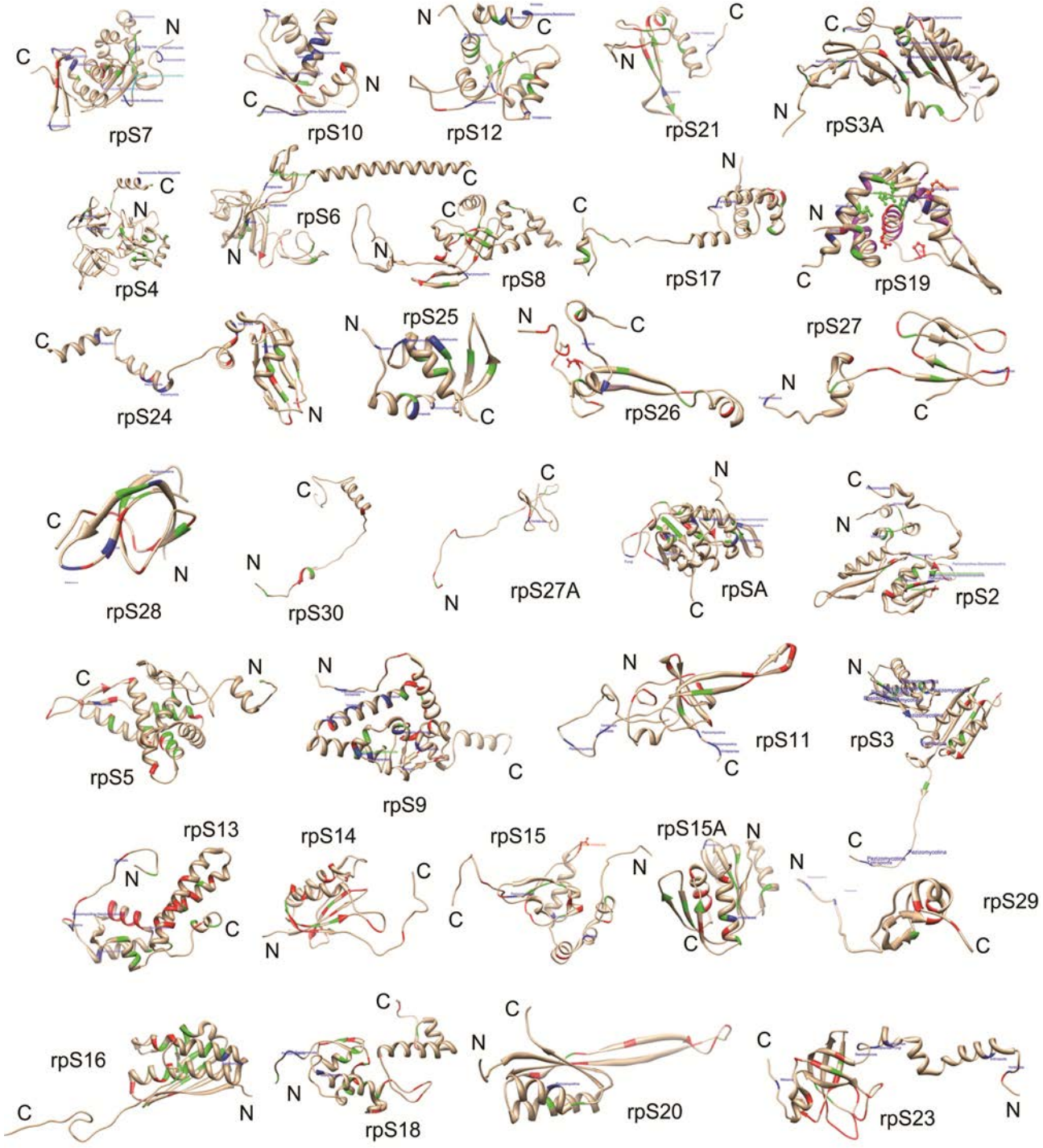

Figure S1. Conserved sites of ribosomal protein 40S subunit labeling on crystal structure. The ribbons colored in red represent the synapomorphic *s. str.* sites of Eukaryota, the green ones represent the synapomorphic *s. lat.* sites of Eukaryota, and the blue ones represent the potential autapomorphic sites of clades lower than Eukaryota. The magenta ribbons represent the sites related to DBA, and the sites presented by sticks and balls are also synapomorphic sites. Because there are too many synapomorphic sites according to clades lower than Eukaryota in the RP, only the synapomorphic sites related to DBA are colored in this figure (shown in orange).

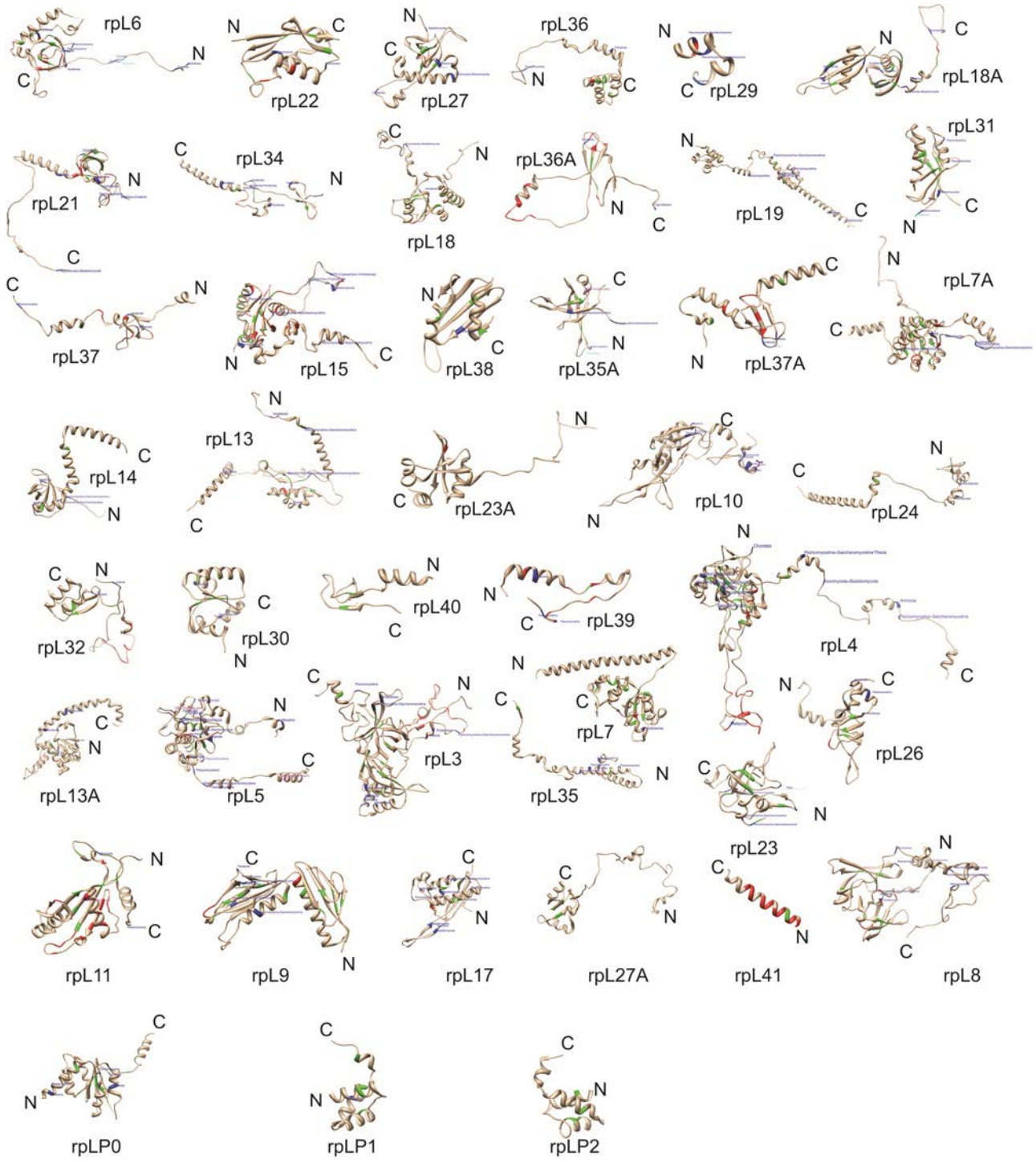

Figure S2. Conserved sites of ribosomal protein 60S subunit labeling on crystal structure. The ribbons colored in red represent the synapomorphic *s. str.* sites of Eukaryota, the green ones represent the synapomorphic *s. lat.* sites of Eukaryota and the blue ones represent the potential autapomorphic sites of clades lower than Eukaryota. The magenta ribbons represent the sites related to DBA, and the sites shown by sticks and balls are also synapomorphic sites. Because there are too many synapomorphic sites according to clades lower than Eukaryota in the RP, only the synapomorphic sites related to DBA are colored in this figure (shown in orange).

Table S1. Synapomorphic *s. str.* sites of Eukaryota in the currently available genomic data from Eukaryota.

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
RPSA	P104	P105	
	G105	G106	
	N109	N110	
	P118	P119	
	E135	E136	
	P142	P143	
	I144	I145	
	D157	D158	
	P161	P162	
RPS2	G105	G131	
	G112	G138	
	G142	G168	
	W144	W170	
	G160	G186	
	P171	P197	
	A172	A198	
	P173	P199	
	G175	G201	
	N209	N235	
RPS3	G133	G133	
	G140	G140	
	R146	R146	
	G155	G155	
	G161	G161	
	G183	G183	
RPS3A	F100	F100	
	K116	K116	
	T129	T129	
	F138	F138	
RPS4	G34	G34	
	D88	D88	
	R100	R100	
	G186	G186	
	G190	G190	
	G229	G229	
RPS5	N128	N107	
	P129	P108	
	E144	E123	
	G153	G132	
	D161	D140	
	P164	P143	
	G178	G157	
	F184	F163	
	E197	E176	
	E218	E197	
	R225	R204	
RPS6	D57	D57	
	G60	G60	
	G99	G99	
	K136	K136	
RPS7	E82	E84	
	R96	R98	
	V121	V125	
	P132	P136	
	G137	G141	

Table S1 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
	R139	R143	
	L154	L158	
RPS8	R47	R47	
	G50	G50	
	K54	K54	
	G63	G63	
	A167	A173	
	S171	S177	
	P173	P179	
	G174	G180	
	Q175	Q181	
	G177	G183	
	G181	G187	
RPS9	P15	P16	
	P18	P19	
	R23	R24	
	E27	E28	
	G35	G36	
	L36	L37	
	K39	K40	
	E41	E42	
	L59	L60	
	L93	L94	
	R107	R108	
	R108	R109	
	L109	L110	
	Q110	Q111	
	R126	R127	
	I134	I135	
	P144	P145	
	S162	S163	
	W61	W64	
RPS10	L68	L71	
	G72	G75	
	L80	L83	
	D55	D57	
RPS11	K57	K59	
	C58	C60	
	P59	P61	
	R67	R69	
	G68	G70	
	I70	I72	
	K79	K81	
	Y90	Y92	
	L91	L93	
	K96	K98	
	Y97	Y99	
	R99	R101	
	R103	R105	
	G120	G123	
	D121	D124	
	R129	R132	
	P130	P133	
	K135	K136	
	T136	T137	

Table S1 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
	G44	G34	
RPS12	L78	L68	
	K114	K102	
	G10	G10	
RPS13	P17	P17	
	W25	W25	
	P47	P47	
	S48	S48	
	I50	I50	
	G51	G51	
	L54	L54	
	R55	R55	
	D56	D56	
	P85	P85	
	E86	E86	
	K93	K93	
	A95	A95	
	H101	H101	
	D108	D108	
	L117	L117	
	E119	E119	
	I122	I122	
	R124	R124	
	R127	R127	
	Y141	Y141	
	I19	I33	
RPS14	N24	N38	
	D25	D39	
	T26	T40	
	H29	H43	
	T31	T45	
	D32	D46	
	E37	E51	
	T38	T52	
	G45	G59	
	Y58	Y72	
	A60	A74	
	A63	A77	
	H80	H94	
	R84	R98	
	G87	G101	
	G88	G102	
	L105	L119	
	R107	R121	
	I112	I126	
	G113	G127	
	R114	R128	
	I115	I127	
	R127	R141	
	G19	G19	
RPS15	R43	R43	
	T78	T78	
	H79	H79	
	R81	R81	
	P87	P87	

Table S1 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
	G99	G99	
	M111	M111	
	H128	H128	
	G132	G132	
	L7	L7	
RPS15A	A17	A17	
	V25	V25	
	S31	S31	
	L38	L38	
	M41	M41	
	G100	G100	
	A18	A21	
RPS16	N32	N35	
	K47	K50	
	E50	E53	
	G71	G74	
	G73	G76	
	A80	A83	
	R82	R85	
	K44	K44	
RPS17	N48	N48	
	N21	N19	
RPS18	G24	G22	
	G37	G35	
	G39	G37	
	R41	R39	
	P82	P80	
	N87	N85	
	D91	D89	
	G95	G93	
	G135	G133	
	Q136	Q134	
	G142	G140	
	R143	R141	
	P47	P47	Diamond-Blackfan anemia(DBA)
RPS19	R63	R62	Diamond-Blackfan anemia(DBA)
	Y66	Y65	
	A136	A135	Diamond-Blackfan anemia(DBA)
	G54	G52	
RPS20	P59	P57	
	L63	L61	
	R68	R66	
	P71	P69	
	G73	G71	
	G75	G73	
	T78	T76	
	K88	K86	
	R89	R87	
	P14	P14	
RPS21	R15	R15	
	I24	I24	
	Q33	Q33	
	G44	G44	
	G57	G57	
	R60	R60	

Table S1 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
	K5	K3	
RPS23	G8	G6	
	A49	A47	
	K62	K60	
	Q63	Q61	
	P64	P62	
	N65	N63	
	S66	S64	
	A67	A65	
	R69	R67	
	K70	K68	
	A85	A83	
	F86	F84	
	P88	P86	
	D90	D88	
	G91	G89	
	N99	N97	
	D100	D98	
	V102	V100	
	G108	G106	
	G111	G109	
	G115	G113	
	D116	D114	
	P118	P116	
	G119	G117	
	L132	L130	
	H29	H29	
RPS24	G65	G65	
	G66	G66	
	G71	G71	
	F72	F72	
	Y76	Y76	
	E86	E86	
	R90	R90	
	K109	K108	
	K25	K33	
RPS25	K26	K34	
RPS25	W27	W35	
	K4	K4	
RPS26	R5	R5	
	G9	G9	
	R10	R10	
	G16	G16	
	C23	C23	
	D33	D33	
	K34	K34	Diamond-Blackfan anemia(DBA)
	D52	D52	
	P65	P65	
	R92	R92	
	K20	K20	
RPS27	P28	P28	
	S30	S30	
	C37	C37	
	C40	C40	
	S48	S48	

Table S1 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
	C56	C56	
	C59	C59	
	G69	G69	
	G76	G76	
RPS27A	K79	K79	
	K82	K82	
	K89	K89	
	L103	L103	
	C141	C141	
	G142	G142	
	G23	G25	
RPS28	Q27	Q29	
	N43	N45	
	G46	G48	
	D52	D54	
	L56	L58	
	E58	E60	
	E60	E62	
	R61	R63	
	E62	E64	
	A63	A65	
	R65	R67	
	R40	R40	
RPS29	F43	F43	
	G51	G51	
	K54	K54	
	A11	A83	
RPS30	G12	G84	
	K13	K85	
	V14	V86	
	G32	G104	
	G185	G187	
RPL3	F209	F211	
	D216	D218	
	G223	G225	
	G225	G227	
	R232	R234	
	L238	L240	
	K241	K243	
	G245	G247	
	R247	R249	
	K248	K250	
	C251	C253	
	G253	G255	
	P257	P259	
	R266	R268	
	G268	G270	
	G271	G273	
	R275	R277	
	G57	G59	
RPL4	A62	A64	
	S64	S66	
	W65	W67	
	G66	G68	
	G68	G70	

Table S1 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
	R69	R71	
	R73	R75	
	P75	P77	
	R76	R78	
	G79	G81	
	A89	A91	
	N92	N94	
	R95	R97	
	G97	G99	
	N114	N116	
	A121	A123	
	L134	L136	
	G139	G141	
	P149	P151	
	N221	N223	
	D72	D72	
RPL5	L83	L83	
	G87	G87	
	Y99	Y99	
	G102	G102	
	L103	L103	
	G37	G147	
RPL6	I41	I151	
	G45	G155	
	G49	G159	
	G67	G177	
	P68	G178	
	T87	T197	
	R88	R94	
RPL7	L65	L70	
RPL7A	P72	P77	
	F78	F83	
	Y97	Y102	
	G135	G139	
	V157	V161	
	P159	P163	
	P167	P171	
	L169	L173	
	C170	C174	
	A78	A78	
RPL8	A94	A94	
	N100	N100	
	P108	P108	
	G110	G110	
	E117	E117	
	G124	G124	
	G131	G131	
	G81	G83	
RPL9	V82	V84	
	H96	H98	
	F97	F99	
	P98	P100	
	D109	D109	
RPL10	43Q	46Q	
RPL11	53T	56T	

Table S1 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
	61R	64R	
	75K	78K	
	84L	87L	
	96F	99F	
	97S	100S	
	102F	105F	
	103G	106G	
	105G	108G	
	109H	112H	
	112L	115L	
	117D	120D	
	121G	124G	
	122I	125I	
	124G	127G	
	135G	138G	
	N66	N66	
RPL12	L85	L85	
	H100	H100	
	S120	S120	
	L133	L133	
	G134	G134	
	G140	G140	
	N112	N113	
RPL13	G146	G146	
RPL13A	M1	M1	
RPL15	E9	E9	
	K14	K14	
	S16	S16	
	R31	R31	
	A48	A48	
	G52	G52	
	Y53	Y53	
	G58	G58	
	R67	R67	
	G69	G69	
	K83	K83	
	A102	A102	
	E103	E103	
	R114	R114	
	S118	S118	
	Y119	Y119	
	D124	D124	
	E131	E131	
	D136	D136	
	R159	R159	
	R162	R162	
	G163	G163	
	T165	T165	
	K27	K27	
RPL17	G119	G118	
RPL18	Y89	Y84	
RPL18A	E101	E96	
	R103	R98	
	P159	P153	
	G73	G73	

Table S1 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
RPL19	L106	L106	
	T11	T11	
RPL21	R12	R12	
	G23	G23	
	V40	V40	
	G51	G51	
	Y57	Y57	
	G59	G59	
	T61	T61	
	G62	G62	
	G73	G73	
	H95	H95	
	S99	S99	
	K42	K48	
RPL22	Y78	Y85	
	90G	93G	
RPL23	97D	100D	
	K120	K134	
RPL23A	G63	G60	
RPL26	G16	G16	
RPL27	G20	G20	
	K111	K110	
RPL27A	L113	L112	
	No data available	G42	
RPL28	M1	M1	
RPL29	A2	A2	
	K3	K3	
	N6	N6	
	H17	H17	
	N19	N19	
	N42	N42	
	W32	W35	
RPL32	G37	G40	
	D39	D42	
	R43	R46	
	G56	G59	
	G27	G27	
RPL34	R74	R76	
	G77	G79	
	L17	L17	
RPL35	L31	L31	
	46G	44G	
RPL35A	69G	72G	
	79G	82G	
	85F	88F	
	G78	G78	
RPL36	H20	H18	
RPL36A	R41	R39	
	Y43	Y41	
	K46	K44	
	Q47	Q45	
	G49	G47	
	G51	G49	
	G52	G50	
	K55	K53	

Table S1 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Physiological Function
	C74	C72	
	C77	C77	
	R87	R87	
	Q30	Q30	
RPL37	W49	W49	
	K52	K52	
	G60	G60	
	G62	G62	
	E30	E30	
RPL37A	G53	G53	
	W55	W55	
	C57	C57	
	G67	G67	
	A68	A68	
	K5	K5	
RPL39	K10	K10	
	N20	N20	
	P24	P24	
	N38	N38	
	W44	W44	
	M1	M1	
RPL41	R2	R2	
	K4	K4	
	W5	W5	
	K7	K7	
	K8	K8	
	R9	R9	
	R11	R11	
	R12	R12	
	L13	L13	
	K16	K16	
	R17	R17	
	R18	R18	

Table S2. Synapomorphic *s. lat.* sites of Eukaryota in the currently available genomic data from Eukaryota.

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
RPSA	A14	V15	V/I/A/L	
	L17	F18	F/M/L	
	I48	I49	I/M/L/F	
	V50	L51	L/V/I	
	L57	L58	L/I/V/F	
	I64	I65	I/V/F/M/L	
	I67	I68	I/V	
	V73	V74	V/I/L	
	A75	V76	V/A	
	L 120	L 121	L/I/A/V	
	V 121	L 122	L/I/M/V	
	I 122	V 123	V/I/L	
	I133	L134	L/I/V	
	L146	L147	L/F/I	
	L176	M177	M/L/I/F	
RPS2	I50	I75	I/L/M	
	L61	L86	L/I/M	
	V63	I88	I/V/F	
	I69	I94	I/V	
	L72	F97	F/L/A	
	V103	A129	V/I/A	
	V104	I130	I/M/V	
	I139	V165	V/I	
	L154	V180	V/I/L	
	K161	R187	R/K	
	L187	L213	L/V/M/F	
	I218	I244	I/V/F/M/L/A	
RPS3	V12	V12	V/I/L	
	F24	F24	F/L/M/V	
	F25	L25	F/L	
	V48	I48	I/V/L	
	I84	V84	V/I/L/F	
	L86	L86	L/I	
	V123	L123	L/V/I	
	V126	I126	I/V/A/L	
	V138	V138	V/I/L	
	L157	M157	M/I/L	
	R178	R178	R/K/P	
	L182	L182	L/I/M	
	I184	I184	I/V	
	V206	V208	I/V	
RPS3A	I32	V33	V/I/M/L/F	
	K50	R51	R/K	
	M103	M103	M/F/L	
	F105	L105	L/F	
	M113	M113	M/L/V/I	
	V114	V114	V/I/M/F	
	I137	L137	L/V/M/I/A/F	
	L181	L181	L/I/M/V	
	L184	V184	L/V/I/A/F	
	V212	V212	V/I	
RPS4	V47	I47	F/L/I/M	
	L48	L48	L/I/V	
	N57	T57	N/T	
	V70	I70	I/V/L	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	V72	I72	V/I	
	I90	I90	I/V/F	
	L92	I92	I/L/M/V	
	L101	L101	L/V/I/M	
	V102	I102	I/V/L/M	
	I195	I195	I/V/L	
	I210	V210	V/I/M/L	
	I236	I236	V/I/L	
	I244	I244	V/I/L	
RPS5	I89	I61	I/L/V	
	I90	V62	V/I	
	L93	L72	L/V/F/M/I	
	I130	L109	L/I/V	
	V133	L112	L/V/A/F	
	I137	I116	V/I	
	L175	L154	L/I/M	
	I190	I169	I/V/M/F	
	L194	L173	L/I/V	
	L198	L177	L/I	
	I199	I178	I/V/M/L	
	L217	L196	L/I/V	
RPS6	L3	L3	L/V/I/F	
	I5	I5	I/L/V/F	
	F27	F27	F/L	
	I32	M32	M/I/L/V	
	F50	V50	F/L/M/V	
	L77	L77	L/V/F/M	
	K93	K93	K/R	
	F144	L144	F/L/M/V	
RPS7	V20	I23	V/I/L	
	F24	L27	F/L/I	
	L41	L43	I/L/F/V	
	I60	I62	I/V	
	F92	F94	M/L/I/V/F	
	L129	L133	M/V/I/L	
	K165	K169	R/K	
RPS8	I43	I43	I/V	
	I78	I78	I/V/L	
	I101	I101	I/V	
	V102	V102	V/I	
	I104	I104	I/V	
	I152	I158	I/V/L	
	V156	L162	L/I/V/M	
	L165	L171	L/I/V/F	
	I169	I175	I/V/L	
	I183	V189	I/V/L	
RPS9	K37	R38	R/K	
	I42	V43	V/I/L	
	I77	L78	L/I/M	
	V85	V86	V/L/F/I	
	96V	97I	I/V	
	L105	L106	L/M/F	
RPS10	M3	M3	M/V/L/I	
	V55	V58	V/I/L	
RPS11	V85	V87	I/V	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	I86	I88	V/I/M	
	I131	L134	L/I	
	V140	L143	L/I/V	
RPS12	L28	L18	L/V/I/M	
	L32	L22	L/I/F/M	
	L41	L31	L/V/I	
	K50	K40	R/K	
	L52	L42	L/I	
	L59	L49	L/M/I/V/F	
	V75	V65	V/I	
	V122	V110	V/A/L	
	V123	V111	V/I	
	I136	I124	L/I/F/A/V	
RPS13	M4	M4	M/L/V	
	R42	K42	K/R	
	V60	V60	V/I	
	I71	I71	I/V	
	L75	L75	L/M/V	
	K76	K76	K/R	
	I84	L84	L/I/V	
	I92	I92	I/V/M	
	V98	V98	V/I/M	
	T145	T145	T/N/Q	
	V150	V150	V/L/I	
RPS14	V28	V42	V/I	
	V81	I95	V/I	
	I83	L97	F/L/I/M	
	L110	M124	M/L/I/V	
	R111	K125	R/K	
RPS15	L22	L22	L/V/I	
	R77	K77	K/R/P	
	I85	I85	I/V	
	V94	V94	I/V/A	
	I96	V96	V/I	
	I107	I107	I/V/L/M	
RPS15A	I14	I14	M/I/L	
	K19	K19	K/R	
	I27	I27	I/L/V	
	I53	I53	V/I	
	I61	I61	I/V/A	
	V63	V63	V/I/L/A	
	L93	L93	L/I/F	
	V102	I102	V/I/L	
	I103	V103	V/I/L	
	I125	I125	I/V/L	
	V129	F129	F/V	
RPS16	S15	T18	T/N/S	
	I29	I32	I/L/V/M	
	V31	V34	V/L/F/I	
	I36	L39	L/I	
	V48	L51	V/L/I/A	
	L52	V55	V/L/I/F	
	L54	L57	L/V/I	
	I63	V66	V/L/I	
	I65	I68	I/M/V/F	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	V67	V70	V/I/L	
	V78	I81	I/V/A/L	
	I81	I84	I/V/A/L	
	I85	I88	I/L/M/V	
	L89	L92	L/I/V	
	L105	I108	I/L/V	
	K106	K109	K/R	
	I118	V121	V/I	
RPS17	L16	I16	V/I/L	
	I38	I38	M/L/I/V/F	
	S43	S43	S/T/N	
	L100	I96	L/I/V	
	L109	M105	M/L/I	
	V110	L106	F/L/I/V	
RPS18	L3	L3	L/I/F/M/V	
	L17	V15	L/V/I/M	
	V22	I20	I/V	
	L32	I3O	L/M/I	
	V38	V36	V/I/L	
	L45	V43	V/I/L/A/M	
	I69	V67	V/I/L/M	
	I72	I70	I/V/L	
	M73	M71	M/I/L/V/F	
	F85	F83	F/M/L	
	V105	L103	L/V/I	
	L116	L114	L/M/A	
	R132	R130	R/K	
RPS19	V6	V6	V/L/I/A	
	V9	V9	V/I/A	
	L28	L28	L/I/M/V	Diamond-Blackfan anemia(DBA)
	I65	L64	F/L/I/M/V	Diamond-Blackfan anemia(DBA)
	L108	L107	L/F/M	Diamond-Blackfan anemia(DBA)
	I124	L123	L/I/M	
	S125	T124	S/T	
	L132	L131	M/L/I/A/V	Diamond-Blackfan anemia(DBA)
RPS20	I41	L39	L/I/M	
	L58	M56	L/A/I/M	
	T66	T64	T/S	
	S76	S74	S/T	
	I108	I106	V/I/M/L	
RPS21	T20	S20	T/S	
	I23	I23	L/I/V	
	I34	M34	I/L/M	
	V36	V36	I/V/L/F	
	V39	V39	V/L/I	
	L55	I55	I/F/L/M	
	V59	I59	I/V/L	
	L69	I69	L/I/F	
RPS23	L57	V55	V/I/L/F/M	
	I59	V57	V/I	
	V83	I81	V/I	
	L104	V102	V/I/L	
RPS24	I7	I7	L/I/V	
	L40	I40	L/I/V/M	
	L44	L44	L/I/V/A/M	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	S69	T69	T/S	
	V75	I75	V/I/A	
RPS25	D34	D42	D/E	
	I50	L58	L/I/A/M/F	
	V60	I68	I/L/V	
	S61	T69	T/S	
	L65	V73	L/V/I	
	L80	L88	L/I/M	
	I92	V100	V/I/L	
	I100	I108	I/V/F	
RPS26	R38	K38	R/K	
	I44	I44	I/M	
	L64	L64	M/L/V/F	
	V75	V75	I/V	
	K93	K93	K/R	
RPS27	V25	V25	V/I	
	L33	M33	M/L/I	
	V46	V46	V/I/L/A	
	K70	K70	R/K	
RPS27A	I3	I3	I/V/L/M	
	V5	V5	V/I/L/F	
	I23	I23	I/V/L/A	
	V26	V26	V/L/I	
	R80	R80	R/K	
	L100	L100	L/M	
	V108	V108	V/L/I/F	
	V114	I114	I/V/A	
	L132	M132	M/L/I/A/F	
RPS28	V15	V17	V/L/I	
	L16	L18	L/I/M/V	
	R29	R31	R/K	
	V41	I43	I/M/L/V	
	I53	V55	I/V/M	
RPS29	L30	L30	L/V/I/M	
RPS30	V4	V76	V/I/A/M	
	K15	R87	R/K	
RPL3	V158	I160	F/L/I/V	
	V160	V162	L/F/V/I/A	
	L161	I163	L/V/I/M	
	L178	L180	F/L/I/V	
	L183	V185	L/I/V	
	V192	L194	L/I/V	
	V203	V205	L/V/I/A/F	
	V205	V207	V/I	
	R244	R246	K/R	
	T306	N315	T/N/S	
	M308	L317	L/I/V/M	
	I317	V326	V/I	
	I322	V331	L/I/V	
	V324	L333	M/L/I/V/F	
	V336	L345	M/L/I/V	
	L338	L347	L/I/M	
	V354	I363	V/I	
	E375	E384	D/E	
RPL4	V6	I8	V/A/I	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	I30	I32	M/L/I/V	
	V42	L44	M/L/I/V	
	K108	R110	R/K	
	K118	K120	R/K	
	I126	L128	L/I/V	
	T129	S131	T/S	
	V135	V137	L/I/V	
	V142	I144	L/I/V	
	I148	L150	F/M/L/I/V	
	V152	V154	F/L/I/V	
	A168	L170	F/L/I/A/V	
	R202	R204	R/K	
	V207	I209	V/I	
	V208	I210	V/I	
	L235	I237	L/I	
	L238	L240	M/L/V	
	F247	F249	F/L/V	
	I249	I251	V/I/M	
	I281	M284	L/I/V/F/M	
	V286	L289	F/A/L/I/V	
RPL5	T56	T56	T/S	
	I60	I60	L/V/I	
	I64	I64	F/I/V	
	I69	I69	I/F/L	
	I88	V88	V/I/L/M	
	L115	M115	M/I/L/V	
	T154	T154	T/N/S	
	R158	K158	K/R	
	V159	V159	V/I/M	
	L163	L163	L/M/I	
	L171	L171	L/V/I/M/F	
	V173	I173	V/I	
	L211	L211	L/V/M	
RPL6	V39	I149	V/I	
	L42	I152	L/I/V/M	
	V53	V163	V/I	
	V65	V175	V/I	
	L76	L186	L/I/V/M	
	V91	I201	F/L/I/V	
	F122	F233	F/L/M	
	L145	I257	L/I/V	
	I149	I261	L/I/V/A	
RPL7	L83	L89	L/V/F	
	V86	V92	A/L/I/V	
	I95	V101	L/I/V/M	
	L103	L109	F/L/I/M	
	L106	L112	F/L/M	
	V117	V123	F/M/L/I/V	
	T120	N126	T/N/S	
	V134	I140	V/I	
	I144	V150	F/I/V	
	V148	I154	L/I/V	
	L163	L169	L/I	
	I169	I175	V/I	
	I178	I184	L/I/V	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	L184	L190	F/M/L/I/V	
	I185	I191	A/L/I/V	
RPL7A	V56	I61	M/V/I	
	L58	L63	L/I/M/V	
	V71	V76	V/I	
	I75	I80	L/I/V	
	L82	L87	A/F/L/I/V	
	L93	L98	F/M/L/I/V	
	V132	L136	A/L/I/V	
	I143	V147	M/L/I/V	
	L150	L154	F/M/L/I/V	
	L152	V156	A/L/I/V	
	I160	I164	L/I/V	
	I180	I184	L/I/V	
	A198	V202	L/I/V/A	
	L200	F204	F/L/I/V	
	L214	L218	F/L/I/V	
RPL8	I41	I41	L/I/V	
	L58	L58	L/M/A/V	
	V61	V61	M/L/V	
	I77	I77	L/I/V	
	I88	V88	L/I/V	
	L104	V104	L/I/V	
	I136	V136	L/I/V	
	I137	I137	L/I/V	
	V148	V148	L/I/V	
	I158	I158	F/M/L/I/V	
	I166	V166	L/I/V	
	I169	V169	V/I/A	
	L179	I179	F/M/L/I/V	
RPL9	I4	I4	L/I/V	
	I10	V10	F/M/L/I/V	
	V25	V25	F/I/V	
	L38	F38	F/L/V	
	F45	L45	F/A/M/L/I/V	
	I53	L55	F/L/I/V	
	V55	V57	F/L/I/V	
	V75	V77	L/I/V	
	M78	M80	L/M	
	I79	I81	F/A/M/L/I/V	
	I112	V112	A/L/I/V	
	V114	I114	V/I	
	L118	L118	F/L/M	
	I144	L144	L/I/V	
	L146	L146	F/L/I/M	
	V151	I151	L/I/V	
RPL10	F159	F159	F/L/I/V	
RPL10A	F68	L68	L/I/M/F	
	V120	I120	F/I/V	
	L159	M159	F/L/M	
	V169	V169	L/I/V	
	V191	V191	V/I/A	
RPL11	L12	L15	L/I/V	
	I14	L17	L/I/V	
	L19	L22	L/I/V	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	V54	V57	L/I/V	
	V71	V74	V/I/L	
	I79	I82	F/L/I/V	
	M125	L128	M/L	
	M131	L134	L/M	
	V138	I141	V/I	
	R141	K144	K/R	
	V148	I151	V/I	
RPL12	I32	I32	A/L/I/V	
	L37	L37	L/V/M/I	
	V42	V42	V/I	
	V58	I58	V/I/A	
	S76	S76	T/S	
	V81	I81	L/I/V	
	I112	I112	V/I	
	I132	I132	V/I/M	
	I155	I155	F/L/I/V	
RPL13	I9	V9	F/L/I/V	
	V22	V22	L/I/V	
	L54	I55	L/I/V	
	K64	R65	K/R	
	R73	R74	K/R	
	T76	S77	T/S	
	L85	I86	F/M/L/I/V	
	I123	L124	F/L/I/V	
RPL13A	F80	F80	F/L/V	
	L138	L138	M/L/I/V	
	L141	L141	F/M/L/I/V	
	V145	V145	F/M/V/I	
RPL14	V81	V73	L/A/V	
	R109	K101	R/K	
RPL15	L7	I7	M/L/I/V	
	L10	L10	F/L/I/V	
	L22	L22	M/L/I/V	
	V25	V25	M/L/I/V	
	T43	T43	T/S	
	K54	K64	K/R	
	I61	I61	V/I	
	V66	V66	M/I/V	
	V89	V89	V/I	
	V121	V121	L/V/I	
	V135	I135	V/I	
RPL17	I36	I36	V/I	
	L41	I41	F/A/L/I/V	
	I58	V58	L/V/I	
	L94	M94	M/L/I/A	
	V114	I114	L/I/V	
	L148	M148	F/M/L/I/V	
RPL18	L32	L31	F/A/L/I/V	
	T40	T39	T/N/S	
	V48	V47	M/L/I/V	
	V64	L63	M/L/I/V	
	I67	M66	M/L/I/V	
	V87	I86	V/I	
	V101	V100	V/I	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	T107	T106	T/S	
	L124	F123	L/I/F	
	L140	L139	F/M/L/I/V	
RPL18A	L32	L27	F/A/L/I/V	
	I66	I61	M/L/I/V	
	V67	V62	F/L/I/V	
	V80	V75	V/I/A	
	V85	I80	V/I	
	V87	L82	F/L/I/V	
	K100	R95	K/R	
	D117	D112	D/E	
	L132	M127	L/I/M	
	I137	I132	L/I/V	
	L156	I150	M/A/L/I/V	
RPL19	K21	K21	K/R	
	I96	M96	L/I/V/M	
	V139	M139	L/I/V/M	
RPL21	R20	R20	R/K	
	I42	I42	V/I	
	V64	V64	V/A/I	
	V67	V67	V/I	
	V72	V72	L/I/V	
	V74	I74	V/I	
	I76	V76	L/I/V	
	L89	I89	L/I/V	
	L91	V91	A/L/I/V	
	I96	I96	F/L/I/V	
RPL22	L50	L56	F/L/V	
	V63	I70	L/I/V	
	I93	L100	L/I/F	
	L105	L112	F/L/V/M	
	V110	I117	F/L/I/V	
RPL23	M57	M60	A/M/L/I/V	
	V58	V61	F/L/I/V	
	M59	M62	F/M/L/V	
	T61	T64	T/S	
	I77	V80	L/I/V	
	V78	V81	V/I	
	L93	L96	M/L/I/V	
	M109	M112	L/V/M	
	V136	I139	V/I	
RPL24	V22	A22	V/I/A	
	R71	R71	K/R	
	I90	I90	M/I/V	
RPL26	I52	I49	L/I/V	
	V60	V57	V/I	
	I72	V70	V/I	
	V82	I80	L/V/I	
	L101	I99	L/I/F/V	
	I108	I106	F/I/V	
RPL27A	L75	L75	L/I/V	
	L78	L78	F/L/I/V	
	V101	I100	L/I/V	
	I102	I101	L/I/V	
	I112	V111	L/V/I	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	I118	L117	L/I/F	
	I124	I123	F/I/V	
	V130	F129	A/F/I/V	
RPL27	V10	V10	A/F/L/I/V	
	V12	L12	F/M/L/I/V	
	V24	V24	F/L/I/V	
	I25	I25	V/I	
	V26	V26	A/M/L/I/V	
	I46	I46	L/I/V	
	I72	V72	L/I/V	
	L87	V87	A/F/L/I/V	
RPL28	No data available	E28	D/E	
	No data available	V51	M/L/I/V	
	No data available	I93	A/L/I/V	
RPL29	T8	T8	T/S	
	R18	R18	K/R	
RPL30	L34	I38	L/I/M	
	I42	V46	V/I	
	I43	I47	F/L/I/V	
	I44	L48	F/M/L/I/V	
	L56	I60	L/I/V	
	V90	L94	F/M/L/I/V	
	I92	I96	M/I/V	
RPL31	I14	I27	V/I	
	L16	I29	A/M/L/I/V	
	V33	L46	L/I/V	
	I36	I49	L/I/V	
	L51	I64	L/I/V	
	L71	I84	A/M/L/I/V	
	L73	V86	L/I/V	
	I75	L88	F/L/I/V	
RPL32	V28	I31	L/I/V	
	K64	K67	K/R	
	V76	V79	V/I/A	
	D81	E84	D/E	
	L82	L85	M/L/I/V	
	V119	V122	L/I/V	
RPL34	K24	R24	K/R	
	L51	L53	L/I	
	V85	V87	L/I/V	
RPL35	I51	I54	V/I	
	V57	V60	V/I	
	I118	V121	L/I/V	
RPL35A	I30	L28	L/I/V	
	I32	I30	L/I/V	
	V52	V50	L/I/V/M	
	V71	V74	V/I	
	I100	V103	V/I	
	I107	I110	V/I	
RPL36	I61	L61	L/I	
	R62	K62	K/R	
	R76	R76	R/K	
	L77	V77	L/I/V	
	M90	L90	L/I/M	
RPL36A	R71	R69	R/K	

Table S2 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Type	Physiological function
	L72	L70	F/L/M	
	L85	I85	L/I/M	
RPL37	L67	L67	L/M	
RPL37A	I10	I10	L/I/V	
	R80	K80	K/R	
RPL38	I5	I5	F/I/V	
	V45	V36	L/I/V	
	L65	L56	L/I/V	
RPL39	L23	I23	M/L/I/V	
RPL40	R106	R106	R/K	
	L121	L121	L/I	
RPL41	R6	R6	R/K	
RPP0	L19	L21	F/L/I/A/M	
	F26	F28	L/I/V/A/F	
	V27	I29	F/L/I/V	
	L45	L47	F/A/L/I/M	
	M53	M55	F/A/L/I/M	
	I63	I65	M/L/I/V	
	V87	V89	V/I/A	
	V121	V123	L/I/V	
	V141	I143	V/I	
	R147	R149	R/K	
	V158	L160	L/I/V	
	L185	L187	M/L/I/V	
	I204	I206	M/L/I/V	
	L209	L211	M/L/I/V	
	I220	V222	F/L/I/V	
	I223	V225	F/M/L/I/V	
	L239	I241	A/F/L/I/V	
	L247	L249	A/F/M/L/I/V	
	V249	L251	A/L/I/V	
RPP1	I13	I15	A/M/L/I/V	
	L26	I28	L/I/M	
	L49	L51	A/F/L/I/V	
	L58	I60	F/M/L/I/V	
RPP2	L4	V4	L/I/V	
	I22	I23	L/I/V	
	I25	I26	L/I/V	
	V38	L39	A/M/L/I/V	

Table S3. Group-specific sites in the large subunit of RP according to the currently available genomic data from Eukaryota.

RPG	Sites in *S. cerevisiae*	Positon in *Homo sapiens*	Position in alignment matrix	Homologous autapomorphic amino acids
RPL3	L14		113	M/Ecdysozoa
	L17		116	Y/Insecta
	T55		155	T/Pezizomycotina+Saccharomycotina
	V73		174	C/Viridiplantae
	S134		236	C/Chordata
	Y137		239	W/Metazoa
	I144		248	L/Vertebrata
	E197		305	G/Pezizomycotina
	H259		367	H/Pezizomycotina+Saccharomycotina
	R369		500	A/Pezizomycotina
RPL4	M1-S2		82	C/Chordata
	G78		166	R/Metazoa
	E157		246	F/Pezizomycotina
	H175		267	W/Bilateria
	L182		274	S/Basidiomycota
	L206		298	C/Deuterostomia
	G214		307	G/Ascomycota+Basidiomycota
	S293		390	S/Pezizomycotina+Saccharomycotina
	S293		390	P/Theria
	G303		401	G/Ascomycota+Basidiomycota
	E337		437	T/Amniota
	G340		440	G/Pezizomycotina+Saccharomycotina
RPL5	A11		521	Q/Neoptera
	F55		846	V/Tetrapoda
	W95		886	W/Ascomycota+Basidiomycota
	K143		953	T/Vertebrata
	S167		977	S/Ascomycota+Basidiomycota
	Y172		982	S/Vertebrata
	F185		1003	P/Basidiomycota
	H203		1025	N/Vertebrata
	E221		1046	S/Pezizomycotina
	F253		1080	Y/Mammalia
	T256		1083	E/Pezizomycotina
	K259		1088	K/Pezizomycotina+Saccharomycotina
	F260		1089	K/Pezizomycotina
RPL6		49	175	V/Tetrapoda
		52	178	R/Vertebrata
		88	272	V/Theria
	T2		294	K/Vertebrata
	A6		298	M/Chordata
	P18--K19		312	L/Tetrapoda
	K20		314	H/Chordata
	T21		315	G/Chordata
	D60		357	Q/Pezizomycotina
	T62		359	V/Pezizomycotina
	F69		366	L/Vertebrata
	V79		376	T/Mammalia
	R82		379	K/Vertebrata
	P123		441	K/Pezizomycotina
RPL7	P99		346	T/Viridiplantae
RPL7A	A76		910	A/Ascomycota+Basidiomycota
	K122		964	K/Pezizomycotina+Saccharomycotina
	Q123		965	E/Pezizomycotina
	D124		966	D/Ascomycota
	Y130		972	Y/Ascomycota

Table S3 (continued)

RPG	Sites in *S. cerevisiae*	Positon in *Homo sapiens*	Position in alignment matrix	Homologous autapomorphic amino acids
	A205		1047	K/SAR
	A256		1099	A/Pezizomycotina+Saccharomycotina
RPL8	N7		35	N/Pezizomycotina+Saccharomycotina
	L22		51	L/Ascomycota
	S52		83	A/Basidiomycota
	N79		110	N/Pezizomycotina+Saccharomycotina
	N115		146	C/Vertebrata
	V168		199	Q/Viridiplantae
	G183		214	S/Pezizomycotina
	F186		217	K/Pezizomycotina
	T199		230	T/Ascomycota+Basidiomycota
	S231		262	Q/Pezizomycotina
RPL9	Y92		173	Y/Pezizomycotina+Saccharomycotina
	E105		186	K/Pezizomycotina
	K106		189	E/Pezizomycotina
	G108		191	G/Pezizomycotina+Saccharomycotina
	T138		221	Q/Tetrapoda
	D160		245	D/Pezizomycotina+Saccharomycotina
RPL10	V50		50	G/Tetrapoda
	K141		141	S/Neoptera
	S143		143	R/Neoptera
	N209		209	W/Metazoa
	R211		211	A/Vertebrata
	F213		213	F/Ascomycota
	P214		214	P/Ascomycota
RPL10A	C80		235	V/Tetrapoda
	N181		338	Y/Vertebrata
	G211		382	S/Pezizomycotina
RPL11		3	10	Q/Amniota
	S2		29	Q/Amniota
	T147		177	C/Mammalia
	L172		207	R/Pezizomycotina
RPL12	G102		176	K/Pezizomycotina
RPL13	N12		32	P/Vertebrata
	H25		51	H/Pezizomycotina+Saccharomycotina
	D52		79	D/Pezizomycotina+Saccharomycotina
	A82		110	V/Theria
	I147		204	L/Basidiomycota
	L172		241	M/Chordata
RPL13A	E46		117	G/Viridiplantae
	V153		226	T/Vertebrata
	A165		238	Q/Basidiomycota
		202	288	L/Vertebrata
RPL14	W12		62	W/Pezizomycotina+Saccharomycotina
	L14		64	L/Pezizomycotina+Saccharomycotina
	Q62		118	C/Vertebrata
	T78		135	Q/Vertebrata
		135	213	A/Basidiomycota
		139	217	S/Theria
		141	219	K/Theria
		142	220	K/Theria
		176	282	Q/Mammalia
RPL15	Y4		16	L/Pezizomycotina
	Y30		42	L/Pezizomycotina
	T80		92	V/SAR+Cryptophyta+Viridiplantae

Table S3 (continued)

RPG	Sites in *S. cerevisiae*	Positon in *Homo sapiens*	Position in alignment matrix	Homologous autapomorphic amino acids
	N86		98	R/Basidiomycota
	N90		103	T/Viridiplantae+Cryptophyta
	A111		125	A/Pezizomycotina+Saccharomycotina
	A185		200	A/Pezizomycotina+Saccharomycotina
RPL17	A6		269	L/Vertebrata
	T9		274	I/Pezizomycotina
	Q34		306	A/Basidiomycota
	W39		311	W/Pezizomycotina+Saccharomycotina
	I67		339	T/Pezizomycotina
	A81		358	A/Ascomycota
	K124		408	C/Neoptera
	Q125		409	T/Basidiomycota
	S141		426	G/Basidiomycota
	V166		454	Q/Vertebrata
	R171		459	R/Pezizomycotina+Saccharomycotina
RPL18	R141		306	S/Vertebrata
RPL18	G156		323	G/Ascomycota+Basidiomycota
	P157---H158		326	T/Vertebrata
RPL18A	R21		71	C/Amniota
	F38		112	W/Viridiplantae
	E65		169	Q/Viridiplantae
	D90		233	Q/Viridiplantae
	A124		268	V/Viridiplantae
	K139		283	K/Ascomycota+Basidiomycota
	Y147		291	A/Vertebrata
	V164		315	H/Holometabola
RPL19	E72		1155	I/Pezizomycotina
	Q92		1197	Q/Pezizomycotina+Saccharomycotina
	E126		1262	L/Pezizomycotina
	H134		1274	H/Ascomycota
	A137		1280	A/Ascomycota
	R172		1364	K/Alveolata
	L184		1376	R/Viridiplantae
			1398	202(*A.thaliana*)-A/Viridiplantae
RPL21	K3		146	N/Bilateria+Cnidaria
	M14		157	A/Pezizomycotina
	V36		179	K/Chordata
	N45		188	M/Tetrapoda
	F56		199	C/Vertebrata
	E104		302	S/Chordata
	T158		388	T/Ascomycota+Basidiomycota
RPL22			270	4(*A.thaliana*)-G/Viridiplantae
	G73		423	G/Ascomycota
	L89		439	V/Viridiplantae
			452	105(*A.thaliana*)-R/Viridiplantae
			453	106(*A.thaliana*)-N/Viridiplantae
RPL23	F11		63	L/Pezizomycotina
	R32		84	R/Pezizomycotina+Saccharomycotina
	S133		186	S/Pezizomycotina+Saccharomycotina
	V135		188	V/Pezizomycotina+Saccharomycotina
RPL23A	M1		161	E/Theria
	S4-A5	16	227	E/Vertebrata
RPL24	Q42		53	S/Vertebrata
	R48		59	Q/Vertebrata
	Q141---Q142	142	164	I/Vertebrata

Table S3 (continued)

RPG	Sites in *S. cerevisiae*	Positon in *Homo sapiens*	Position in alignment matrix	Homologous autapomorphic amino acids
RPL26	K110		133	R/Vertebrata
	K118		141	E/Pezizomycotina
	G125----------G126		154	Q/Vertebrata
RPL27	L5		75	Y/Basidiomycota
	K55		178	A/Vertebrata
	S94-----V95		224	94(*A.thaliana*)-V/Viridiplantae
	V95		226	K/Insecta
	Q106		239	Q/Ascomycota+Basidiomycota
RPL28		13	21	C/Tetrapoda
		34	49	A/Vertebrata
		38	53	R/Pezizomycotina
		41	56	A/Pezizomycotina
		60	81	G/Basidiomycota
		69	91	G/Vertebrata
		75	97	T/Tetrapoda
		82	107	L/Basidiomycota
		83	108	K/Basidiomycota
		133	178	P/Theria
RPL29	V35		64	T/Pezizomycotina
	H43		76	H/Pezizomycotina+Saccharomycotina
	H45		86	H/Pezizomycotina+Saccharomycotina
	A57		98	M/Tetrapoda
		60	101	N/Tetrapoda
		75	116	L/Theria
		86	127	P/Theria
		115	157	G/Theria
		147	201	K/Theria
		154	208	K/Theria
RPL30	S33		116	M/Chordata
	V81		165	A/Basidiomycota
		111	191	Q/Vertebrata
		112	192	T/Amniota
		114	196	E/Vertebrata
		115	197	K/Vertebrata
RPL31		9	2296	E/Vertebrata
	K5		2308	A/Pezizomycotina
	T9		2312	A/Pezizomycotina
	E54		2358	Q/Pezizomycotina
	K61		2365	E/Pezizomycotina
RPL32	P5		37	Y/Insecta
	H88		123	C/Mammalia
RPL34	R16		59	A/Vertebrata
RPL34	K43--C44	44	87	S/Vertebrata
	K43—C44	45	88	A/Vertebrata
	G48		93	P/Vertebrata
	Q61		106	V/Vertebrata
	R80		129	M/Vertebrata
	V90		139	K/Vertebrata
RPL35	R10		38	W/Pezizomycotina
	S13		41	D/Insecta
	T46		79	D/Pezizomycotina
	E60		93	A/Pezizomycotina
	A65		98	Q/Pezizomycotina
RPL35A	H5		146	H/Pezizomycotina+Saccharomycotina
	H13		154	H/Pezizomycotina+Saccharomycotina

Table S3 (continued)

RPG	Sites in *S. cerevisiae*	Positon in *Homo sapiens*	Position in alignment matrix	Homologous autapomorphic amino acids
	K57--E58	56	199	N/Tetrapoda
	G61--------S62	64	212	K/Pancrustacea
RPL36	K4		5	E/Pezizomycotina
	L43		67	M/Tetrapoda
		105	132	D/Pezizomycotina
RPL36A	K78		191	R/Vertebrata
	F106		222	F/Fungi+Metazoa
RPL37	R24		112	S/Vertebrata
	S35		123	G/Vertebrata
	A86		179	G/Pezizomycotina
RPL37A	S58--C59	58	152	G/Vertebrata
	C59		153	R/Insecta
	K61		155	M/Vertebrata
RPL38	V55		177	C/Viridiplantae
RPL39	K12		180	F/Vertebrata
	R46		214	K/Pezizomycotina
	K48		216	R/Pezizomycotina
RPP0	K7		80	W/Metazoa
	A9		82	E/Basidiomycota
	R48		131	E/Pezizomycotina
	V99		205	M/Tetrapoda
	G131		237	E/Metazoa
	P142		248	T/Vertebrata
	V156		262	L/Pezizomycotina
	D160		266	E/Pezizomycotina
	P198		304	N/Amniota
	D245		351	K/Pezizomycotina
	D245		351	R/Vertebrata
	Y257		364	W/Pezizomycotina
	R266		385	R/Pezizomycotina+Saccharomycotina
RPP1	F45		163	L/Basidiomycota

Table S4. Group-specific sites in the small subunit of RP according to currently available genomic data from Eukaryota.

	Position in *S. cerevisiae*	Position in *H.sapiens*	Position in Alignment	Homologous autapomorphic amino acids
RPSA	Q15		90	E/Pezizomycotina
	H32		107	H/Pezizomycotina+Saccharomycotina
	V37		112	L/Pezizomycotina
	V74		149	F/Neoptera
	F102		178	H/Viridiplantae
	R113		189	R/Fungi
	E180		256	M/Viridiplantae
	E209		289	A/Pezizomycotina
	E221		304	P/Pezizomycotina
	E222		305	G/Pezizomycotina
RPS2	R4		44	D/Vertebrata
	G48		130	M/Vertebrata
	F66		148	S/Vertebrata
	N108		198	E/Pezizomycotina
	N147		237	N/ Pezizomycotina+Saccharomycotina
	S153		243	S/ Pezizomycotina+Saccharomycotina
	L154		244	L/ Pezizomycotina+Saccharomycotina
	T158		248	E/Pezizomycotina
	K161		251	R/Vertebrata
	Q189		279	M/Vertebrata
	Q250		351	G/Pezizomycotina
	K251		352	K/ Pezizomycotina+Saccharomycotina
	R253		354	Y/Pezizomycotina
RPS3	T26		38	Q/ Pezizomycotina
	P43		56	H/Basidiomycota
	K45		58	V/ Pezizomycotina
	E47		60	D/ Pezizomycotina
	N62		75	Q/ Pezizomycotina
	V85		127	S/ Pezizomycotina
	E89		131	A/ Pezizomycotina
	D166		209	Y/Vertebrata
	E215		262	Q/ Pezizomycotina
	V222		269	I/Tetrapoda
	K223		270	Q/ Pezizomycotina
	V238		306	R/Pezizomycotina
RPS3A	N56		65	N/Ascomycota+Basidiomycota
	H79		92	R/Pancrustacea
	D132		145	D/Pezizomycotina+Saccharomycotina
	Q146		159	D/Insecta
	T185		201	T/ Pezizomycotina+Saccharomycotina
	L217		247	M/Vertebrata
	L217		247	L/ Pezizomycotina+Saccharomycotina
	G235		277	Y/Viridiplantae
RPS4	L207		229	V/Vertebrata
	L222		244	E/Pezizomycotina
	R256		279	R/ Ascomycota+Basidiomycota
RPS5	T146		173	A/Viridiplantae
RPS6	P86		95	G/Viridiplantae
	L133		150	R/Viridiplantae
	F144		161	F/ Ascomycota+Basidiomycota
	I158		184	Y/Viridiplantae
	T163		189	N/Amniota
	A224		255	S/Viridiplantae
	I226		280	S/Viridiplantae
RPS7	A6		22	N/Pezizomycotina

Table S4 (continued)

	Position in *S. cerevisiae*	Position in *H.sapiens*	Position in Alignment	Homologous autapomorphic amino acids
	K7		23	R/Basidiomycota
	Q42		106	N/Tetrapoda
	D87		155	D/Ascomycota+Basidiomycota
	R104		172	R/Ascomycota+Basidiomycota
	-	R109	176	S/Pezizomycotina
	-	T110	177	R/Pezizomycotina
	V144		216	E/Pezizomycotina
	S156		228	E/Pezizomycotina
	Q160		234	G/Pezizomycotina
	I162		236	T/Viridiplantae
RPS8	G39		80	S/Viridiplantae
RPS9	R6		8	W/Vertebrata
	T7		9	S/Pezizomycotina
	S46		48	K/Metazoa
	R54		56	N/Viridiplantae
	T61		63	K/Basidiomycota
	I77		79	I/ Ascomycota+Basidiomycota
	V81		83	N/Viridiplantae
	K92		102	R/Basidiomycota
	D103		113	K/Stramenopila+Alveolata
	Q112		122	C/Pezizomycotina
	G137		147	R/Metazoa
RPS10	K10		21	E/Viridiplantae
	H12		23	H/Ascomycota
	Y14		25	L/Vertebrata
	Q29		49	M/Tetrapoda
	T57		85	T/ Ascomycota+Basidiomycota
	92		133	I/ Pezizomycotina+ Saccharomycotina
	P97		138	S/Pezizomycotina
		E112	155	Y/Basidiomycota
RPS11	K26		51	R/Pezizomycotina
	T31		56	L/Chordata
	K32		57	P/Vertebrata
	S143		171	L/Pezizomycotina
	A145		173	R/Pezizomycotina
	A146		174	G/Viridiplantae
RPS12	V30		97	G/Pezizomycotina
	R43		110	H/insecta
	N70		137	D/Viridiplantae
	G117		186	N/Pezizomycotina
	M137		207	E/Vertebrata
	Q142		212	C/Amniota
	Q143		213	E/Pezizomycotina
	Q143		213	R/Basidiomycota
RPS13	I11		16	L/Chordata
	S19		24	S/ Pezizomycotina+Saccharomycotina
	A22		27	V/Metazoa
	S32		37	Q/Pezizomycotina
	L45		51	A/Pezizomycotina
RPS15	A37		62	Y/Chordata
	A54		80	H/Amniota
	N98		126	S/Pezizomycotina
RPS15A	N15		40	Y/Viridiplantae
	A108		133	Y/Alveolata
RPS16	K102		157	A/Basidiomycota

Table S4 (continued)

	Position in *S. cerevisiae*	Position in *H.sapiens*	Position in Alignment	Homologous autapomorphic amino acids
RPS17	G64		78	S/Pancrustacea
	P65		79	Q/Insecta
	R128		147	V/Vertebrata
		T130	153	T/Vertebrata
RPS18	K8		41	Q/Pezizomycotina
	G9		42	Q/Basidiomycota
	T62		95	Q/Basidiomycota
RPS19	K69		101	S/Arthropoda
	Q70		104	T/Pezizomycotina
	Q70		104	P/Ecdysozoa
	N101		136	D/Pezizomycotina
	N127		167	D/Basidiomycota
	R134		174	Q/Viridiplantae
	Q138		178	R/Viridiplantae
	Q138		178	T/Pezizomycotina
RPS20	T104		120	I/Pezizomycotina
RPS21	I34		49	M/Theria
	A37		52	S/Sauropsida
	D67		83	D/Fungi+Metazoa
	W83		104	W/Fungi
RPS23	P6		111	C/Vertebrata
	L15		120	H/Arthropoda
	A37		142	E/Viridiplantae
	S41		146	K/Viridiplantae
	S41		146	S/Fungi
	F43		148	T/Basidiomycota
	K142		281	R/Metazoa
RPS24	L74		92	M/Vertebrata
	F85		103	N/Vertebrata
	Q106		124	Q/Ascomycota
	Q110		128	Q/Ascomycota
	L125		143	N/Tetrapoda
RPS25	Q5		5	D/Vertebrata
		K8	11	S/Arthropoda
	A10		16	G/Vertebrata
	A16		22	D/Vertebrata
	A16		22	K/Ecdysozoa
		D21	25	D/Vertebrata
		V23	27	V/Vertebrata
		N24	28	N/Vertebrata
	A39		51	Q/Neoptera
	L51		64	C/Tetrapoda
	P55		68	Q/Pezizomycotina
	V62		78	Q/Basidiomycota
	V66		82	I/Basidiomycota
	G72		88	R/Metazoa
	E108		132	G/Tetrapoda
RPS26	A81		85	G/Pezizomycotina
	V86		90	N/Metazoa
	A118		143	M/Vertebrata
RPS27	L3		5	L/Fungi+Metazoa
	P38		43	Q/Viridiplantae
RPS27A	V147		154	C/Vertebrata
RPS28		R5	5	R/Vertebrata
	K4		8	V/Vertebrata

Table S4 (continued)

	Position in *S. cerevisiae*	Position in *H.sapiens*	Position in Alignment	Homologous autapomorphic amino acids
	K14		20	R/Pezizomycotina
	R22		28	Q/Metazoa
	V25		31	C/Metazoa
	V41		52	M/Viridiplantae
RPS29	N5		9	S/Pezizomycotina
	F8		12	W/Vertebrata
RPS30		L54	58	V/Insecta

Table S5. Proportion of sites under negative selection in the RP large subunit.

RPG	Proportion	Protein length	Number of negative selection
RPL3	0.90339	543	490.54077
RPL4	0.67578	766	517.64748
RPL5	0.71337	1171	835.35627
RPL6	0.83757	408	341.72856
RPL7	0.74483	430	320.2769
RPL7A	0.93644	1027	961.72388
RPL8	0.97395	335	326.27325
RPL9	0.93095	272	253.2184
RPL10	0.85756	230	197.2388
RPL10A	0.81398	447	363.84906
RPL11	0.87873	272	239.01456
RPL12	0.82579	313	258.47227
RPL13	0.95748	361	345.65028
RPL13A	0.9713	293	284.5909
RPL14	0.75224	329	247.48696
RPL15	0.99329	234	232.42986
RPL17	0.99503	222	220.89666
RPL18	0.9802	550	539.11
RPL18A	0.91241	325	296.53325
RPL19	0.93092	1816	1690.55072
RPL21	0.92334	626	578.01084
RPL22	0.73594	494	363.55436
RPL23	0.97914	190	186.0366
RPL23A	0.50896	436	221.90656
RPL24	0.88296	250	220.74
RPL26	0.96274	189	181.95786
RPL27	0.97028	346	335.71688
RPL27A	0.79138	269	212.88122
RPL28	0.97874	200	195.748
RPL29	0.31355	290	90.9295
RPL30	0.86342	275	237.4405
RPL31	0.94178	2628	2474.99784
RPL32	0.98075	167	163.78525
RPL34	0.90211	226	203.87686
RPL35	0.95349	165	157.32585
RPL35A	0.70858	281	199.11098
RPL36	0.80535	142	114.3597
RPL36A	0.77907	250	194.7675
RPL37	0.65936	274	180.66464
RPL37A	0.59253	197	116.72841
RPL38	0.50423	224	112.94752
RPL39	0.50173	236	118.40828
RPL40	0.71468	634	453.10712
RPL41	0.56158	67	37.62586
RPP0	0.93219	459	427.87521
RPP1	0.9048	399	361.0152
RPP2	0.9295	161	149.6495
	0.844985399	20419	17253.75687

Table S6. Proportion of sites under negative selection in the RP small subunit.

RPG	Proportion	Protein length	Number of negative selection
RPSA	0.60162	476	286.37112
RPS2	0.83265	353	293.92545
RPS3	0.76562	331	253.42022
RPS3A	0.94452	306	289.02312
RPS4	0.92199	304	280.28496
RPS5	0.81781	252	206.08812
RPS6	0.83726	502	420.30452
RPS7	0.97313	274	266.63762
RPS8	0.90548	615	556.8702
RPS9	0.95371	220	209.8162
RPS10	0.91583	337	308.63471
RPS11	0.88937	192	170.75904
RPS12	0.93372	222	207.28584
RPS13	1	188	188
RPS14	0.876	265	232.14
RPS15	0.93852	175	164.241
RPS15A	0.98462	160	157.5392
RPS16	0.87337	184	160.70008
RPS17	0.78576	174	136.72224
RPS18	0.92676	192	177.93792
RPS19	0.87408	225	196.668
RPS20	0.77395	141	109.12695
RPS21	0.93667	181	169.53727
RPS23	0.94573	332	313.98236
RPS24	0.95067	171	162.56457
RPS25	0.94633	163	154.25179
RPS26	0.87155	152	132.4756
RPS27	1	106	106
RPS27A	0.86712	202	175.15824
RPS28	0.91896	80	73.5168
RPS29	0.96033	144	138.28752
RPS30	0.66634	177	117.94218
	0.874321811	7796	6816.21284

Table S7. Apomorphic sites in ribosomal proteins that are suggested to be under neutral selection.

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Apomorphy
RPL4	M1-S2	C3	C in Chordata(Aut)
	T7		T in Pezizomycotina+Saccharomycotina(Syn)
	G214		G in Ascomycota+Basidiomycota(Aut)
	A231		N in Sauropsida(Syn)
	T266		R in Chordata(Syn)
	S270-K271	L273	L in Vertebata(Syn)
	Q307		H in Amniota(Syn)
	K332		A in Pezizomycotina(Syn)
	A336		K in Pezizomycotina(Syn)
	G340		G in Pezizomycotina+Saccharomycotina(Aut)
		K353	K in Vertebrata(Syn)
RPL5	E177		T in Vertebrata(Syn)
	F185		P in Basidiomycota(Aut)
	E210		Y in Amniota(Syn)
	D232		G in Basidiomycota(Syn)
	S235		D in Basidiomycota(Syn)
	F253		Y in Mammalia(Aut)
	A287		A in Vertebrata(Syn)
RPL10	T42		K in Vertebrata(Syn)
	I210		R in Tetrapoda(Syn)
	R211		A in Vertebrata(Aut)
	F213		H in Vertebrata(Syn)
	F213		F in Ascomycota(Aut)
	P214		S in Mammalia(Syn)
	P214		P in Ascomycota(Aut)
RPL10A	K18		E in Pezizomycotina(Syn)
RPL11	T147		C in Mammalia(Aut)
RPL12	K51		K in Viridiplantae+Cryptophyta(Syn)
	K51		G in Vertebrata(Syn)
	K51		G in Pezizomycotina(Syn)
	S72		E in Vertebrata(Syn)
	K146		R in Vertebrata(Syn)
	N147		S in Pezizomycotina(Syn)
RPL14		L135	A in Basidiomycota(Aut)
RPL21		Q127	Q in Theria(Syn)
RPL23A	S4-A5	E16	E in Vertebrata(Aut)
RPL24	E119		K in Vertebrata(Syn)
RPL27A	A94		K in Mammalia(Syn)
	L133		E in Pezizomycotina(Syn)
RPL29		L75	L in Theria(Aut)
		P86	P in Theria(Aut)
RPL30		T112	T in Amniota(Aut)
RPL34	E110		Q in Vertebrata(Syn)
RPL35A	N87		S in Viridiplantae(Syn)
RPL37	K32		S in Vertebrata(Syn)
	A51		E in Pezizomycotina(Syn)
	A88		R in Vertebrata(Syn)
RPL37A	C59		R in Insecta(Aut)
RPP0	E117		L in Holometabola(Syn)
	P198		N in Amniota(Aut)
	E299		S in Vertebrata(Syn)
RPP2	T20		K in Amniota(Syn)
RPSA	T6		V in Vertebrata (Syn)
	P42		A in Basidiomycota (Syn)
	E209		A in Pezizomycotina (Aut)

Table S7 (continued)

RPG	Site in *S. cerevisiae*	Site in *Homo sapiens*	Apomorphy
	V210		E in Pezizomycotina (Syn)
	E222		G in Pezizomycotina (Aut)
RPS2	R4		D in Vertebrata (Aut)
	P234		K in Pezizomycotina (Syn)
	S238		S in *S.cerevisiae*+Pezizomycotina (Syn)
	S248		V in Theria (Syn)
	Q250		G in Pezizomycotina (Aut)
	K251		K in *S.cerevisiae*+Pezizomycotina (Aut)
	R253		Y in Pezizomycotina (Aut)
	F254		R in Vertebrata (Syn)
		V280	V in Vertebrata (Syn)
RPS3		W192	W in *C.paramecium*+Viridiplantae (Syn)
	T197		I in Chordata (Syn)
	E215		Q in Pezizomycotina (Aut)
	A219		Q in Pezizomycotina (Syn)
	V222		I in Tetrapoda (Aut)
	V238		R in Pezizomycotina (Aut)
RPS8	A139		I in Amniota (Syn)
RPS11		R22	R in Vertebrata (Syn)
RPS14	A40		C in Vertebrata (Syn)
RPS16	E42		R in Amniota (Syn)
RPS17	S112		L in Tetrapoda (Syn)
	P118		S in Vertebrata (Syn)
	R128		V in Vertebrata (Aut)
RPS18	K8		Q in Basidiomycota (Aut)
RPS19	E126		P in Chordata (Syn)
RPS20	S2		A in Vertebrata (Syn)
	Q15		P in Mammlia (Syn)
RPS30		T16	T in Mammlia (Syn)

Table S8. Intersection between the dataset of conservative sites and the cSNPs.

Conserved class	Sites in *Homo sapiens*[*]
Synapomorphy of Eukaryota	Gly140Arg(RPS3), Lys22Arg(RPS4), Gly34Ser(RPS4), Leu158Phe(RPS7), Arg109Ser(RPS9), Lys98Asn (RPS11), Arg101Ser/Gly(RPS11), Asn38Ser(RPS14), Ala74Gly(RPS14), Pro47Leu(RPS19), Arg62Trp/ Gln(RPS19), Ala135Thr(RPS19),Gly52Arg(RPS20, Pro69Ala(RPS20), Lys34Arg(RPS25), Phe99Ile(RPL11), Ser100Leu(RPL11), His100Tyr(RPL12), Ser120Thr(RPL12), Gly118Cys(RPL18)
Synapomorphy of Eukaryota*	Phe24Ser(RPS3),Leu86Pro(RPS3), Ile121Thr(RPS3A), Lys98Gln(RPS8), Val102Leu(RPS8), Glu194Lys(RPS8), Val111Phe(RPS12), Met4Val(RPS13),Thr64Ile(RPS20),Ile23Thr(RPS27A), Asp108Gly(RPS27A), Met71Ile (RPS30), Leu180Pro(RPL3), Phe249Ser(RPL4), Ile140Val(RPL7), Leu58Phe(RPL8), Leu22Pro(RPL11), Ile141 Thr(RPL11), Phe123Leu(RPL18), Leu82Val(RPL18A), Leu96Phe/Pro(RPL23), Leu78Val(RPL27A), Ile100Val (RPL27A), Leu117Phe(RPL27A), Lys80Asn/Arg(RPL37A)
Synapomorphy of the (Cnidaria+ Bilateria) clade	Ser160Phe(RPS9)
Synapomorphy of Deuterostomia	Ser137Gly(RPS7), Met181Thr(RPL17), Thr79Met(RPL18)
Synapomorphy of Chordata	Leu110Ile(RPS3), Glu256Lys(RPL8), Arg21His(RPL13), Gln205His(RPL13), Arg21Leu(RPL13A), Arg6Cys (RPL18), Ile97Thr(RPL28), Lys80Asn/Arg(RPL37A)
Synapomorphy of Vertebrata	Val280(RPS2)Phe, His138Tyr(RPS4), Ile9Thr(RPS10), Met70Val(RPS15), Leu70Val/Phe(RPS21), Arg110His (RPL6), Asn7Ser(RPL9), Leu20Pro(RPL9), Ile141Thr(RPL11), Lys152Glu(RPL19), Ala179Val/Thr(RPL19), Ile185Phe(RPL19), Ser38Leu(RPL24), Ala34Thr(RPL28), Glu30Asp(RPL29), Ser24Phe(RPL37), Leu261Phe (RPLP0)
Synapomorphy of Tetrapoda	Arg102Pro(RPS19), Asn124Ser(RPS24), Thr37Ala(RPL36), Ile51Met(RPLP2)
Synapomorphy of Amniota	Met335Leu(RPL4), Tyr210Cys(RPL5), Arg121Ser(RPL14), Arg79Pro(RPL28), Thr112Ile(RPL30), Val36Gly/ Glu/Leu/Met(RPL31), Phe185Leu(RPLP0), Asn200Lys(RPLP0)
Synapomorphy of Mammlia	Ser214Leu(RPL10)
Synapomorphy of Theria	Pro121Ala(RPL29)
Synapomorphy of Eutheria	Pro233Arg/Gln(RPS3), Ser17Pro(RPL4), Thr278Met(RPL6), Ser209Leu(RPL7A), Gln45Leu(RPL18), Ala11Thr (RPL23A), Arg128Gly(RPL28), Ile99Phe(RPL29), Arg117Trp(RPL29), Lys66Gln(RPL35A), Ser6Cys(RPLP2), Asp37Glu(RPLP2)
Autapomorphy of the (Fungi + Metazoa) group	Asp67Asn(RPS21)
Autapomorphy of Vertebrata	Pro34Thr(RPS11)
Autapomorphy of Eutheria	Ala18Val(RPL6)

* indicates the amino acids belonging to the same biochemical type serve as synapomorphy of Eukaryota.