Identification and Evolutionary Analysis of the Nucleolar Proteome of Giardia lamblia

Feng Jin-mei
Jianghan University
Yang Chun-Lin
Kunming Institute of Zoology Chinese Academy of Sciences
Tian Hai-Feng
Kunming Institute of Zoology Chinese Academy of Sciences
Wang Jiang-Xin
Kunming Institute of Zoology Chinese Academy of Sciences
Jianfan Wen (wenjf@mail.kiz.ac.cn)
Kunming Institute of Zoology Chinese Academy of Sciences https://orcid.org/0000-0002-5246-1664

Research article

Keywords: Giardia lamblia, Protist, Nucleolar proteome, Evolution, Primitiveness, Parasitic reduction

Posted Date: December 16th, 2019

DOI: https://doi.org/10.21203/rs.2.18939/v1

Version of Record: A version of this preprint was published on March 30th, 2020. See the published version at https://doi.org/10.1186/s12864-020-6679-9.

Abstract

Background: The nucleoli, including their proteomes, of higher eukaryotes have been extensively studied, while few studies about the nucleoli of the lower eukaryotes – protists were reported. *Giardia lamblia*, a protist with the controversy of whether it is an extreme primitive eukaryote or just a highly evolved parasite, might be an interesting object for carrying out the nucleolar proteome study of protists and further examining the controversy.

Results: Using bioinformatics methods, we reconstructed *G. lamblia* nucleolar proteome (*Gi*NuP) and the common nucleolar proteome of the three higher eukaryotes (human, *Arabidopsis*, yeast) (HEBNuP). Comparisons of the two proteomes revealed that: 1) *Gi*NuP is much smaller than HEBNuP, but 78.4% of its proteins have orthologs in the latter; 2) More than 68% of the proteins in *Gi*NuP are involved in the "Ribosome related" function, and the others participate in the other functions, and these two groups of proteins are much larger and much smaller than those in HEBNuP, respectively; 3) Both *Gi*NuP and HEBNuP have their own specific proteins, but HEBNuP has a much higher proportion of such proteins to participate in more categories of functions.

Conclusion: For the first time the nucleolar proteome of a protist - *Giardia* was reconstructed. The results of comparison of it with the common proteome of three representative higher eukaryotes – HEBNuP indicated that the relatively simple *Gi*NuP should reflect the primitiveness but not the parasitic reduction of *Giardia*, and simultaneously revealed some interesting evolutionary phenomena about the nucleolus and even the eukaryotic cell, compositionally and functionally.

Background

Nucleolus, the most prominent sub-nuclear compartment in the interphase nucleus of eukaryotic cells, is a ribosome factory, where most of the ribosome biogenesis events take place, such as ribosome RNA (rRNA) synthesis, processing, and subsequent assembly of ribosome subunits. Accumulated studies in the past decades have shown that this organelle is also involved in many other cellular processes, such as DNA repair, regulation of mitosis, stress response, biogenesis of multiple ribonucleoprotein particles, cancer, protein quality control [1–6]. Although the multiple functions of nucleoli have been recognized gradually, when and how they arose in the evolution of eukaryotic cells is still elusive.

The functions of the nucleolus have been studied extensively and deeply in model organisms from the three so-called higher eukaryote groups (animals, plants, and fungi) such as human, Arabidopsis, and budding yeast, and the nucleolar proteomes of the three model eukaryotes have already been identified [7–9]. Continuous high-throughput and individual case studies in these higher eukaryotic model organisms have revealed a lot of nucleolar proteins, indicating potential multiple functions of their nucleoli [10]. However, few studies of nucleoli were carried out in the so-called lower eukaryotes, protists, much less the study of their nucleolar proteomes. However, protists occupy pivotal positions in the evolution of eukaryotes because they are the link between prokaryotes and multicellular/higher

eukaryotes. Therefore, studies on their nucleoli will be valuable for understanding the origin and evolution of the nucleolus and even the eukaryotic cells.

G. lamblia is an intestinal protozoan parasite responsible for widespread diarrheal disease in humans and animals worldwide [11]. Besides medical importance, its significance in the study of eukaryotic evolution was first proposed in 1980s but has been debated for many years. It was once thought to be the most primitive extant eukaryote because of having many so-called primitive peculiarities: lack of some eukaryotic typical cellular structures such as mitochondrion [12] and nucleolus [13, 14], and its early branching position on some phylogenetic trees [15–18]. However, the later discoveries of mitochondrion-derived organelle – mitosome [19] and nucleolus [20] in its cells, and the non-early branching positions on some other phylogenetic trees [21, 22] tend to refute the primitivity of Giardia but prove that it is just a highly evolved parasite with many parasitic reductions [23, 24]. On the other hand, some authors found that some simple traits of Giardia cannot be attributed to its parasitic reduction, and thus still persisted in that Giardia is one of the most primitive extant eukaryotes, and emphasized that it is of significance to the study of the evolution of the eukaryotic cell [25–28]. Therefore, investigating the nucleolar proteome of G. lamblia may be useful either to the re-examining of the debate above or to the understanding of the evolution of the nucleolus and the eukaryotic cell.

However, high quality isolation of nucleoli from nuclei is always a challenge even for higher eukaryotic cells, and the isolation of the nucleoli from G. lamblia cells is much more difficult using the available experimental techniques because of the smallest size of Giardia nucleolus and probably other reasons such as fragility. Accordingly, it is almost impossible to use mass spectrometry, the best efficient method for proteome studies, to identify nucleolar proteins of G. lamblia. Fortunately, the nucleolar proteomes and genome databases of three higher eukaryotic representatives of animals, plants, and fungi mentioned above are available, and the completely sequenced genome of G. lamblia has also been determined and reported. Therefore, here we used a series of bioinformatics tools to identify nucleolar protein genes of G. lamblia and reconstruct the nucleolar proteome (GiNuP) and also to reconstruct the ‘Higher Eukaryote Basic Nucleolar Proteome (HEBNuP)’, then a comprehensively comparative proteomics analysis between the GiNuP and the HEBNuP were performed, and some significant implications for the evolution of nucleolar protein components and functions and for the evolutionary position of Giardia were obtained.

Results

Reconstruction of the giardial nucleolar proteome (Gi NuP)

To obtain a relatively complete nucleolar proteome of G. lamblia, we have used two independent methods to bioinformatically identify putative nucleolar proteins in the genome of this protist: homology search based on the known nucleolar proteins of the three higher eukaryote representatives and de novo prediction by analyzing protein sequence features. For homology search, 38 candidate Giardia orthologs were obtained when blasting with 246 yeast nucleolar proteins as queries. Analogously, 57 and 189

candidate orthologs were obtained when blasting with 281 A. thaliana and 4705 human nucleolar proteins as queries, respectively. All the Giardia nucleolar proteins orthologous to those of H. sapiens, A. thaliana, and S. cerevisiae were collected together. After discarding the redundant ones, 237 Giardia nucleolar protein candidates were obtained finally. Subsequent domain analyses of these obtained protein sequences by using PFAM online service showed that 216 ones possess characteristic domains of various nucleolar proteins. They were further confirmed to be nucleolar proteins by Blast searching against the nr protein database in NCBI. Finally, 216 orthologs to the nucleolar proteins of the three representative eukaryotes were identified in the G. lamblia genome database by the homology search approach (Supplementary Table S1).

Since all the available nucleolar proteomes of the three higher eukaryotes each possess their own specific proteins that do not have any homologs in the other two proteomes, it is reasonable to image that G. lamblia, though much more ancient, also has its specific nucleolar proteins, which are not present in other species. Therefore, to identify such putative Giardia specific nucleolar proteins, we investigated all the Giardia proteins in the genome database to identify those ones that would be predicted to localize to the nucleolus from all the nuclear proteins. First, we got 172 Giardia nuclear proteins by predicting to have nuclear location signal. We also used 'nucleus/nuclear' or "nucleolus/nucleolar" as key words to screen the G. lamblia genome database, and obtained 25 annotated nuclear/nucleolar proteins. Then all the 197 (172 + 25) nuclear proteins were further subjected to the protein sub-localization prediction, and 55 of them were predicted to be most likely localized to the nucleolus.

Altogether, finally 255 (216 + 39) nucleolar proteins were identified in the G. lamblia genome database after discarding the redundant ones, which includes 216 orthologs to the nucleolar proteins of the three representative eukaryotes and 39 Giardia-specific nucleolar proteins (Table S1). Thus, we have reconstructed a putative nucleolar proteome of G. lamblia (GiNuP), which contains 255 individual nucleolar proteins.

Reconstruction of the 'Higher Eukaryote Basic Nucleolar Proteome (HEBNuP)'

To compare the GiNuP with the nucleolar proteomes of the three representatives of higher eukaryotes, we investigated the orthologous relationships between either two or among all the three higher eukaryotes by identifying the nucleolar proteins that are present in all the three genomes. Because of the relatively far less protein numbers in both the nucleolar proteomes of Arabidopsis and budding yeast, to avoid the possible incompleteness of them, we collected all the ortholog groups with the presence of human nucleolar proteins. This investigation revealed the following orthologous relationships: 1) there are 1058 orthologous groups between human nucleolar proteome and Arabidopsis whole proteome, containing 2341 human nucleolar proteins and 2780 Arabidopsis proteins, respectively; 2) there are 856 orthologous groups between human nucleolar proteome and budding yeast whole proteome, containing 1946 human nucleolar proteins and 1078 yeast proteins, respectively; 3) there are 799 orthologous groups among human nucleolar proteome, the whole proteome of Arabidopsis, and budding yeast proteome, containing 1848 human nucleolar proteins, 2227 Arabidopsis proteins, and 1015yeast proteins, respectively (Fig. 1

and Supplementary Table S2). As a whole, we called these 799 orthologous groups as 'Higher Eukaryote Basic Nucleolar Proteome (HEBNuP)'.

The functional inventories of the proteins in the HEBNuP and GiNuP

The results of functional inventory of the 1848 human nucleolar proteins in HEBNuP is as follows (Fig. 2A): 1) 218 (12%) belong to the "Ribosome related" class; 2) 220 (12%) belong to the "mRNA related" class; 3) 222 (12%) belong to the "Translation related" class; 4) 176 (9.5%) belong to the "DNA binding" proteins; 5) 69 (4%) belong to the "Chromatin related" class; 6) 86 (5%) belong to the "Mitotic cell cycle related" class; 7) 857 (46.5%) belong to none of the six classes, and thus we classify them as "undefined function" class.

The results of functional inventory of the 255 proteins in GiNuP is as follows (Fig. 2B): 1) 73 (29%) proteins are classified among the "Ribosome related" proteins; 2) three (1%) belong to the "mRNA related" class; 3) 12 (5%) belong to the "Translation related" class; 4) 12 (5%) belong to the "DNA binding related" class; 5) six (2%) belong to the "Chromatin related" class; 6) one (0.4%) belong to the "Mitotic cell cycle related" class; 7) 148 (57.6%) belong to the "undefined function" class.

Comparative analysis between the GiNuP and HEBNuP

To explore the evolution of nucleolus, we compared the GiNuP and HEBNuP in terms of protein homology and function. From the above results, we know that the HEBNuP consists of 799 orthologous groups, which contains 1848 individual human nucleolar proteins – the HEBNuP-Hu protein dataset, and that the GiNuP dataset contains 255 orthologous groups and Giardia nucleolar proteins. Since the nucleolar proteome of human seems to be the most complete one among those of the three higher eukaryotes, thus the nucleolar protein groups in HEBNuP-Hu protein dataset were used as representatives of HEBNuP to compare with those in GiNuP in the following analysis.

Comparison of the GiNuP with the HEBNuP in terms of protein homology shows that: 1) 200 orthologous groups (containing 200 individual Giardia nucleolar proteins) are shared by GiNuP and HEBNuP, which make up the HEBNuP-GiNuP-shared dataset, indicating that 78.4% (200 out of 255) of the Giardia nucleolar protein orthologous groups (also the individual proteins) all have their orthologs in the HEBNuP, but these orthologs only occupy 25.0% of the orthologous protein groups of the HEBNuP (and the Giardia nucleolar proteins only occupy 13.8% of the individual human nucleolar proteins in the HEBNuP and HEBNuP-Hu), which means that the majority of Giardia nucleolar proteins belong to the common/basic nucleolar proteins of the higher eukaryotes, and in higher eukaryotes the common/basic nucleolar proteins are much more than in Giardia; 2) 55 Giardia nucleolar orthologous groups (containing 55 individual Giardia nucleolar proteins) are specific to GiNuP, which make up the dataset we call GiNuP-specific datase; 599 orthologous groups (containing 1253 individual human nucleolar proteins) in HEBNuP are specific to HEBNuP, which make up the dataset we call HEBNuP-specific dataset.

The functional distributions of the nucleolar orthologous protein groups in the five datasets mentioned above are shown in Fig. 3, and the proportions of the annotated proteins for each nucleolar functional class are shown in Fig. 4. Functional distribution comparison of the proteins in the GiNuP with those in the HEBNuP shows that: 1) 68.2% of the annotated proteins in the GiNuP dataset and 68.9% in the HEBNuP-GiNuP-shared dataset are involved in the "Ribosome related" function, respectively, implying that the majority of the annotated Giardia's nucleolar proteins participate in the "Ribosome related" function, and that these proteins still perform this function in higher eukaryotes; the other about 31% of the annotated proteins in these two datasets are involved in the other five functions, respectively, implying that besides the major "Ribosome related" function, the other five nucleolar functions also exist in Giardia's nucleolus, though with a very few proteins to perform them, and that these few proteins still perform the five functions in higher eukaryotes. 2) Half (50%) of the annotated proteins in GiNuP-specific dataset are classified into the "Ribosome related" functional class, 25% are classified into the "DNA binding related" functional class, and the other 25% are classified into the "Translation related" functional class, and none are classified into the other three functional classes; 22.7%, 25%, 27.7%, 10.6%, 2.7%, and 11.2% of the annotated proteins in HEBNuP-specific dataset are classified into the "Ribosome related", "DNA binding related", "Translation related", "Chromatin related", "mRNA related", and "Mitotic cell cycle related" functional classes, respectively, which means that the basic "Ribosome related" function of nucleolus also needs lineage- and even species-specific protein components to perform it in a certain lineage or species, and so do the other five nucleolar functions; and that such specific proteins, especially those for the other five functions, continuously increased in the evolution of eukaryotes. Besides, obviously, for both the GiNuP and the GiNuP-specific datasets, the proportions of annotated proteins involved in the other five functional classes all are much fewer than those involved in the "Ribosome related" function, while for the HEBNuP-Hu dataset and the HEBNuP-specific dataset, the proportions of nucleolar proteins involved in the other five functions increase much more substantially, compared with those involved in the "Ribosome related" function. This implies that the "Ribosome related" function should arise and consummate earlier than the other five functions, and the other five ones became more and more consummate and complicated latter, especially in the evolution of higher eukaryotes.

Discussion

The nucleolus of G. lamblia seems to be the smallest one described so far [29] and atypical when compared with those of higher eukaryotes [20], and they are very difficult to isolate in high quality for mass spectrometry, thus, here we tried to use bioinformatics methods to identify its proteome based on its genome database and the already-existing nucleolar proteome databases of three representative eukaryotes, human, Arabidopsis, and yeast. In order to exhaustively identify the putative nucleolar proteins in Giardia, Giardia-specific nucleolar proteins were also identified by a combined computational approach. Thus we reconstructed the first nucleolar proteome of unicellular protist – Giardia's nucleolar proteome, GiNuP.

When comparing with any one of the nucleolar proteomes of human, A. thaliana, and yeast [7–9], the GiNuP was found to contain far fewer nucleolar proteins. Thus, in terms of protein number, the nucleolar

protein components of G. lamblia are much simpler than those of higher eukaryotes. However, since many species-specific nucleolar proteins have been found in the nucleolar proteomes of human, A. thaliana, and yeast [7–9], and also in Giardia (please see those we identified above), to reasonably compare the component and function of nucleolar proteins between GiNuP and the nucleolar proteomes of typical eukaryotes, here we reconstructed the HEBNuP, which consists of the nucleolar protein orthologous groups shared by the proteomes of the three representative eukaryotes and thus to a certain degree can represent the common/basic protein components of the nucleolus of higher eukaryotes, and then compared it with the GiNuP in two aspects -- orthologous group and functional category. Compared with that of human, which was obtained by using multiple mass spectrometry to analyze highly purified preparations of human nucleoli from different cell lines, the nucleolar proteomes of Arabidopsis and yeast are remarkably smaller and thus might have been underestimated due to the less sensitive mass spectrometric techniques used and the dynamic behavior of nucleolar proteins [8, 9, 30]. Thus in the present work, for Arabidopsis and yeast, we used their putative whole proteome (downloaded from the genome database) instead of their nucleolar proteomes in the reconstruction of HEBNuP. Comparisons of protein components between the GiNuP and the HEBNuP revealed that the majority of Giardia nucleolar proteins belong to this common/basic nucleolar proteins of higher eukaryotes, but the individual protein number (and also the orthogous group number) of these Giardia nucleolar proteins is far fewer than those in the higher eukaryotes, which suggests that Giardia's simplified nucleolus should be a reflection of its primitiveness rather than its parasitic reduction. Because (1) in general, the common/basic nucleolar proteins should emerge earlier than other proteins in the evolution of the nucleolus (and also of the eukaryote), thus our findings that GiNuP is mainly composed of the common/basic nucleolar proteins (namely, the proportion of the other proteins in GiNuP is much lower than that in HEBNuP), and that the main and basic function of nucleolus -- "Ribosome related" function is the major function of the GiNuP, both imply that Giardia's nucleolus is a very primitive one; (2) the parasitic reduction should not be necessary to occur on the common/basic nucleolar proteins which take part in the basic nucleolar function in all eukaryotes but are not directly related to parasitic life-style, and the much smaller number of the common/basic nucleolar proteins in Giardia must be due to the primitive status of nucleolus of this organism, and later more and more proteins were recruited into the nucleolus as common/basic nucleolar proteins during eukaryotic evolution after the divergence of Giardia from the eukaryote trunk (our data shows that the common/basic nucleolar proteins have increased about 300% from GiNuP to the HEBNuP), on the contrary, it is much less likely that Giardia lost so much of the common/basic nucleolar proteins of the eukaryotic essential structure due to parasitism. Actually, our previous studies have also revealed that Giardia's unusual and simple 5S rRNA system should be a reflection of its primitiveness but not be due to parasitic degeneration [27], and that Giardia possesses 89 orthologs to the 129 conserved common ribosomal biogenesis proteins of higher eukaryotes, which can carry out all the steps of ribosome biogenesis, indicating that the ribosome biogenesis system of Giardia is similar to that of higher eukaryotes but just simpler [31]. Moreover, it was reported that compared with its counterparts of higher eukaryotes, the nucleolar organizer regions (NORs) of Giardia gather much less copies of much shorter rDNA repeat units and participate in the formation of the structurally simpler nucleolus of this organism [32]. Therefore, the nucleolus of G. lamblia is simpler than those of higher

eukaryotes in structure, composition, and function, and such a simplified nucleolus in G. lamblia should be due to its primitiveness but not secondary parasitic reduction. Our recent work on Giardia's GPL biosynthesis pathways revealed that these pathways of it are evolutionarily primitive, but with many secondary parasitic adaptation 'patches' including gene loss, rapid evolution, and horizontal gene transfer, which implies Giardia might be a mosaic of 'primary primitivity' and 'secondary parasitic adaptability' [28]. This is also consistent with the present work.

Based on the above understanding that Giardia's nucleolus is a primitive one, our results of comparison of GiNuP with HEBNuP thus can reveal some interesting evolutionary phenomena. First, the majority of Giardia nucleolar proteins have orthologs to the common/basic nucleolar proteins of higher eukaryotes (HEBNuP) but occupy a very small proportion of the latter, and that the majority of the Giardia's nucleolar proteins participate in the "Ribosome related" function, both the observations should imply that the "Ribosome related" function, as the major/basic function of the nucleolus, must have arisen earlier than the other nucleolar functions, and that this major/basic function became more and more consummate and complicated in the evolution of eukaryotes by increasing more and more functional protein components. Second, there are some proteins in GiNuP (though very few compared to those of higher eukaryotes) involved in the other five nucleolar functions should mean that besides the major "Ribosome related" function, the other five nucleolar functions also have arisen in Giardia, though with a very few proteins to perform them, and they also became more and more consummate and complicated in the evolution of eukaryotes, especially in the evolutionary process from primitive unicellular protists to higher multicellular eukaryotes. Third, in either Giardia or the higher eukaryotes, either the major "Ribosome related" function or the other five functions, all contain some (quite a proportion in higher eukaryotes) species- and linage-specific proteins, and that such specific proteins, especially those for the other five functions, increased remarkably in higher eukaryotes, should mean that in all eukaryotic species and lineages, specific protein components are also necessary to evolve to participate the performance of all the common functions of nucleolus. This might be a very interesting evolutionary biology finding, which probably implies that the evolution from lower to higher organisms, especially in the divergence of species and lineages, does not simply mean the increase of common components on the basis of the relatively lower organisms but the evolution of species- and lineage-specific components for a cellular structure or a function so as to became more efficient and consummate in a certain species and lineage.

Conclusions

To sum up, in the present work for the first time the nucleolar proteome of a lower eukaryote (protist) – Giardia (GiNuP) was reconstructed. The results of comparison of it with the common proteome of three representative higher eukaryotes – HEBNuP indicated that the relatively simple GiNuP should reflect the primitiveness but not the parasitic reduction of Giardia, and revealed some interesting evolutionary phenomena about the nucleolus and even the eukaryotic cell, compositionally and functionally.

Methods

Data collection

The International Protein Index (IPI) IDs of 4,749 available Homo sapiens nucleolar proteins and their corresponding sequences were retrieved from the Nucleolar Proteome Database NOPdb3.0 [7], and the non-redundant 4057 IDs and sequences were used in this study. The whole human genome data was downloaded from Ensembl. The IDs and sequences of 281 available A. thaliana nucleolar proteins were downloaded from the Arabidopsis Information Resource [33, 34] and the Arabidopsis nucleolar protein database (AtNoPDB) [35]. The IDs and sequences of 246 available S. cerevisiae nucleolar proteins were downloaded from the Saccharomyces Genome Database [36–38] and the Comprehensive Yeast Genome Database [39]. G. lamblia genome data was downloaded from the GiardiaDB (http://giardiadb.org/giardiadb/) [11]. The Gene Ontology (GO) functional annotations of human proteins were downloaded from the Gene Ontology (http://www.geneontology.org/).

Identification of Giardia nucleolar proteins by homology search

We used the Best Reciprocal Hit (BRH) method to identify nucleolar protein orthologs in G. lamblia genome. Briefly, the nucleolar protein sequences from human, Arabidopsis, and budding yeast were used as queries to BLASTP search against G. lamblia genome (E-value $\leq$ 0.001, coverage $\geq$ 25%, and identity $\geq$ 25%). The obtained hit protein sequences were collected and used as queries to BLASTP search against genomes of human, Arabidopsis and budding yeast following the same standards, respectively. Reciprocal best hits between G. lamblia and either of human, Arabidopsis and budding yeast were established, and those Giardia proteins that have reciprocal hit in either of these three reference genomes were considered as candidate nucleolar proteins in G. lamblia. Then, the obtained candidate protein sequences were assessed by domain analysis by using PFAM online service [40], and those ones that contain known nucleolar protein domains were considered as putative nucleolar proteins. Further validation of these putative nucleolar proteins was performed by using them as queries to BLASTP search against GenBank non-redundant (nr) protein database to investigate the annotations of their identified homologs in nr database.

Identification of Giardia-specific nucleolar proteins by subcellular localization prediction and reconstruction of G. lamblia Nucleolar Proteome (GiNuP)

The nucleolar proteins specific to G. lamblia were identified by a combined computational approach. First, two approaches were used to screen for nuclear proteins in the G. lamblia genome data: 1) Using “nucleus/nuclear” or “nucleolus/nucleolar” as key words to search against the genome database to collect all the related annotated proteins; 2) Using PredictNLS program (https://rostlab.org/owiki/index.php/PredictNLS) and Psort II program (http://psort.hgc.jp) [41] to predict the nuclear location signal (NLS) in all the proteins in the G. lamblia genome data and collecting the proteins with NLS. Putting the results of 1) and 2) together, we obtained all the nuclear proteins in the G. lamblia genome data. Then, the ProLoc prediction program [42] was used to predict the subnuclear localizations of them, and those ones that were predicted to be localized to the nucleolus were considered

as nucleolar protein candidates. Finally, after removing those ones overlapping with those identified by BRH above, Giardia-specific nucleolar proteins were obtained. Combining the orthologs identified by BRH above with these specific ones, we obtained the nucleolar proteins and genes in G. lamblia genome data, and put them together and reconstructed G. lamblia Nucleolar Proteome (GiNuP). The general approach for identifying G. lamblia nucleolar proteins and reconstructing the GiNuP is summarized in Fig. 5.

Reconstruction of the 'Higher Eukaryote Basic Nucleolar Proteome (HEBNuP)'

The orthologous relationships between any two of the three eukaryotes, H. sapiens, A. thaliana, and S. cerevisiae, were obtained from InParanoid database (http://inparanoid.sbc.su.se/cgi-bin/index.cgi) [43]. Orthologous nucleolar protein groups among all the three species were generated by MultiParanoid [44] based on the pairwise orthologous relationships. The IPI IDs of human nucleolar proteins were used to replace their corresponding Ensembl IDs in the orthologous groups through a local BLASTP search against the whole human proteome database in Ensembl with the 4057 human nucleolar proteins as queries (E-value cutoff 1e-10). The orthologous groups shared by human nucleolar proteome and the whole proteomes (in genome databases) of Arabidopsis and yeast were put together to reconstruct the 'Higher Eukaryote Basic Nucleolar Proteome (HEBNuP)'.

Functional inventory of the proteins in the GiNuP and HEBNuP

The GO functional annotation of each human nucleolar protein was from the Gene Ontology database. Because no GO functional annotation of G. lamblia proteins is available to date, the GO functional annotations of G. lamblia nucleolar protein orthologs were classified according to the GO functional annotations of corresponding human nucleolar proteins in the same ortholog group. Ortholog groups among the G. lamblia, H. sapiens, A. thaliana, and S. cerevisiae, were generated by MultiParanoid as described above. Based on the identified nucleolar functions previously [1–5], we classified the nucleolar proteins into the following six main functional categories: 1) "ribosome related", for example, 'rRNA processing'; 2) "mRNA related", for example, 'mRNA processing'; 3) "translation related", for example, 'translation initiation factor'; 4) "DNA binding related", for example, 'DNA binding'; 5) "chromatin related", for example, 'chromatin remodeling complex'; and 6) "mitotic cell cycle related", for example, 'M/G1 transition of mitotic cell cycle'. Then the nucleolar proteins in GiNuP and HEBNuP were inventoried by the six categories.

Comparative analysis between GiNuP and HEBNuP

Perl scripts were written to compare the GiNuP with HEBNuP compositionally and functionally. Besides the GiNuP dataset, four other datasets of nucleolar proteins were constructed: 1) HEBNuP-GiNuP-shared dataset: the common proteins shared by both the GiNuP and HEBNuP; 2) GiNuP-specific dataset: the proteins being exclusively present in GiNuP; 3) HEBNuP-specific dataset: the proteins being exclusively present in HEBNuP; 4) HEBNuP-Hu dataset: all the human nucleolar proteins in HEBNuP. Functional inventories of the proteins in all the five datasets were also carried out as above. Finally, comparisons of the six main and well-known nucleolar functional classes among the five datasets were implemented.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work is supported by the National Natural Science Foundation of China (NSFC) (Grant No. 31572256, 31772452) and the open foundation of the State Key Laboratory of Genetic Resources and Evolution (Grant No. GREKF16-02).

Authors' contributions

JFW, JMF and CLY conceived and designed the experiment. JMF, CLY, HFT and JXW analyzed the data. JFW, JMF, CLY and HFT wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Ms Yasmeen Ahmad (University of Dundee) for her help about the use of NOPdb3.0.

References

1.
Boisvert FM, van Koningsbruggen S, Navascues J, Lamond AI. The multifunctional nucleolus. Nat Rev Mol Cell Biol. 2007;8:574–85.
2.
Feng JM, Sun J, Wen JF. Advances in the study of the nucleolus. Zoological Research. 2012;33:8.
3.
Larsen DH, Stucki M. Nucleolar responses to DNA double-strand breaks. Nucleic Acids Res. 2016;44:538–44.
4.
Shaw P, Brown J. Nucleoli: composition, function, and dynamics. Plant Physiol. 2011;158:44–51.

5.
Takada H, Kurisaki A. Emerging roles of nucleolar and ribosomal proteins in cancer, development, and aging. Cell Mol Life Sci. 2015;72:4015–25.
6.
Frottin F, Schueder F, Tiwary S, Gupta R, Korner R, Schlichthaerle T, Cox J, Jungmann R, Hartl FU, Hipp MS. The nucleolus functions as a phase-separated protein quality control compartment. Science. 2019;365:342–7.
7.
Ahmad Y, Boisvert FM, Gregor P, Cobley A, Lamond AI. NOPdb: Nucleolar Proteome Database–2008 update. Nucleic Acids Res. 2009;37:D181–4.
8.
Pendle AF, Clark GP, Boon R, Lewandowska D, Lam YW, Andersen J, Mann M, Lamond AI, Brown JW, Shaw PJ. Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. Mol Biol Cell. 2005;16:260–9.
9.
Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. Global analysis of protein localization in budding yeast. Nature. 2003;425:686–91.
10.
Ogawa LM, Baserga SJ. Crosstalk between the nucleolus and the DNA damage response. Mol Biosyst. 2017;13:443–55.
11.
Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, et al. Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science. 2007;317:1921–6.
12.
Gillin FD, Reiner DS, McCaffery JM. Cell biology of the primitive eukaryote Giardia lamblia. Annu Rev Microbiol. 1996;50:679–705.
13.
Narcisi EM, Glover CV, Fechheimer M. Fibrillarin, a conserved pre-ribosomal RNA processing protein of Giardia. J Eukaryot Microbiol. 1998;45:105–11.
14.
Guo J, Chen YH, Zhou KY, Li JY. Distribution of rDNA in the nucleus of Giardia lamblia - Detection by Ag-I silver stain. Anal Quant Cytol Histol. 2005;27:79–82.
15.
Sogin ML, Gunderson JH, Elwood HJ, Alonso RA, Peattie DA. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from Giardia lamblia. Science. 1989;243:75–7.
16.
Cavalier-Smith T, Chao EE. Molecular phylogeny of the free-living archezoan Trepomonas agilis and the nature of the first eukaryote. J Mol Evol. 1996;43:551–62.
17.

Hashimoto T, Nakamura Y, Kamaishi T, Nakamura F, Adachi J, Okamoto K, Hasegawa M. Phylogenetic place of mitochondrion-lacking protozoan, Giardia lamblia, inferred from amino acid sequences of elongation factor 2. Mol Biol Evol. 1995;12:782–93.
18.
Hashimoto T, Nakamura Y, Nakamura F, Shirakura T, Adachi J, Goto N, Okamoto K, Hasegawa M. Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, Giardia lamblia. Mol Biol Evol. 1994;11:65–71.
19.
Tovar J, Leon-Avila G, Sanchez LB, Sutak R, Tachezy J, van der Giezen M, Hernandez M, Muller M, Lucocq JM. Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. Nature. 2003;426:172–6.
20.
Jimenez-Garcia LF, Zavala G, Chavez-Munguia B, Ramos-Godinez Mdel P, Lopez-Velazquez G, Segura-Valdez Mde L, Montanez C, Hehl AB, Arguello-Garcia R, Ortega-Pierres G. Identification of nucleoli in the early branching protist Giardia duodenalis. Int J Parasitol. 2008;38:1297–304.
21.
Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci U S A. 2009;106:3859–64.
22.
Burki F. The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb Perspect Biol. 2014;6:a016147.
23.
Lloyd D, Harris JC. Giardia: highly evolved parasite or early branching eukaryote? Trends Microbiol. 2002;10:122–7.
24.
Cernikova L, Faso C, Hehl AB. Five facts about Giardia lamblia. PLoS Pathog. 2018;14:e1007250.
25.
Nino CA, Chaparro J, Soffientini P, Polo S, Wasserman M. Ubiquitination dynamics in the early-branching eukaryote Giardia intestinalis. Microbiology open. 2013;2:525–39.
26.
Gourguechon S, Holt LJ, Cande WZ. The Giardia cell cycle progresses independently of the anaphase-promoting complex. J Cell Sci. 2013;126:2246–55.
27.
Feng JM, Sun J, Xin DD, Wen JF. Comparative Analysis of the 5S rRNA and Its Associated Proteins Reveals Unique Primitive Rather Than Parasitic Features in Giardia lamblia. PLoS One. 2012;7:e36878.
28.
Ye Q, Tian H, Chen B, Shao J, Qin Y, Wen J. Giardia's primitive GPL biosynthesis pathways with parasitic adaptation 'patches': implications for Giardia's evolutionary history and for finding targets against Giardiasis. Sci Rep. 2017;7:9507.

29.
Lara-Martinez R, De Lourdes Segura-Valdez M, De LMora-De La Mora, Lopez-Velazquez I, Jimenez-Garcia G. LF: Morphological Studies of Nucleologenesis in Giardia lamblia. Anat Rec (Hoboken). 2016;299:549–56.
30.
Andersen JS, Lam YW, Leung AKL, Ong SE, Lyon CE, Lamond AI, Mann M. Nucleolar proteome dynamics. Nature 2005, 433.
31.
Xin DD, Wen JF. Ribosome Biogenesis System of Giardia Inferred from Analysis of Giardial Genome. Zoological Research. 2005;26:484–91.
32.
Adam RD. Biology of Giardia lamblia. Clin Microbiol Rev. 2001;14:447–75.
33.
Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W, et al. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. 2001;29:102–5.
34.
Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, et al. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res. 2003;31:224–8.
35.
Brown JW, Shaw PJ, Shaw P, Marshall DF. Arabidopsis nucleolar protein database (AtNoPDB). Nucleic Acids Res. 2005;33:D633–6.
36.
Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, et al. SGD: Saccharomyces Genome Database. Nucleic Acids Res. 1998;26:73–9.
37.
Christie KR, Weng S, Balakrishnan R, Costanzo MC, Dolinski K, Dwight SS, Engel SR, Feierbach B, Fisk DG, Hirschman JE, et al. Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. Nucleic Acids Res. 2004;32:D311–4.
38.
Hirschman JE, Balakrishnan R, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hong EL, Livstone MS, Nash R, et al. Genome Snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the Saccharomyces cerevisiae genome. Nucleic Acids Res. 2006;34:D442–5.
39.

Guldener U, Munsterkotter M, Kastenmuller G, Strack N, van Helden J, Lemer C, Richelles J, Wodak SJ, Garcia-Martinez J, Perez-Ortin JE, et al. CYGD: the Comprehensive Yeast Genome Database. Nucleic Acids Res. 2005;33:D364–8.
40.
Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. The Pfam protein families database. Nucleic Acids Res. 2012;40:D290–301.
41.
Horton P, Nakai K. Better prediction of protein cellular localization sites with the k nearest neighbors classifier. Proc Int Conf Intell Syst Mol Biol. 1997;5:147–52.
42.
Huang WL, Tung CW, Huang HL, Hwang SF, Ho SY. ProLoc: Prediction of protein subnuclear localization using SVM with automatic selection from physicochemical composition features. Biosystems. 2007;90:573–81.
43.
Remm M, Storm CEV, Sonnhammer ELL. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. Journal Of Molecular Biology. 2001;314:1041–52.
44.
Alexeyenko A, Tamas I, Liu G, Sonnhammer ELL. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. Bioinformatics. 2006;22:E9–15.

Figures

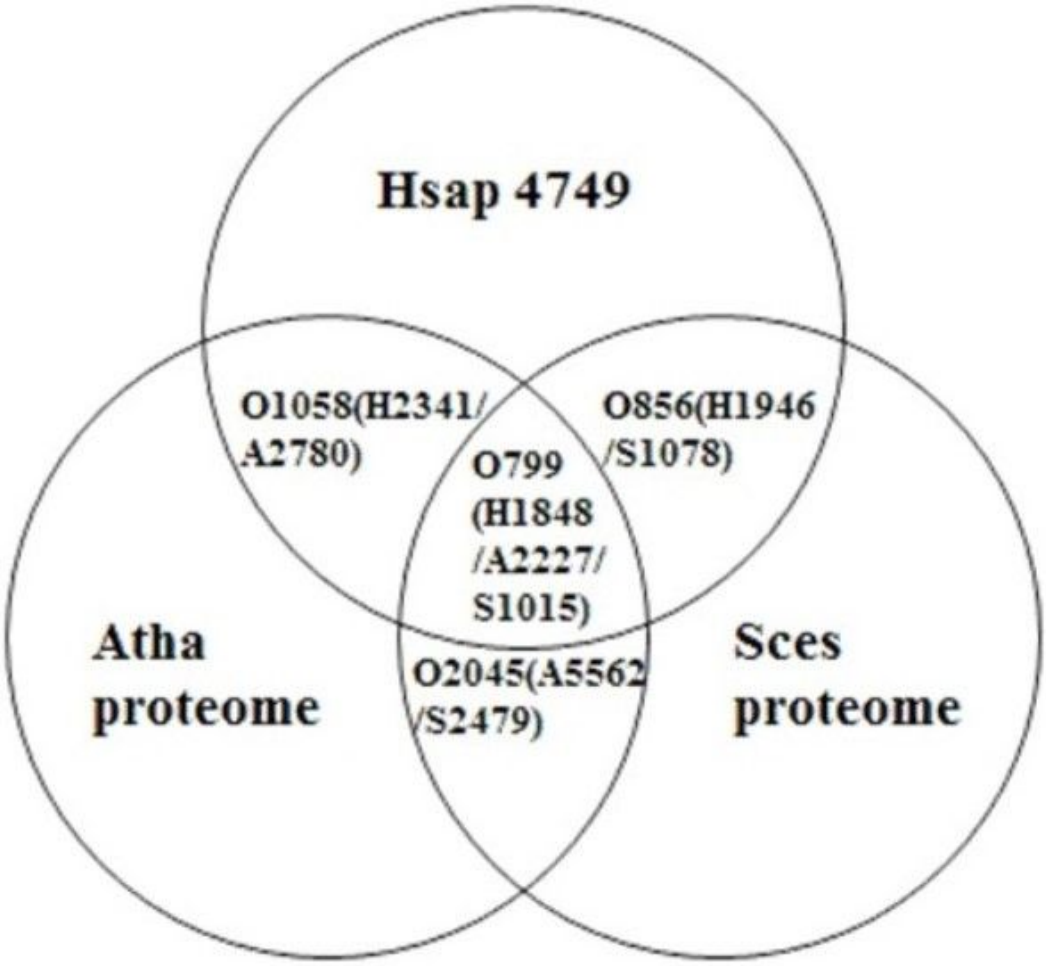

Figure 1

Orthologous relationships of nucleolar proteomes among Human (Hsap, H) and Arabidopsis (Atha, A), Yeast (Sces, S).

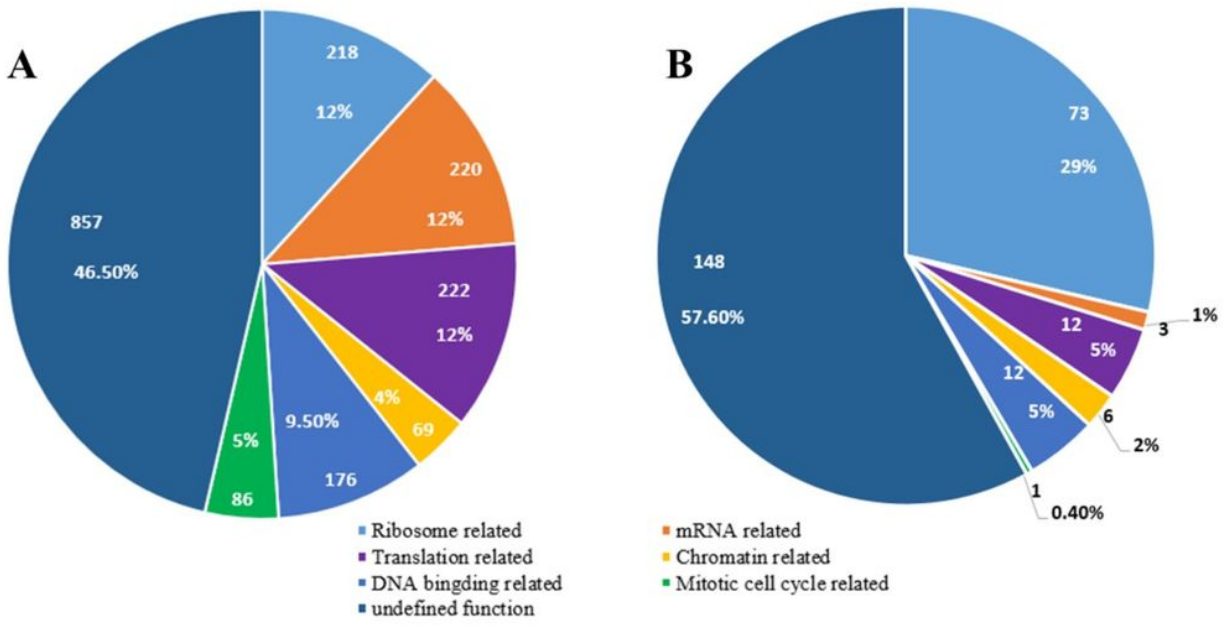

Figure 2

The functional inventories of nucleolar proteins in HEBNuP (A) and GiNuP (B).

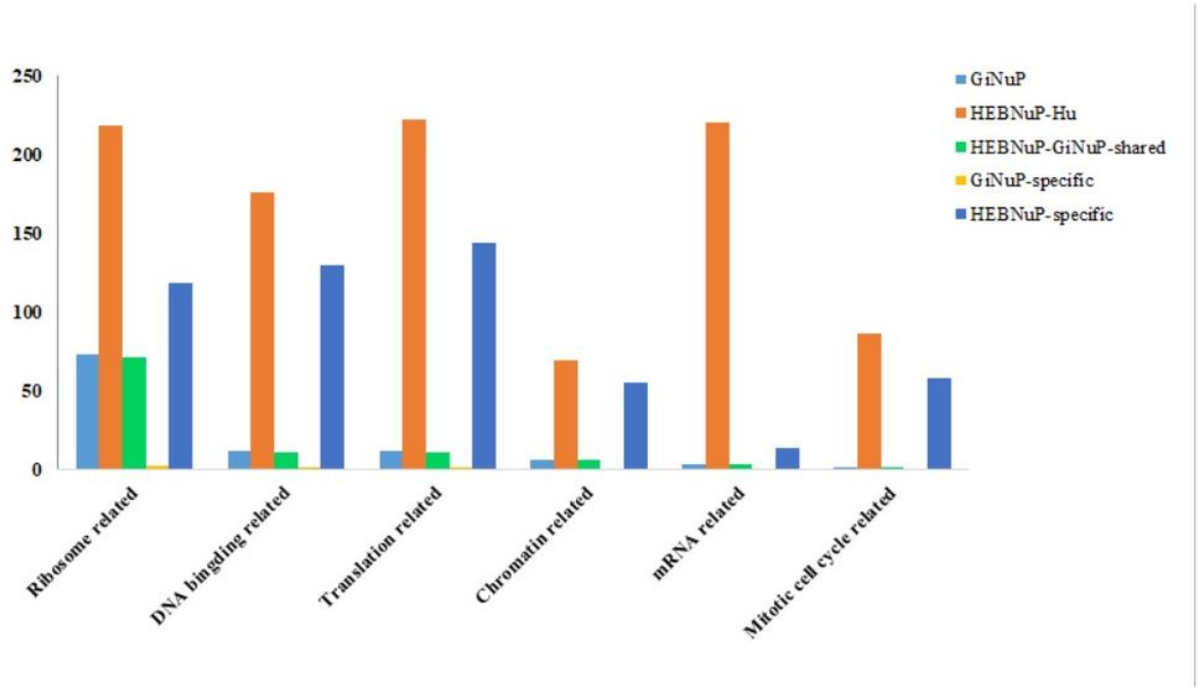

Figure 3

Functional distribution of nucleolar proteins in the five datasets. The five different colors refer to the five datasets, respectively; Horizontal axis, six main and well-known nucleolar functional classes; Vertical axis, Number of proteins.

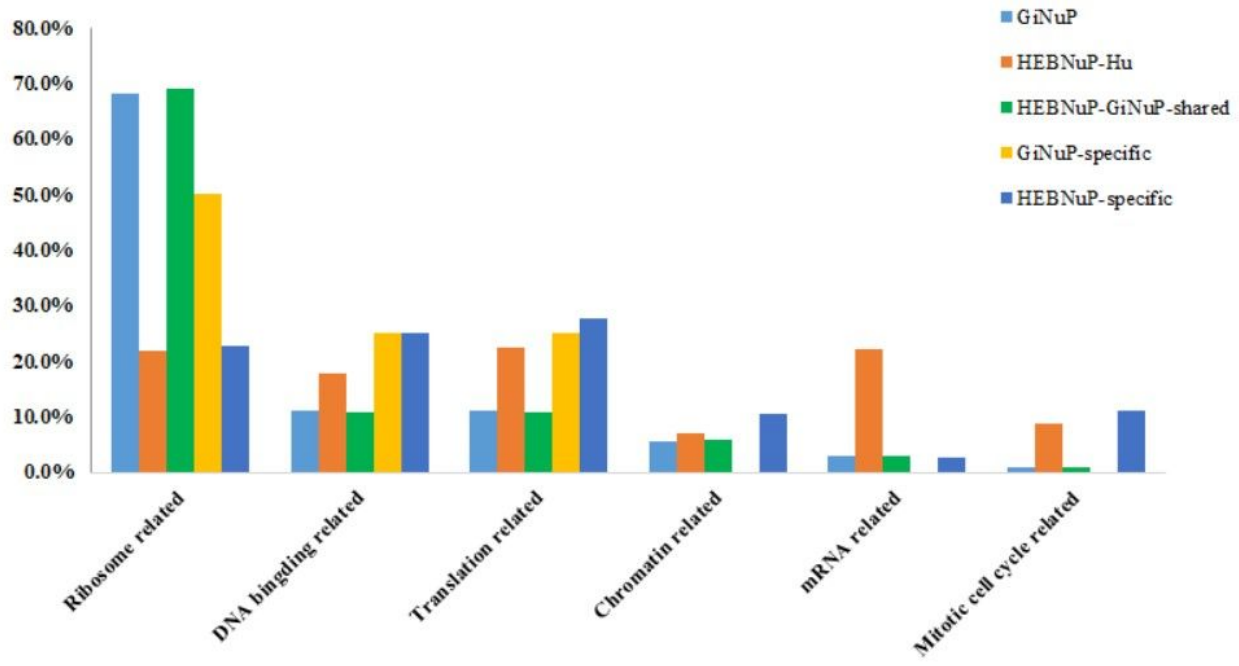

Figure 4

Comparisons of the proportions of the proteins in each nucleolar functional class of the five datasets. The five different colors refer to the five datasets, respectively; Horizontal axis, six main and well-known nucleolar functional classes; Vertical axis, Ratio.

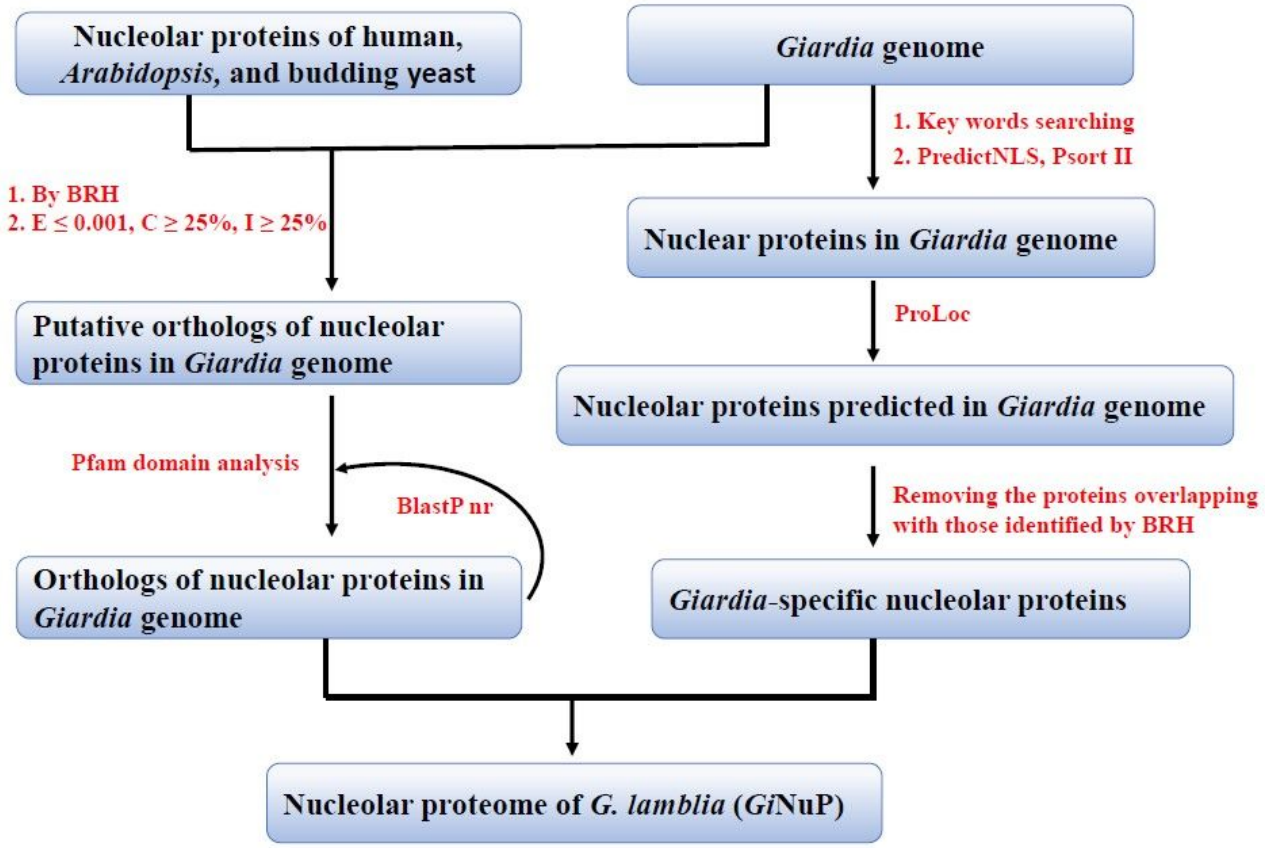

Figure 5

The flow chart of the computational identification of G. lamblia nucleolar proteins and the reconstruction of G. lamblia nucleolar proteome (GiNuP). E: E-value, C: coverage value, I: Identity value. BRH: Best Reciprocal Hit.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS2.799ortholog.xlsx
- TableS1.GiNuP.xlsx