

Distribution, organization an evolutionary history of La and LARPs in eukaryotes

Jean-Marc Deragon

▶ **To cite this version:**

Jean-Marc Deragon. Distribution, organization an evolutionary history of La and LARPs in eukaryotes. RNA Biology, 2020, pp.1-9. 10.1080/15476286.2020.1739930 . hal-02521074

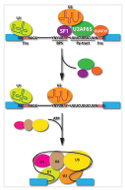
HAL Id: hal-02521074
https://univ-perp.hal.science/hal-02521074

Submitted on 27 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RNA Biology



ISSN: 1547-6286 (Print) 1555-8584 (Online) Journal homepage: https://www.tandfonline.com/loi/krnb20

Distribution, organization an evolutionary history of La and LARPs in eukaryotes

Jean-Marc Deragon

To cite this article: Jean-Marc Deragon (2020): Distribution, organization an evolutionary history of La and LARPs in eukaryotes, RNA Biology, DOI: 10.1080/15476286.2020.1739930

To link to this article: https://doi.org/10.1080/15476286.2020.1739930

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 19 Mar 2020.

Submit your article to this journal

Article views: 80

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=krnb20

RNA BIOLOGY
https://doi.org/10.1080/15476286.2020.1739930

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS Check for updates

Distribution, organization an evolutionary history of La and LARPs in eukaryotes

Jean-Marc Deragon

LGDP-UMR5096, Université de Perpignan Via Domitia, Perpignan, France; CNRS LGDP-UMR5096, Perpignan, France; Institut Universitaire de France, Paris, France

ABSTRACT

The fate of any cellular RNA is largely influenced by the nature and diversity of its interactions with various RNA-binding proteins (RBPs) leading to the formation of a biologically significant ribonucleoprotein (RNP) complex. La motif-containing proteins (composed of genuine La and La-related proteins (LARPs)) represent an evolutionary conserved family of RBPs that encompass a large range of crucial functions, involving coding and non-coding RNAs. In this work, we provide data that extend our previous knowledge on the distribution, organization and evolutionary history of this important protein family. Using a repertoire of 345 La motif-containing proteins from 135 species representing all major eukaryotic lineages, we were able to pinpoint many lineage-specific variations in the structural organization of La and LARPs and propose new evolutive scenarios to explain their modern genomic distribution.

ARTICLE HISTORY
Received 24 October 2019
Revised 24 February 2020
Accepted 4 March 2020

KEYWORDS
RNA-binding proteins; genuine La; La-related proteins; LARP; evolution

Introduction

RNA-binding proteins (RBPs) are key determinants of coding and non-coding RNA metabolism. La motif-containing proteins are part of an important RBP family that emerged early in eukaryotic evolution and acquired specialized functions [1–3]. The genuine La protein (also named the La autoantigen) is the most ubiquitous member of this family. Genuine La proteins have been shown to protect the 3′-end of nascent RNA polymerase III transcripts and, in many occasions, to act as an RNA chaperone to prevent RNA misfolding [4,5]. La-Related Protein (LARP) 7 targets specialized non-coding RNAs (Vertebrates 7SK [6,7] or Ciliates [8,9] and Fungi [10–12] telomerase RNA) to promote their assembly into a functional ribonucleoprotein complex. Other LARPs (1, 4 and 6), as well as the cytoplasmic isoform of the genuine La, target mRNAs and regulate their cellular localization, stability and translation potential [13–19]. Several extensive recent reviews have been published on the functions of genuine La and LARPs [1,13,20] and this aspect will not be developed further here. In light of the recent accumulation of genomic data, the objective of this work is to re-evaluate the distribution of genuine La and LARPs using data from 135 species, representing all major eukaryotic lineages (Metazoans, Embryophytes, Chlorophytes, Fungi, Amoebozoa, Alveoloates, Excavates, Stramenopiles, Rhizaria and Hacrobia). We also determined the phylogenetic relationships of the 345 La and LARP proteins collectively and revisited the structural organization of each subfamily.

Results

Genuine La

A ‘classical’ genuine La protein contains a La motif, closely followed by a first RRM (RRM1), by a more distal second atypical RRM (RRM2) and by a C-terminal GK/R-rich basic motif. The combination of the La/RRM1 is usually referred to as the La Module (LaM). This classical organization is found in all Chlorophyte, Embryophyte and animal (vertebrate and invertebrate) species without any detected exception. All animal genomes possess a single genuine La gene. However, in Embryophytes, although Bryophytes, Lycophytes, basal vascular plants and most Eudicots present a single La gene per species, all Poaceae (BstLa, OsLa, PhLa, SiLa) and Brassicaceae (AhLa, AlLa, AtLa, BsLa, EsLa) species have two (Figure 1). In Fungi, the organization of genuine La proteins can be different from the classical one described above. For example, Ascomycota La are usually shorter and do not possess an RRM2. However, other Fungi possess longer genuine La, many of which present a conserved RRM2. This is the case for many Basidiomycota (CcLa, LbiLa, MosLa, PgLa, WhLa, XdLa), Mucoromycota (EspLa, LtrLa, RiLa) and Zoopagomycota (BmeLa) genuine La proteins that have a ‘classical’ organization (Figure 1). The situation in basal eukaryote lineages is also variable. While genuine La proteins from Stramenopiles (PsLa, TcLa), Ameobozoa (HaLa, DdLa), Rhizaria (PbaLa) and Excavates (NgLa) have the classical structure (LaM, RRM2, GK/R-rich motif), Hacrobia (EhLa, GtLa) La miss the RRM2 (Figure 1). The organization of what appears to be Alveolates genuine La proteins is quite unusual. These proteins (PmLa, SleLa, ThLa, TgLa) contain a La motif, clearly related to other eukaryotic genuine La proteins, but no RRM nor GK/R-rich motif. Instead, they possess a conserved C-terminal region of unknown function named DUF3223 (see Figure S1). The La/DUF3223 association is specific to Alveolates, with a single-identified exception for the Hacrobia species *Chrysochromulina tobinii* that also possesses a La/DUF3223 containing protein (Figure S1)

CONTACT Jean-Marc Deragon jean-marc.deragon@univ-perp.fr LGDP-UMR5096, Université de Perpignan Via Domitia, 58 Av. Paul Alduy, Perpignan 66860, France

Supplemental data for this article can be accessed here

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

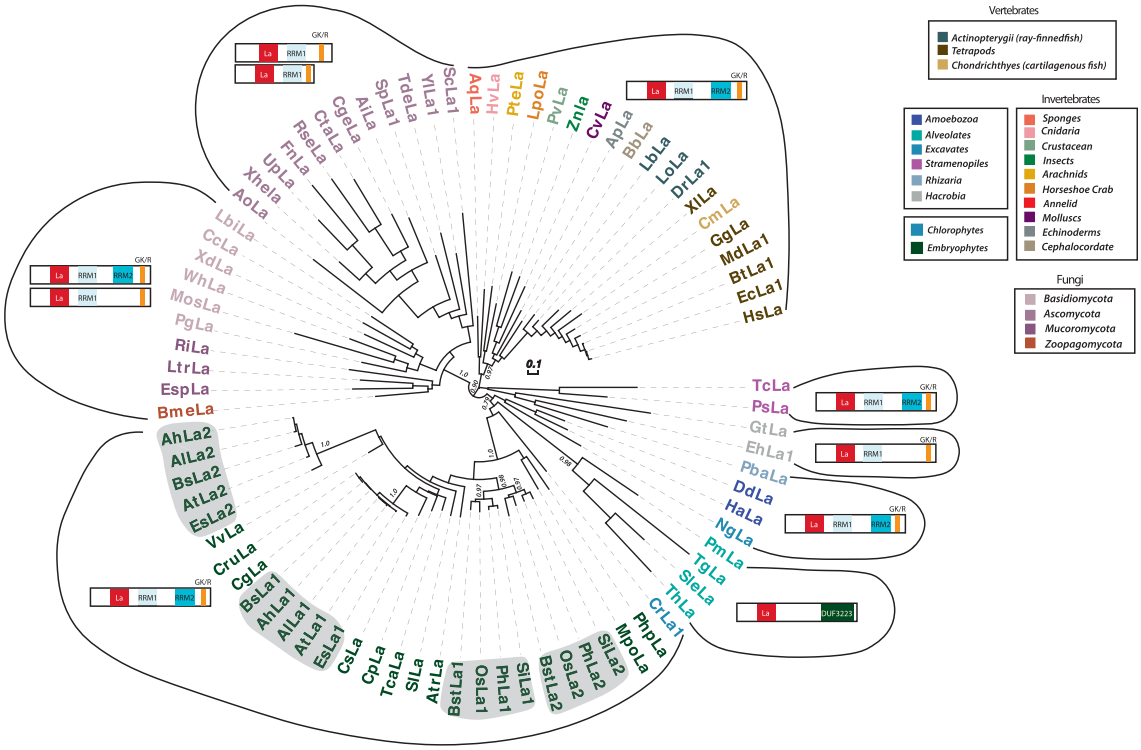


Figure 1. Distribution, phylogenetic relationships and structural organization of genuine La in eukaryotes. Full-length La protein sequences from the different eukaryotic lineages were aligned and used to construct the phylogenetic tree (with the exception of Alveolates La for which the La motif alone was used). Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated. Scale bar indicates length of 0.1 substitution/site. For a description of species and protein sequences used, see Supplemental Table 1. Boxes on the right define how the different lineages were colour-coded. The grey shading is used to highlight the presence of two genuine La protein subgroups in Poaceae and Brassicaceae. Positions of the different conserved regions (on the Human La sequence (NP_003133.1)) are the following: La motif (10–91), RRM1 (110–185), RRM2 (232–326) and GK/R-repeats (332–361).

LARP1

Genes encoding LARP1 were detected in Amoebozoa, Fungi, Embryophytes, Chlorophytes and Metazoans. Since genomic data from several basal eukaryotic lineages are still scarce, the presence of LARP1 in other eukaryote lineages than the ones reported here cannot be completely excluded at this time. LaM, DM15, as well as two additional regions (named C-terminal Region 1 and 2 (CR1 and CR2)), are conserved in all full-length eukaryotic LARP1 s (Figure 2 and S2). The RRM region, adjacent to the La motif, is conserved at the primary sequence level only in closely related vertebrate species. In other eukaryotes, its presence was inferred solely from secondary structure predictions[2]. The putative unconventional PAM2 motif, identified as part of the human LARP1 RRM region[21], is only conserved in vertebrates (see Figure S2). In Embryophytes and Fungi, in addition to full-length LARP1, shorter LARP1 versions, with a C-terminal La motif as the sole conserved region, can also be found. While basal Embryophyte (AtrLARP1, MpoLARP1, PhpLARP1), as well as Chlorophyte (CrLARP1) species have only one full-length LARP1, all Angiosperms possess one full-length and one to several of these solo La motif versions of LARP1. In Fungi, the situation is more complex. In basal Fungi lineages (Cryptomycota (PsaLARP1), Zoopagomycota (BmeLARP1, SmeLARP1, TspLARP1) and Mucoromycota (ArLARP1, LtrLARP1, RiLARP1)), only full-length LARP1 is present while in Basidiomycota, only shorter versions (with a C-terminal solo La motif) are present. In Ascomycota, the situation is intermediate, with two sub-lineages (Saccharomycotina and Taphrinomycotina) having only shorter versions (with the curious exception of one Taphrinomycotina species Neolecta irregularis (NiLARP1) that has only a full-length LARP1), and one sub-lineage (Pezizomycotina, DfLARP1, EgrLARP1, GlLARP1, McLARP1, UpLARP1) having both the full-length and shorter types. However, full-length LARP1 from Pezizomycotina are less conserved compared to their orthologues from basal Fungi lineages (illustrated by long branches for these species in Figure 2). Other LARP1 lineage-specific variations include a very short N-terminal region for Amoebozoa LARP1 s (DdLARP1, HaLARP1, TlLARP1) and the presence of conserved glycine-arginine (GR) repeats, N-terminal to the La motif, for Metazoans and (full-length) Fungi LARP1. A gene duplication event also likely occurred in the common ancestor of vertebrates generating two distinct versions of full-length LARP1 (named LARP1A and B, grey shading in Figure 2). This is the only situation where two distinct full-length LARP1 versions co-exist in the same lineage.

LARP4

Genes coding for LARP4 are found in Amoebozoa, Stramenopiles, Excavates as well as Metazoans, but are apparently absent from other eukaryotic lineages, including Embryophytes, Chlorophytes and Fungi. Most vertebrates

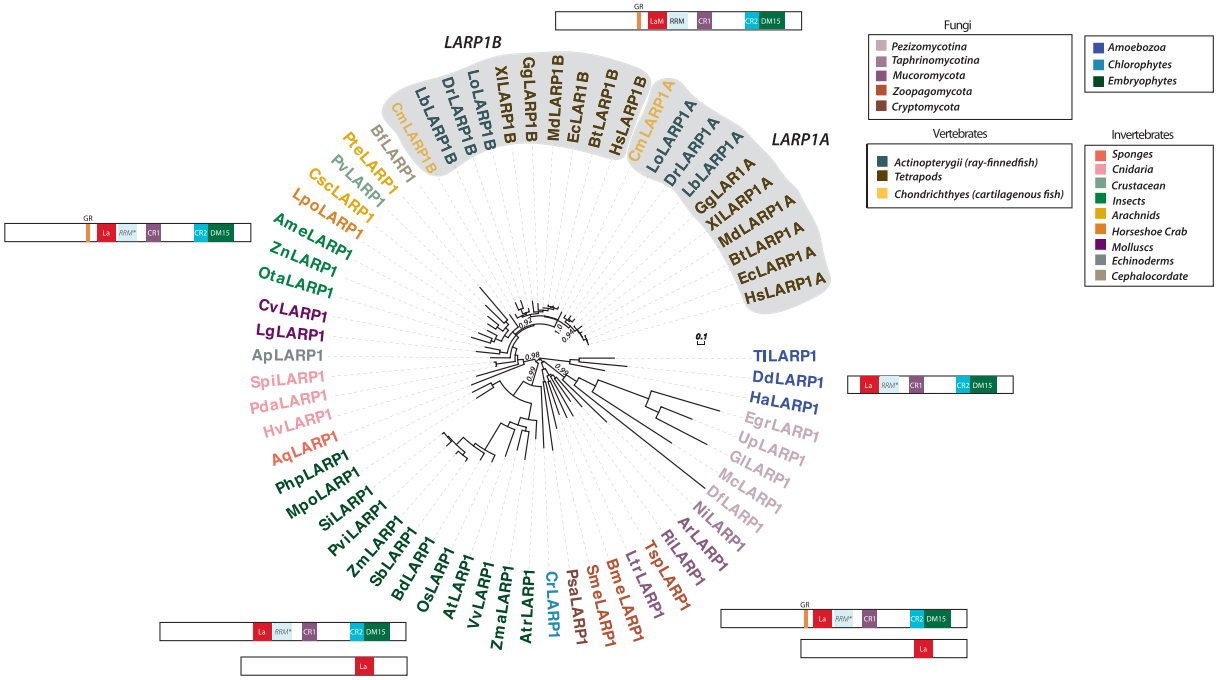


Figure 2. Distribution, phylogenetic relationships and structural organization of LARP1 in eukaryotes. Only full-length LARP1 protein sequences from the different eukaryotic lineages were aligned and used to construct the phylogenetic tree. Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated. Scale bar indicates length of 0.1 substitution/site. For a description of species and protein sequences used, see Supplemental Table 1. Boxes on the right define how the different lineages were colour-coded. The grey shading is used to highlight the presence of two LARP1 (LARP1A and LARP1B) subgroups in vertebrates. Positions of the different conserved regions (on the Human LARP1A sequence (NP_291029.2)) are the following: GR repeats (338–353), LaM (401–478), RRM (500–559), CR1 (603–687), CR2 (825–884) and DM15 (885–999). The RRM region, adjacent to the LaM, is conserved at the primary sequence level only in vertebrates. In other eukaryotes, its presence was inferred from secondary structure prediction only[2] and is represented by *RRM**.

possess two related but distinct LARP4 versions (LARP4A and B) while some fishes (but not all) have three versions (two of the A types (A1 and A2) and one of the B types) (see Figure 3). The LARP4 LaM module is the only region conserved in all LARP4s, while Metazoans present one to several other conserved motifs. Indeed, invertebrate LARP4s have a single supplementary conserved region (CR1, Figure S3), while vertebrate LARP4s have four (PAM2 w, PBM, CR1 and CR2). The existence of a consensus PAM2 motif in some invertebrates was observed[22], but overall this motif is not present in invertebrate LARP4s. Human LARP4A PAM2 w and PMB have been shown to be functional PolyA-Binding Protein (PABP)-interacting regions[23]. Also, a receptor for activated C kinase (RACK1) interacting region (RIR) was identified near the C-terminus of human LARP4A[1]. It remains to be seen if this RIR corresponds to the conserved CR2 identified here (Figure S3). LARP4A and B version have similar conserved regions but with some significant differences in the precise organization of their La module[1], in the amino acid sequence of their CR1 and CR2 regions (Figure S3) and in the size and primary amino acids conservation of their N-terminal regions (NTRs). LARP4A has a short unconserved region upstream of the PAM2 w, while LARP4B corresponding region is longer (designated by P2 + in Figure 3) and very well conserved (Figure S3). In a recent work, the human LARP4A NTR (position 1–111) was showed to adopt a semi-disordered state and function as RNA-binding platform, in combination with the RRM1 region[24]. The secondary structure propensity of this NTR and the conservation of the PAM2 w motif are both important determinants for RNA-binding. It would be interesting to test if differences observed in LARP4A and LARP4B NTRs reflect different RNA-binding properties of these two proteins.

LARP6

LARP6 genes are present basal Stramenopiles, in Chlorophytes, Embryophytes and Metazoans but are apparently absent from other eukaryotic lineages. Most Embryophytes possess three distinct LARP6 versions (green shading in Figure 4). For Metazoans, although invertebrates possess a single LARP6 version, most vertebrates have two (LARP6A and B, grey shading in Figure 4), with two exceptions: some fishes that have three versions (LARP6A, B1 and B2) and Eutherians that have only one (of the A type) (see Figure 4). All eukaryote LARP6 s present three highly conserved regions, the La motif and adjacent RRM (forming the LaM) and the C-terminal **L**AM and **S**1 **A**ssociated motif (LSA), that represent a specific diagnostic region of the LARP6 group. All Embryophyte LARP6s present an additional conserved GK/R-rich motif between the RRM and the LSA. Bony fish LARP6As also present a much longer N-terminal region compared to other LARP6s, but without any clear supplementary conserved regions.

LARP7

Full-length LARP7 s have a similar organization as genuine La proteins with a La motif, an RRM1 and an atypical RRM2. However, the spacer region between the LaM and the RRM2 is

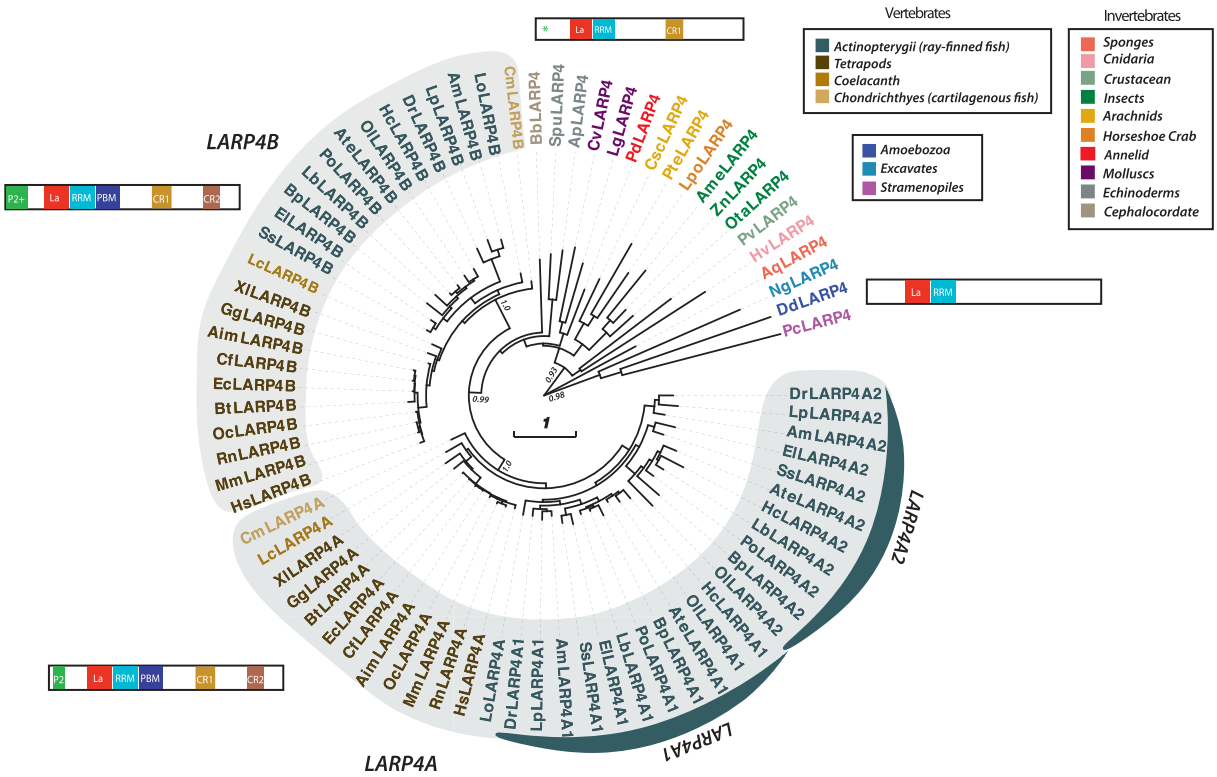


Figure 3. Distribution, phylogenetic relationships and structural organization of LARP4 in eukaryotes. Full-length LARP4 proteins from the different eukaryotic lineages were aligned and used to construct the phylogenetic tree. Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated. Scale bar indicates length of 1 substitution/site. For a description of species and protein sequences used, see Supplemental Table 1. Boxes on the right define how the different lineages were colour-coded. The grey shading is used to distinguish between vertebrate LARP4As and LARP4Bs from other LARP4s. Positions of the different conserved regions (on the Human LARP4A sequence (NP_443111.4)) are the following: P2 (including the PAM2 w, 8–31), La (117–195), RRM (200–271), PMB (281–349), CR1 (477–489), CR2 (613–629). For LARP4B, sequence conservation was observed from the N-terminal end to the PAM2 w motif (Position 1 to 71 on the Human LARP4B sequence (NP_055970.1)) and is indicated as P2 +. A few invertebrate LARP4s have a consensus PAM2 motif (indicated by a green asterisk)[22] but overall this region is not conserved in invertebrates.

much longer for LARP7 compared to genuine La, and the RRM2 is closer to the C-terminal end of the protein. LARP7s are present in all animals and in Amoebozoa (HaLARP7, DdLARP7), Alveolates (EaLARP7, SleLARP7, ThLARP7) and Stramenopiles (PsLARP7, TcLARP7) (Figure 5), but not all LARP7s belonging to these three later lineages host a conserved RRM2 region. Recently, several reports described the presence of a LARP7 (named Pof8) in fission yeast[10–12]. The only significant primary sequence conservation between Pof8 and other eukaryote LARP7s resides in the short RRM2 region, and the presence of a LaM (La motif/RRM1) can only be inferred from secondary sequence predictions. Accordingly, apart from the RRM2 region, primary sequence homology to Pof8 was not found outside of a few *S. pombe* sister species. Using a phylogenetic approach, we were able to confirm that the Pof8 RRM2 is indeed of the LARP7 type (and not of the genuine La type) (not shown). We next used Pof8 RRM2 as a bait, to detect proteins containing a similar RRM2 in species outside of the Taphrinomycotina lineage. Pof8-related RRM2 regions were detected in proteins from most Basidiomycota species. Surprisingly, except for species belonging to Pezizomycetes, no such protein could be detected in Ascomycota nor in other more basal Fungi lineages. These Pof8-related RRM2 regions are not only similar at the primary sequence level but also likely at the secondary structure level (see Figure 6). Some of these Pof8-like proteins are also predicted to adopt a LaM structure upstream of the RRM2 (not shown), suggesting that they could be a true orthologue of the *S. pombe* Pof8 protein. Others are however too short to host a LaM or are not predicted to fold as a LaM.

Discussion

This work provides new information on the detailed distribution and evolutionary history of La and LARPs in all major eukaryotic lineages. It also identifies several new conserved domains in genuine La and LARP1, 4 and 6 (see Figures 1 to 4 and S1 to S4) that could be the target of future functional studies.

One of the most surprising results of this study is the discovery that classical genuine La proteins are apparently absent from Alveolates, a major eukaryotic lineage. LARP7 proteins belonging to this lineage have been well characterized in two species (*Tetrahymena thermophila* and *Euplotes aediculatus*) [8,9], however, no genuine La from Alveolates have been reported so far. We find here that Alveolates possess an atypical La protein, with a conserved La motif, clearly grouping with classical genuine La motifs (see Figure 1). However, this 'genuine La-like motif' is not associated with RRMs but with a conserved motif of unknown function called DUF3223 (also named DeCl). DUF3223 is also detected in the C-terminal region of two plant-specific RNA Polymerases

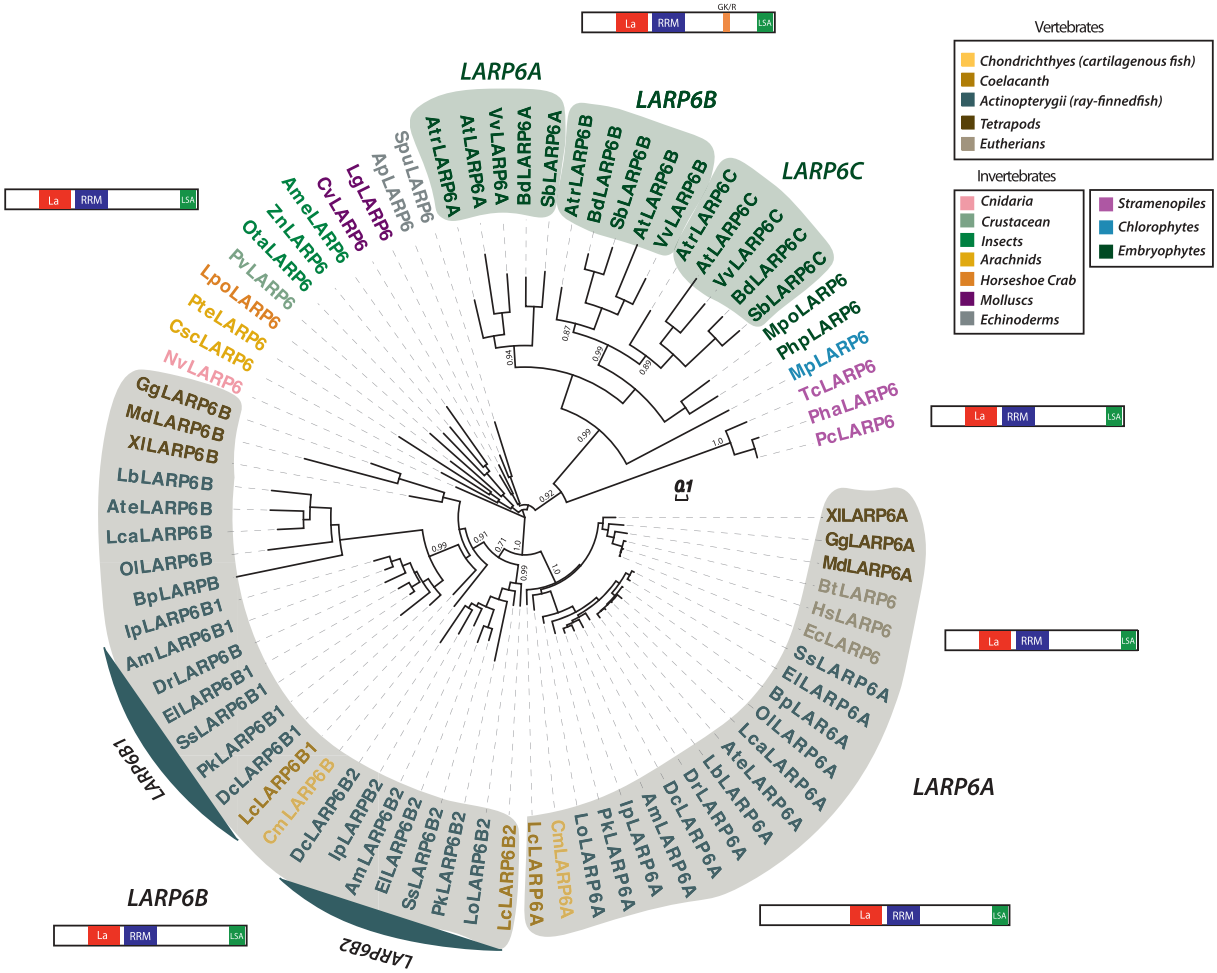


Figure 4. Distribution, phylogenetic relationships and structural organization of LARP6 in eukaryotes. Full-length LARP6 proteins from the different eukaryotic lineages were aligned and used to construct the phylogenetic tree. Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated. Scale bar indicates length of 0.1 substitution/site. For a description of species and protein sequences used, see Supplemental Table 1. Boxes on the right define how the different lineages were colour-coded. The grey shading is used to distinguish between vertebrate LARP6As and LARP6Bs from other LARP6s. The green shading is used to highlight the presence of three plant specific LARP6 subgroups (LARP6A, B and C). Positions of the different conserved regions (on the Human LARP6 sequence (NP_060827.2)) are the following: La (90–169), RRM (185–275), LSA (467–482). Position of the Embryophyte specific GK/R conserved region (on the *Arabidopsis thaliana* LARP6C sequence (At3g19090)) is from amino acids 379 to 388.

(Pol IV/V)[25], in three plant proteins likely involved in nuclear, chloroplastic and mitochondrial rRNA processing and ribosome assembly [26–28] and in many bacteria (Figure S1). The function of the DUF3223 is unknown, but this conserved region has been proposed to be involved in binding proteins and RNA, at least in the context of the plant PolV enzyme[25]. It is therefore tempting to speculate that these unusual La/DUF3223 containing proteins may represent the atypical Alveolates genuine La and that this combination of conserved regions may form a new and original RNA-binding surface. Functional studies will be needed to determine if La/DUF3223 proteins are able to bind RNA and fulfill classical genuine La functions in Alveolates.

Another intriguing aspect is the presence of two genuine La versions in two Embryophyte lineages (Brassicaceae and Poaceae) (see Figure 1). This situation likely results from two independent gene duplication events, one in the common ancestor of Poaceae (around 76MY ago), the other in the common ancestor of Brassicaceae (around 37MY ago). Although this second event is much younger than the previous one, Brassicaceae paralogues are much more divergent compared to the ones from Poaceae (see Figure 1). This suggests that a neofunctionalization event occurred in the Brassicaceae lineage that included a fast-evolving period. This hypothesis is supported by data showing that only one of the two *Arabidopsis thaliana* genuine La proteins is involved in tRNA maturation and is essential for viability[29]. It remains to be seen what the function of the second Brassicaceae La copy could be, and if the closely related Poaceae La paralogues are fully redundant or not.

The study of LARP1 confirmed the presence in all vertebrates of two distinct LARP1 versions (LARP1A and B), a situation previously described in human. All works describing LARP1 function in mammals have been done on LARP1A and the function of the LARP1B version in vertebrates remains to be studied. An intriguing aspect of LARP1 evolution is the presence in two lineages (Fungi and Embryophytes) of proteins containing a solo La motif (of the LARP1 type) in a C-terminal position. *Saccharomyces cerevisiae* bears two of these solo La motif LARP1, named Slf1 and Sro9p, that are both involved in the regulation of mRNA localization, stability and translation

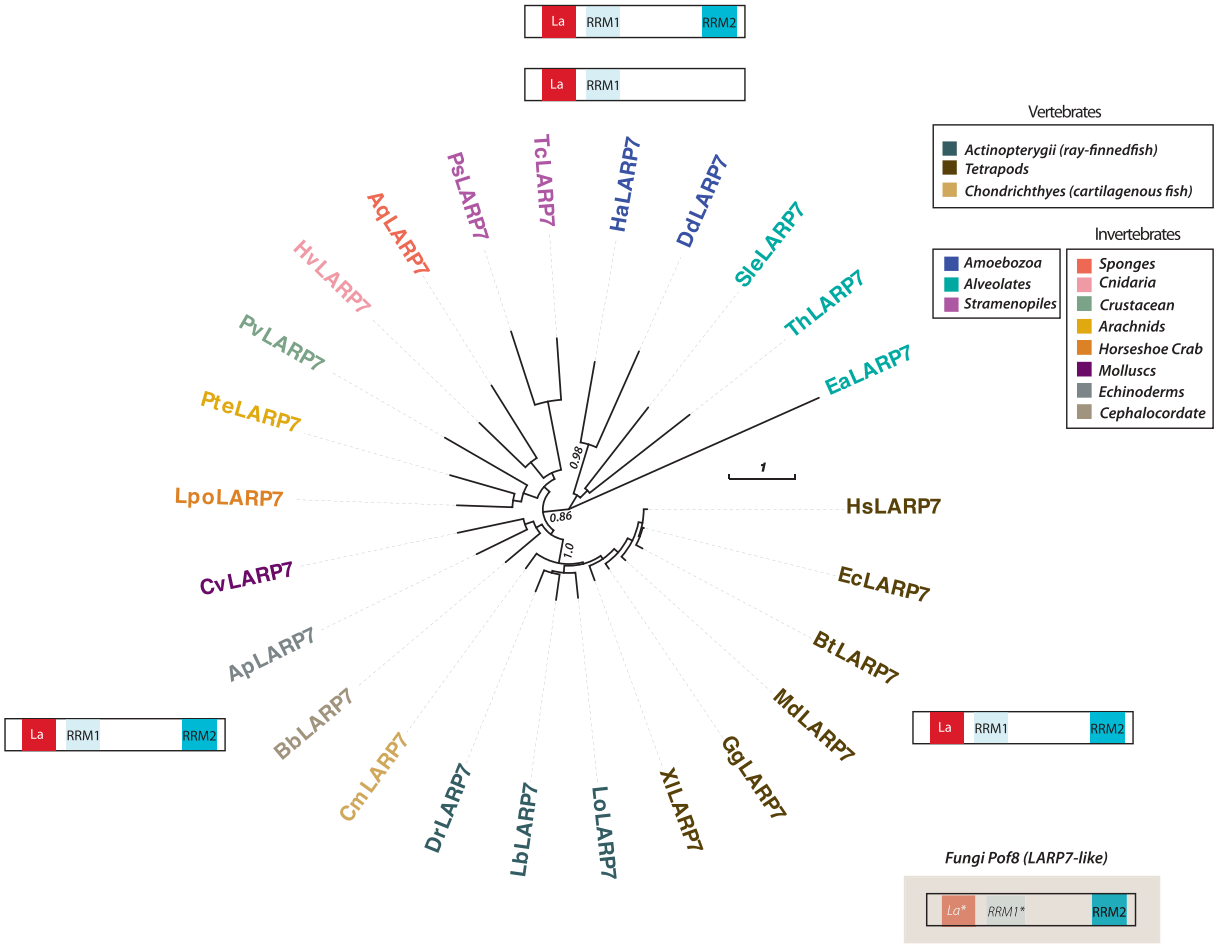


Figure 5. Distribution, phylogenetic relationships and structural organization of LARP7 in eukaryotes. Full-length LARP7 proteins from the different eukaryotic lineages were aligned and used to construct the phylogenetic tree. The Fungi LARP7-related (Pof8) proteins were not used for the phylogenetic analysis since their homology with other LARP7 is limited to the short RRM2 region. Fungi LARP7-like La motif and RRM1 were only inferred from secondary structure predictions and are represented by *La** and *RRM1** in the corresponding cartoon. Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated. Scale bar indicates length of 1 substitution/site. For a description of species and protein sequences used, see Supplemental Table 1. Boxes on the right define how the different lineages were colour-coded. Positions of the different conserved regions (on the Human LARP7 sequence (NP_056269)) are the following: La (40–111), RRM1 (124–199), RRM2 (450–558).

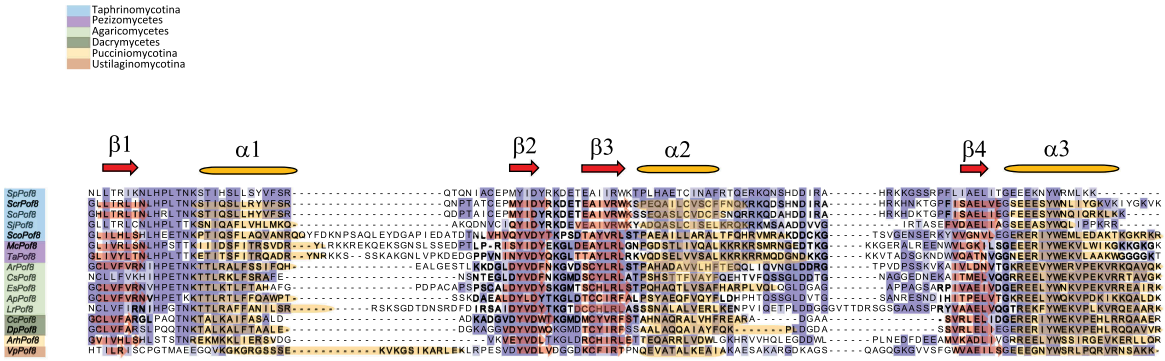


Figure 6. Sequence alignment of Pof8 (LARP7-like) RRM2 from different Fungi lineages showing primary and secondary sequence conservation. The amino acids are coloured in blue based on their conservation. Experimentally determined positions of the *S. Pombe* Pof8 (SpPof8) secondary structure[39] are schematically represented above the alignment. For other sequences, secondary structure predictions, inferred using the Phyre2 software are superimposed. The origin of each sequence is colour-coded.

[13,30–32]. The La motif of Slf1 and Sro9p is necessary but not sufficient to bind RNAs, and the additional *cis* and/or *trans* factor(s) needed to generate the required RNA-binding surface have not been detected so far[30]. In Embryophyte genomes, solo La motif LARP1 are systematically present in association with classical full-length LARP1. LARP1 distribution in Embryophytes and Fungi suggests that solo La motif LARP1 versions appeared two times independently during eukaryote

evolution, once in the common ancestor of Angiosperms and once in the common ancestor of Dikarya (at the origin of Basidiomycota and Ascomycota). While in Angiosperms, both full-length and solo La motif versions were kept in all species, in Dikarya the solo version is the rule, as full-length LARP1 were most often lost or largely reorganized (in Pezizomycotina). It remains to be seen if Embryophytes and Fungi solo La motif LARP1s have similar functions (i.e. convergent evolution). Also, the rapid divergence of full-length LARP1 in Pezizomycotina could be symptomatic of a neofunctionalization event. It would therefore be interesting to determine if these more divergent proteins are still fulfiling the same functions as other full-length eukaryotic LARP1s.

The evolutionary history of LARP4 and LARP6 is more complex compared to the ones of genuine La and other LARPs (the two evolutionary scenarios described below are summarized in Figure 7). For LARP4, its unusual distribution in eukaryotes suggests the following evolutionary scenario (Figure 7A). A LARP4 gene was present very early in eukaryote evolution, as suggested by its presence today in some basal eukaryotic lineages (Excavates, Amoebozoa, Stramenopiles). However, contrarily to LARP1, LARP4 gene was loss in Embryophytes, Chlorophytes and Fungi. A first gene duplication event occurred in the common ancestor of vertebrates generating the two LARP4 versions (A and B). A second gene duplication, targeting the LARP4A version, occurred in fishes but not at the onset of this class (around 500MY ago) but later in the common ancestor of the Clupeocephala group (around 250MY ago). This results in cartilaginous fishes (Chondrichthyes) and some bony fishes belonging to basal orders (such as Coelacanthiformes, Polypteriformes, Acipenserriformes, Lepisosteiformes and Osteoglossomorpha) having a single LARP4A gene while other fishes have two (see Figures 3 and 7).

The evolutionary history of the LARP6 family is also complex and includes several lineage-independent gains and losses (see Figures 4 and 7B). A LARP6 gene was also likely present early in eukaryote evolution, as suggested by its presence today in Stramenopiles. In Embryophytes, in the common ancestor of Angiosperms, gene duplication events led to the emergence of three LARP6 versions (A, B and C). These three versions are systematically present in all modern Angiosperm species. The LARP6A version is always encoded by a single copy gene, but LARP6B and C versions are sometimes encoded by multiple paralogous genes. In Metazoans, while invertebrates possess a single LARP6 gene, a gene duplication event (independent from the one proposed in Embryophytes) likely occurred in the common ancestor of vertebrates, generating LARP6A and B types. Vertebrates LARP6B type was further duplicated in fishes, in the common ancestor of Actinopterygii (around 375 MY), to generate the B1 and B2 types. However, the LARP6B2 type was lost two times in evolution, once precisely in Cypriniformes (that include the model species *Danio rerio*), and a second time in the common ancestor of Neoteleostei (around 180 MY). Therefore, Cypriniformes and Neoteleostei present only a single LARP6B of the B1 type. The loss of the LARP6B2 type in the common ancestor of Neoteleostei was apparently associated with the divergence of the remaining LARP6B (see the longer branch length for AteLARP6B, LbLARP6B, LcaLARP6B, BpLARP6B and OlLARP6B on the tree of Figure 4 compared to other fish LARP6 s). Another independent gene loss, this time of the LARP6B type, occurred in the common ancestor of Eutherians, leaving a single LARP6 (of the A type) in these species.

Recently, a LARP7-like protein (Pof8) was described in *Schizosaccharomyces pombe* and in a few Taphrinomycotina sister species [10–12]. Primary sequence homology of Pof8 to other eukaryotic LARP7s is limited to the RRM2 region and consequently, Pof8 LaM was only inferred from secondary structure predictions. Although most LARP1 proteins possess RRM1 regions conserved only at the secondary sequence level, this is

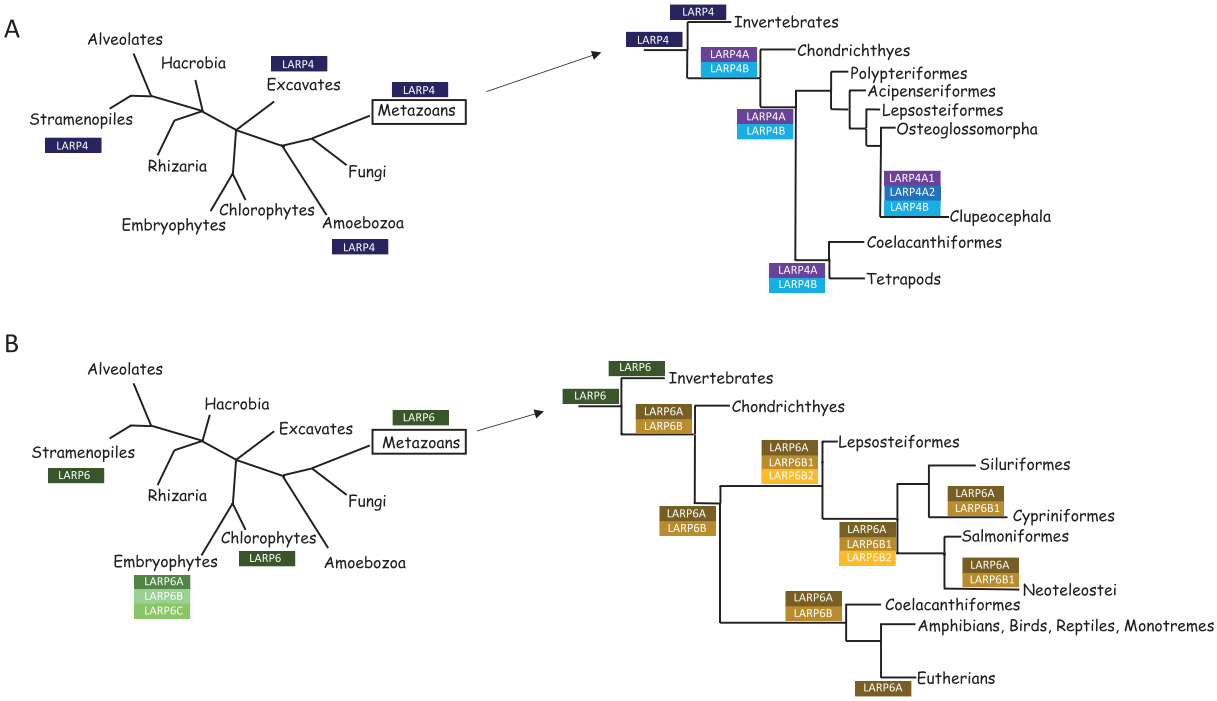


Figure 7. Diagrammatic representation of the LARP4 (A) and LARP6 (B) putative evolutionary scenarios. See the Discussion section for a description of the two scenarios.

the first time that a La-related protein presents only secondary structure conservation of the complete LaM. This suggests that some functions fulfilled by LARPs require only partial (most LARP1s) or no (Pof8) primary conservation of the LaM, while still needing the correct folding of this region. Using the *S. pombe* pof8 RRM2 sequence, we were able to identify several Pof8-related proteins in three other Fungi lineages (see Supplemental Table 1 and Figure 6). Pof8-related proteins all possess primary and likely secondary structure homologies to the *S. pombe* RRM2 (see Figure 6). However, some of these Pof8-related proteins are too short to host a LaM or were not predicted to fold as a LaM. In these cases, one possibility would be that they could interact *in trans* with a yet to identify partner to provide the missing structures. The patchy distribution of Pof8-like proteins in Fungi still remains difficult to explain. Indeed, although Pof8-like proteins are very well represented in Basidiomycota, they are absent from most Ascomycota lineages (except for Taphrinomycotina and Pezizomycetes) and from basal Fungi lineages. Since Pof8 is involved in telomerase assembly[10–12], the loss of Pof8-like proteins may indicate functional differences in the way telomerase associates with its telomerase RNA in these Pof8-free lineages.

Methods

Blast searches (blastp and tblastn) were performed starting from known La and LARPs conserved domains on the genome of 135 species representing all major eukaryotic lineages using the JGI (https://jgi.doe.gov) and NCBI (https://www.ncbi.nlm.nih.gov) genomic resources. Full-length protein sequences were aligned using the multiple sequence comparison by log-expectation (MUSCLE v3.7) software[33]. The resulting alignment was processed using the TRIMAI software [34] and trees were reconstructed using the fast maximum likelihood tree estimation program PHYML [35] using the LG amino acids replacement matrix [36]. Statistical support for the major clusters was obtained using the approximate likelihood-ratio test (aLRT) [37]. Protein secondary structure predictions were obtained using the Phyre2 (V2) software[38].

Acknowledgments

Thanks to Remy Merret and Jean-Jacques Favory, Mark Bayfield and Jennifer Porat for helpful discussion and critical reading of the manuscript. This work is set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP (ANR-10-LABX-41).

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by the Centre National de la Recherche Scientifique [2019]; Institut Universitaire de France (IUF) [2016-2021]; Université de Perpignan Via Domitia [2019]

References

1. Maraia RJ, Mattijssen S, Cruz-Gallardo I, et al. The La and related RNA-binding proteins (LARPs): structures, functions, and evolving perspectives. Wiley Interdiscip Rev RNA. 2017;8(6):e1430.
2. Bousquet-Antonelli C, Deragon J. A comprehensive analysis of the La-motif protein superfamily. RNA. 2009;15(5):750–764.
3. Bayfield MA, Yang R, Maraia RJ. Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). Biochim Biophys Acta. 2010;1799(5–6):365–378.
4. Wolin SL, Cedervall T. The La protein. Annu Rev Biochem. 2002;71(1):375–403.
5. Chakshusmathi G. A La protein requirement for efficient pre-tRNA folding. EMBO J. 2003;22(24):6562–6572.
6. Krueger BJ, Jeronimo C, Roy BB, et al. LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic Acids Res. 2008;36(7):2219–2229. DOI:10.1093/nar/gkn061.
7. Markert A, Grimm M, Martinez J, et al. The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. EMBO Rep. 2008;9(6):569–575. DOI:10.1038/embor.2008.72.
8. Aigner S, Postberg J, Lipps HJ, et al. The Euplotes La motif protein p43 has properties of a telomerase-specific subunit. Biochemistry. 2003;42(19):5736–5747.
9. Witkin KL. Holoenzyme proteins required for the physiological assembly and activity of telomerase. Genes Dev. 2004;18(10):1107–1118.
10. Páez-Moscoso DJ, Pan L, Sigauke RF, et al. Pof8 is a La-related protein and a constitutive component of telomerase in fission yeast. Nat Commun. 2018;9(1):587.
11. Mennie AK, Moser BA, Nakamura TM. LARP7-like protein Pof8 regulates telomerase assembly and poly(A)+TERRA expression in fission yeast. Nat Commun. 2018;9(1):586.
12. Collopy LC, Ware TL, Goncalves T, et al. LARP7 family proteins have conserved function in telomerase assembly. Nat Commun. 2018;9(1):557. DOI:10.1038/s41467-017-02296-4.
13. Deragon J-M, Bousquet-Antonelli C. The role of LARP1 in translation and beyond. Wiley Interdiscip Rev RNA. 2015;6(4):399–417.
14. Merret R, Descombin J, Juan YT, et al. XRN4 and LARP1 are required for a heat-triggered mRNA decay pathway involved in plant acclimation and survival during thermal stress. Cell Rep. 2013;5(5):1279–1293. DOI:10.1016/j.celrep.2013.11.019.
15. Zhang Y, Stefanovic B. LARP6 Meets Collagen mRNA: Specific Regulation of Type I Collagen Expression. Int J Mol Sci. 2016;17:419.
16. Intine RV, Tenenbaum SA, Sakulich AL, et al. Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs. Mol Cell. 2003;12(5):1301–1307.
17. Mattijssen S, Arimbasseri AG, Iben JR, et al. LARP4 mRNA codon-tRNA match contributes to LARP4 activity for ribosomal protein mRNA poly(A) tail length protection. Elife. 2017;6. DOI:10.7554/eLife.28889.
18. Küspert M, Murakawa Y, Schäffler K, et al. LARP4B is an AU-rich sequence associated factor that promotes mRNA accumulation and translation. RNA. 2015;21(7):1294–1305. DOI:10.1261/rna.051441.115.
19. Fonseca BD, Lahr RM, Damgaard CK, et al. LARP1 on TOP of ribosome production. Wiley Interdiscip Rev RNA. 2018;9(5):e1480.
20. Stavraka C, Blagden S. The La-Related Proteins, a Family with Connections to Cancer. Biomolecules. 2015;5(4):2701–2722.
21. Fonseca BD, Zakaria C, Jia JJ, et al. La-related Protein 1 (LARP1) Represses Terminal Oligopyrimidine (TOP) mRNA Translation Downstream of mTOR Complex 1 (mTORC1). J Biol Chem. 2015;290:15996–16020.
22. Merret R, Martino L, Bousquet-Antonelli C, et al. The association of a La module with the PABP-interacting motif PAM2 is a recurrent evolutionary process that led to the neofunctionalization of La-related proteins. RNA. 2013;19(1):36–50. DOI:10.1261/rna.035469.112.

23. Yang R, Gaidamakov SA, Xie J, et al. La-related protein 4 binds poly(A), interacts with the poly(A)-binding protein MLLE domain via a variant PAM2w motif, and can promote mRNA stability. Mol Cell Biol. 2011;31:542–556.
24. Cruz-Gallardo I, Martino L, Kelly G, et al. LARP4A recognizes polyA RNA via a novel binding mechanism mediated by disordered regions and involving the PAM2w motif, revealing interplay between PABP, LARP4A and mRNA. Nucleic Acids Res. 2019;47(8):4272–4291. DOI:10.1093/nar/gkz144.
25. Wendte JM, Haag JR, Singh J, et al. Functional Dissection of the Pol V Largest Subunit CTD in RNA-Directed DNA Methylation. Cell Rep. 2017;19(13):2796–2808.
26. Lahmy S, Guilleminot J, Cheng CM, et al. DOMINO1, a member of a small plant-specific gene family, encodes a protein essential for nuclear and nucleolar functions. Plant J. 2004;39(6):809–820. DOI:10.1111/tpj.2004.39.issue-6.
27. Bellaoui M, Keddie JS, Gruissem W. DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. Plant Mol Biol. 2003;53:531–543.
28. Wang Y, Ma H. Step-wise and lineage-specific diversification of plant RNA polymerase genes and origin of the largest plant-specific subunits. New Phytol. 2015;207(4):1198–1212.
29. Fleurdedine S, Deragon J, Devic M, et al. A bona fide La protein is required for embryogenesis in Arabidopsis thaliana. Nucleic Acids Res. 2007;35(10):3306–3321. DOI:10.1093/nar/gkm200.
30. Kershaw CJ, Costello JL, Castelli LM, et al. The yeast La related protein Slf1p is a key activator of translation during the oxidative stress response. PLoS Genet. 2015;11:e1004903.
31. Röther S, Burkert C, Brünger KM, et al. Nucleocytoplasmic shuttling of the La motif-containing protein Sro9 might link its nuclear and cytoplasmic functions. RNA. 2010;16(7):1393–1401.
32. Schenk L, Meinel DM, Strässer K, et al. La-motif-dependent mRNA association with Slf1 promotes copper detoxification in yeast. RNA. 2012;18(3):449–461.
33. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–1797.
34. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–1973.
35. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52:696–704.
36. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25(7):1307–1320.
37. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 2006;55(4):539–552.
38. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 2009;4(3):363–371.
39. Riwika Basu CDE, Cheng R, Feigon J. Structure of S. pombe telomerase protein Pof8 C-terminal domain is an xRRM conserved among LARP7 proteins. bioRxiv. 2019;739532. DOI:10.1101/739532.