

Proteomic LC–MS analysis of *Arabidopsis* cytosolic ribosomes: Identification of ribosomal protein paralogs and re-annotation of the ribosomal protein genes



Maureen Hummel^{a,b}, Thomas Dobrenel^{c,d}, Jan (J.H.G.) Cordewener^e, Marlène Davanture^f, Christian Meyer^d, Sjeff (J.C.M.) Smeekens^{a,g}, Julia Bailey-Serres^b, Twan (A.H.P.) America^{g,h}, Johannes Hanson^{a,c,*}

^a Molecular Plant Physiology, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

^b Center for Plant Cell Biology, Department of Botany and Plant Sciences, University of California, Riverside, CA 92521-0124, USA

^c Umeå Plant Science Center, Department of Plant Physiology, Umeå University, 90187, Umeå, Sweden

^d Institut Jean-Pierre Bourgin, UMR 1318 INRA AgroParisTech, Saclay Plant Sciences, F-78026 Versailles, France

^e BU Bioscience, Plant Research International, P.O. Box 619, 6700 AP Wageningen, The Netherlands

^f Plateforme PAPPSO, UMR de Génétique Végétale, Ferme du Moulon, Gif sur Yvette, France

^g Centre for BioSystems Genomics, P.O. Box 98, 6700 AB Wageningen, The Netherlands

^h Netherlands Proteomics Centre, P.O. Box 80082, 3508 TB Utrecht, The Netherlands

ARTICLE INFO

Article history:

Received 12 February 2015

Received in revised form 17 June 2015

Accepted 1 July 2015

Available online 29 July 2015

Keywords:

A. thaliana

LC–MS

DIA

Data-independent acquisition

Ribosomes

Ribosomal protein

Paralogs

ABSTRACT

Arabidopsis thaliana cytosolic ribosomes are large complexes containing eighty-one distinct ribosomal proteins (r-proteins), four ribosomal RNAs (rRNA) and a plethora of associated (non-ribosomal) proteins. In plants, r-proteins of cytosolic ribosomes are each encoded by two to seven different expressed and similar genes, forming an r-protein family.

Distinctions in the r-protein coding sequences of gene family members are a source of variation between ribosomes. We performed proteomic investigation of actively translating cytosolic ribosomes purified using both immunopurification and a classic sucrose cushion centrifugation-based protocol from plants of different developmental stages. Both 1D and 2D LC–MS^E with data-independent acquisition as well as conventional data-dependent MS/MS procedures were applied. This approach provided detailed identification of 165 r-protein paralogs with high coverage based on proteotypic peptides. The detected r-proteins were the products of the majority (68%) of the 242 cytosolic r-protein genes encoded by the genome. A total of 70 distinct r-proteins were identified. Based on these results and information from DNA microarray and ribosome footprint profiling studies a re-annotation of *Arabidopsis* r-proteins and genes is proposed. This compendium of the cytosolic r-protein proteome will serve as a template for future investigations on the dynamic structure and function of plant ribosomes.

Biological significance: Translation is one of the most energy demanding processes in a living cell and is therefore carefully regulated. Translational activity is tightly linked to growth control and growth regulating mechanism. Recently established translational profiling technologies, including the profiling of mRNAs associated with polysomes and the mapping of ribosome footprints on mRNAs, have revealed that the expression of gene expression is often fine-tuned by differential translation of gene transcripts. The eukaryotic ribosome, the hub of these important processes, consists of close to eighty different proteins (depending on species) and four large RNAs assembled into two highly conserved subunits.

In plants and to lesser extent in yeast, the r-proteins are encoded by more than one actively transcribed gene. As r-protein gene paralogs frequently do not encode identical proteins and are regulated by growth conditions and development, *in vivo* ribosomes are heterogeneous in their protein content. The regulatory and physiological importance of this heterogeneity is unknown. Here, an improved annotation of the more than two hundred r-protein genes of *Arabidopsis* is presented that combines proteomic and advanced mRNA expression data. This proteomic investigation and re-annotation of *Arabidopsis* ribosomes establish a base for future investigations of translational control in plants.

© 2015 Elsevier B.V. All rights reserved.

Abbreviations: 60S, 60 Svedberg; DDA, data-dependent acquisition (also called MS/MS); DIA, data-independent acquisition (also called MS^E); mRNA, messenger RNA; MS^E, data-independent acquisition MS with alternating low and high collision energy; MS medium, Murashige-Skoog medium; nt, nucleotide; PEB, polysome extraction buffer; PLGS, ProteinLynx™ Global server; RPKM, reads per kilobase per million reads; r-protein, ribosomal protein; RPG, ribosomal protein gene database; RPL, ribosome protein large subunit; RPP, ribosome acidic phosphoprotein; RPS, ribosome protein small subunit; rRNA, ribosomal RNA; TAIR, The Arabidopsis Information Resource; UPLC, ultra performance liquid chromatography.

* Corresponding author at: Umeå Plant Science Center, Department of Plant Physiology, Umeå University, 90187 Umeå, Sweden.

E-mail address: johannes.hanson@umu.se (J. Hanson).

1. Introduction

Ribosomes are large two-subunit complexes composed of ribosomal proteins (r-proteins), rRNAs and associated (non ribosomal) proteins. In eukaryotes, the cytosolic 60S large ribosomal subunit contains the 5S, 5.8S and 23S rRNAs and the 40S small ribosomal subunit harbors the 18S rRNA. Up to 48 and 33 distinct proteins bound to the rRNAs form the large and small ribosomal subunits, respectively (counting RACK1 as a r-protein). Together, the 60S and 40S ribosomal subunits form the 80S ribosome. The crystal structures of bacterial and archaeal subunits at atomic resolution identified the locations of r-proteins within the prokaryotic ribosome [1–3] and provided information on the structural aspects of protein translation by ribosomes (reviewed in [4, 5]). The recently reported atomic structures of eukaryotic cytoplasmic ribosomes exposed both similarities and differences between eukaryotic 80S and prokaryotic 70S ribosomes (reviewed in [6,7]).

The *Arabidopsis* (*Arabidopsis thaliana* Col-0) genome encodes 81 different types of r-proteins (48 large subunit (RPL) and 33 small subunit (RPS) proteins) [1,8]. Each r-protein type is encoded by multiple expressed genes (two to seven gene paralogs), forming an r-protein family. In rice (*Oryza sativa*) and other plants, and to a lesser extent in Baker's yeast (*S. cerevisiae*), ribosomal-proteins (r-proteins) are encoded by multiple paralogs [1,6]. Some r-protein gene duplication events occurred following separation of the vascular plant and the bryophyte lineages, with subsequent lineage specific loss of paralogs [1,8,9]. By contrast, mammalian r-proteins are generally encoded by single expressed genes, although mammalian genomes often harbor multiple r-protein pseudogenes [10,11]. In *Arabidopsis*, the amino acid sequence conservation of the r-protein paralogs within each r-protein family is generally very high, up to 100% [12]. Most r-proteins are conserved between *Arabidopsis* and mammals, with the exception of the acidic phosphoprotein (RPP) family P3 found only in higher plants [8,13,14]. Public expressed sequence tag (EST) and DNA microarray hybridization data showed that most r-protein family gene members are expressed, indicating that r-proteins with distinctions in their amino acid sequence encoded by paralogs could be present in the ribosome at different points of development, in different cell types, or under different growth conditions [13–15].

Mass spectrometry of cytosolic r-proteins of rat, yeast and humans shed light on both the composition and post-translational protein modifications of this complex organelle [16–20]. Gene sequence information confirmed a common evolutionary history of the eukaryotic r-protein families [6,21,22] that was recognized in *Arabidopsis* based on gene sequence data [8]. Subsequent proteomic studies of *Arabidopsis* ribosomes identified up to 78 r-protein families and individual r-protein paralogs [23] and references therein). Covalent modifications of *Arabidopsis* and maize (*Zea mays* L.) r-proteins are highly conserved and similar to those in yeast and mammals [12,24–27]. Altogether, the information indicated a heterogenic nature of plant cytosolic ribosomes. Evidence for differential functionality of r-protein paralogs was provided by phenotypic analysis of partial and complete loss-of-function mutant plants. Mutations in r-protein gene paralogs result in developmental phenotypes including embryo-lethality [28,29] and a recurring set of pleiotropic developmental defects [30] such as altered leaf shape [31,32]. Specific r-protein mutants enhance the phenotype of other developmental mutants such as *asymmetric leaves1* (*asl1*) [33,34]. The basis for the mutant phenotypes can include differential function of r-proteins, ribosome insufficiency or functions outside of the ribosome. Extra-ribosomal functions of r-proteins have been documented in plants as well as in other eukaryotes [35–37]. A recent review of the *Arabidopsis* cytosolic ribosomal proteome defined by MS analyses [23] exposed the need for a systematic re-annotation of cytosolic r-proteins and the genes that encode them. The complexity is due to the high sequence similarity of r-proteins and limited biological data. The increasing interest in translational regulation of individual and cohorts of mRNAs in plants and other organisms further necessitates updated annotation

and analysis of the *Arabidopsis* ribosomal proteome. Here we report on the identification of r-protein paralogs using mass spectrometry (MS)-based proteomic analysis of functional cytosolic ribosomes. Transgenic *Arabidopsis* plants expressing the His₆-FLAG-tagged RPL18B r-protein were grown and leaf ribosomes immunoprecipitated. Tryptic digests of ribosomes were analyzed by 1D and 2D nanoUPLC-MS^E, supplemented with MS/MS analysis. This analysis was complemented with proteomic investigation of ribosomes purified by sucrose cushion centrifugation. Database searching of the peptide fragmentation data resulted in the identification of the majority (68%) of the expected 242 encoded r-proteins (excluding thirteen pseudogenes) with proteotypic peptides. The proteomic data were combined with information from different sources, including dedicated databases of r-protein sequences and transcriptome data that included mRNAs engaged in translation, to generate a re-annotation of *Arabidopsis* cytosolic r-proteins.

2. Materials and methods

2.1. Plant material and growth conditions

Three independent batches of transgenic 35S:His₆FLAG-RPL18B plants [10] were grown on soil in long day (16 h light, 8 h dark) conditions under cool white fluorescent light (150 $\mu\text{Em}^{-2} \text{s}^{-1}$) at 22 °C, 70% relative humidity. Prior to leaf detachment, half of the plants were dark treated for four days to induce starvation. Rosette leaves of 4–5 week old plants were harvested either in a dark room (starved plants) or under normal light conditions. Leaves were incubated for 24 h in a Microclima cabinet (Snijders, Tilburg, NL) on a rotary shaker (60 rpm) under constant white fluorescent light (100 $\mu\text{Em}^{-2} \text{s}^{-1}$) in liquid 0.5× Murashige-Skoog (MS) media (Duchefa, Haarlem, NL), pH 5.8 supplemented with 0.5 g/l 1,2-(N-morpholino) ethanesulfonic acid (MES) with or without the addition of 6% sucrose (w/v). The rosette leaves from the starved plants were incubated in flasks wrapped in aluminum foil. For sedimented ribosomes, seven-day-old seedlings grown *in vitro* for 7 days on 0.2× solid Murashige and Skoog medium supplemented with sucrose 0.3% (w/v) under long day conditions (16 h light, 8 h dark) at a constant temperature of 25 °C and a light intensity of 75 $\mu\text{Em}^{-2} \text{s}^{-1}$ was used. Five biological replicates of transgenic *Ler* plants harboring either the p35S:AlcR and pAlcA:TOR RNAi constructs or the p35S:AlcR and pAlcA:GUS constructs [38]. Plant material was snap frozen in liquid nitrogen, pulverized and stored at –80 °C prior to immunopurification or purification by differential density centrifugation.

2.2. Immunopurification of ribosomes

Ribosomes were isolated by immunopurification according to Zanetti et al. [10] with minor modifications. Frozen, pulverized leaf tissue was homogenized with 2 volumes of polysome extraction buffer (PEB). Homogenates were clarified by centrifugation at 16,000 g for 10 min, and approximately 300 OD₆₀₀ units of the supernatant were incubated with approximately 200 μl ANTI-FLAG® M2 Affinity Gel beads (Sigma, St. Louis, USA) at 4 °C for 2 h with gentle rotation. The supernatant fraction was recovered, and the beads were rapidly rinsed and then washed for 5 min with 4 ml PEB per 200 μl ANTI-FLAG® M2 Affinity Gel beads, followed by a 5 min wash with 4 ml wash buffer (100 mM Tris-HCl, pH 8.5, 200 mM KCl, 25 mM EGTA, 36 mM MgCl₂, 5 mM DTT, 1 mM PMSF, 50 mg $\mu\text{g/ml}$ cycloheximide, 50 mg $\mu\text{g/ml}$ chloramphenicol and 50 U/ml RNase inhibitor (Fermentas, Leon-Rot, GmBH)) per 200 μl ANTI-FLAG® M2 Affinity Gel beads. Elution was performed by incubation of the agarose beads with 400 μl of wash buffer containing 200 ng/ μl of 3X FLAG® Peptide (Sigma) at 4 °C for 30 min. Aliquots of eluted material were stored at –80 °C prior to trypsin digestion and mass spectrometry analysis.

2.3. Purification of ribosomes using differential density centrifugation

Ribosomal subunits (40S and 60S), monoribosomes (80S) and poly-ribosomes were isolated from frozen tissue according to Bailey-Serres and Freeling [39] with minor modifications. Freshly harvested and ground seedlings were homogenized at a final concentration of 10% (w/v) in the ice-cold extraction buffer (0.2 M Tris–HCl pH 9, 400 mM KCl, 25 mM EGTA, 35 mM MgCl₂, 0.2 M sucrose) supplemented with 2% (v/v) Triton X-100, 2% (v/v) Tween 20, 2% (v/v) NP-40 and 1% (w/v) sodium deoxycholate. The extracts were incubated on ice for 10 min to solubilize membrane-bound ribosomes and centrifuged at 2880 g for 15 min at 4 °C. The supernatants were layered over a sucrose cushion (400 mM Tris–HCl pH 9, 200 mM KCl, 5 mM EGTA, 30 mM MgCl₂, 1.75 M sucrose) and ultracentrifuged at 225,000 g for 14 h. The ribosome pellet was resuspended directly in 300 µl of Laemmli buffer [40] and denatured at 100 °C for 10 min.

2.4. Trypsin digestion and peptide purification

Affinity purified proteins (100 µl) were precipitated by addition of 8 volumes of cold acetone and incubated at –20 °C for at least 1 h. After centrifugation the pellet was washed with 80% acetone and dried in air. The pellet was resuspended in 50 µl 0.1% (w/v) RapiGest SF Surfactant (Waters Corporation, Milford, USA), 5 mM DTT (Sigma) in 0.1 M ammonium bicarbonate and incubated at 60 °C for 1 h. Alkylation was performed by incubation with 15 mM iodoacetamide (IAA) (GE Healthcare, Waukesha, USA) for 30 min at room temperature in the dark. Proteolytic digestion was initiated by adding 2 µl of modified porcine trypsin (0.02 µg/µl) (Promega, Madison, USA). The sample was incubated overnight at 37 °C, after which the digestion was terminated by addition of trifluoroacetic acid (TFA) (Fluka, Buchs, GmbH) to a final concentration of 0.5% (v/v). After centrifugation at 15,000 × g for 10 min the supernatant was cleaned by binding to a SupelClean™ LC-18 1 ml SPE column (Supelco, Bellefonte, USA) equilibrated with 0.1% (v/v) TFA. The peptides were eluted from this column with 84% (v/v) acetonitrile (ACN) (HPLC Supra-gradient, Biosolve, Valkenswaard, NL) containing 0.1% (v/v) Formic acid (FA) (Merck, Darmstadt, Germany) and were then dried down during vacuum centrifugation and dissolved in 40 µl in 0.1% (v/v) FA prior to 1D LC–MS analysis or 0.1 M ammonium formate (pH 10) prior to 2D LC–MS analysis.

For sedimented ribosomes, proteins were briefly run through the stacking gel to the top of a resolving gel of a SDS-PAGE set-up in order to remove the contaminating RNA and detergents and subsequently visualized by Coomassie blue staining and the band of concentrated protein was excised from the gel and cut into five pieces. Proteins were then fixed with ethanol:acetic acid (40%:10% (v/v)) and samples were subsequently washed with acetonitrile in 25 mM of ammonium bicarbonate. Further reduction was performed by incubation in 10 mM DTT for 30 min at 56 °C and alkylation was performed by incubation with 55 mM iodoacetamide for 45 min at room temperature in the dark. Trypsin digestion was done in 120 ng of trypsin (Promega) for 5 h at 37 °C. Peptides were resuspended in 50% (v/v) ACN containing 0.5% (v/v) TFA, pelleted by vacuum centrifugation and dissolved in 0.05% (v/v) FA, 0.05% TFA, 3% (v/v) ACN.

2.5. Comparative 1D LC–MS^E and data dependent LC–MS/MS

For peptide separation of immunopurified ribosome preparations a nanoAcquity UPLC system (Waters Corporation, Manchester, UK) was used with a BEH C18 column (75 µm × 25 cm with 1.7 µm particles, Waters, UK) and a 65 min linear gradient from 3 to 40% (v/v) ACN (in 0.1% FA) at 200 nl/min; the eluting peptides were injected on-line into a Synapt Q-TOF HDMS instrument (Waters Corporation). MS analyses were performed in positive mode using ESI with a NanoLockSpray source. As lock mass, [Glu¹] fibrinopeptide B (1 pmol/µl) (Sigma) was delivered from a syringe pump (Harvard Apparatus, Holliston, USA) to

the reference sprayer of the NanoLockSpray source at a flow rate of 0.2 µl/min and the lock mass channel was sampled every 30 s. Accurate mass LC–MS data were collected with the instrument operating in either the MS/MS or MS^E mode for data-dependent acquisition (DDA) or data-independent acquisition (DIA) using low (6 eV) and elevated (ramp from 15 to 35 eV) energy spectra every 0.6 s over a 140–1900 m/z range, respectively. DDA was performed by peptide fragmentation on the three most intense multiply charged ions that were detected in the MS survey scan (0.6 s) over a 140–1900 m/z range and a dynamic exclusion window of 60 s with an automatically adjusted collision energy based on the observed precursor m/z and charge state.

Trypsin digested peptides from sedimented ribosomes were separated on a NanoLC-Ultra instrument (Eksigent, Dublin, CA). Samples were desalted in 0.1% (v/v) FA on a pre-column (stationary phase: C18 PepMap100, 100 µm × 1 cm with 5 µm particles, Dionex, Sunnyvale, CA) with a flow rate of 7.5 µl/min prior to separation on an analytical PepMapC18 column (stationary phase: C18 PepMap100, 75 µm × 150 mm with 3 µm particles, Dionex) using a 37 min linear gradient from 5 to 80% (v/v) ACN (in 0.1% (v/v) FA) and were directly injected in the Q-Exactive (Thermo Scientific, San Jose, CA) mass spectrometer using a nano-electrospray source. Ionization (1.5 kV) was performed with a liquid junction and a non-metallic needle. Ion measurement was performed with Xcalibur 2.07 (Thermo Scientific) on every ion with a mass-to-charge (m/z) ratio between 300 and 1400.

2.6. Comparative 2D LC–MS^E and data dependent LC–MS/MS

For high-resolution separation of the complex tryptic peptide samples a nanoAcquity 2D UPLC system (Waters Corporation, Manchester, UK) was used employing orthogonal reversed phase separation at high and low pH, respectively. With this 2D set up, the pool of peptides was eluted from the first dimension XBridge C₁₈ trap column (in 20 mM ammonium formate, pH 10) using a discontinuous step gradient of 12%, 15%, 18%, 20%, 25%, 35% and 65% (v/v) ACN. For the second dimension an acidic ACN gradient was applied using a BEH C₁₈ column (75 µm × 25 cm, Waters, UK) and a 65 min linear gradient from 3 to 40% (v/v) ACN (in 0.1% FA (v/v)) at 200 nl/min. The eluting peptides were on-line injected into a Synapt Q-TOF MS instrument (Waters Corporation, Manchester, UK) using a nanospray device coupled to the second dimension column output. The Synapt MS was operated in positive mode with [Glu¹] fibrinopeptide B (1 pmol/µl; Sigma) as reference (lock mass) sampled every 30 s. Data collection was performed similarly as the comparative 1D LC–MS data collection described above.

2.7. Data base construction and data analysis

LC–MS/MS and MS^E data were processed using ProteinLynx™ Global Server software (PLGS version 2.4, Waters Corporation) and the resulting list of masses containing all the fragment information was searched against the TAIR9 protein sequence database (TAIR9_pep_20090619 downloaded from The Arabidopsis Information Resource (TAIR, www.arabidopsis.org) as part of the TAIR9 Arabidopsis genome release). The annotation of r-protein genes in TAIR9 was manually updated. For MS^E, the search was performed using the following parameters: a minimum of five fragment ions per peptide and a minimum of nine fragment ions per protein, a minimum of one peptide match per protein and a maximum of one missed trypsin cleavage. Furthermore, we used (i) Carbamidomethylation (Cys) as fixed modification, (ii) Deamidation (NQ), Oxidation (M) and acetylation (N-terminus of the protein) as variable modifications, and (iii) a false discovery threshold of 4%. The false discovery rate was determined automatically in PLGS by searching the randomized TAIR9 database. Following analysis the annotation was manually curated to TAIR10. For DDA analysis the peptide tolerance was set to 30 ppm and a fragment tolerance of 0.05 Da. Carbamidomethylation (Cys) was used as fixed modification and Deamidation (NQ), Dioxidation (M), and Oxidation (M) as variable

modifications. The AutoMod option was applied as secondary search to the database search results with a maximum of one missed trypsin cleavage and non-specific secondary digest reagent were chosen. Finally, the DDA and MS^E outputs were merged in Excel. Protein identification was considered to be accurate when a protein (produce of a gene paralog) was assigned based on at least two proteotypic peptides (excluding insource decay fragments). In the DDA analyses single peptides were only kept if the ladder score was above 50. For MS^E analyses, proteins that were certified with a single peptide were accepted at a peptide-score above 5 in addition to filtering at a false discovery rate of 4%, yielding a maximum false discovery rate of 2% after filtering. Proteins were only considered to be identified if more than one peptide corresponding to the protein were identified using the above-mentioned criteria. Coverage was estimated based on the maximized coverage per sample.

Peptide and protein identity of sedimented ribosomes were searched against the TAIR8 protein sequence database with the X!tandem CYCLONE software (2011.12.01) (<http://www.thegpm.org/tandem>) based on the mass matching technique. Peptides with an e-value below 0.03 were filtered and proteins identified with a minimum of two peptides were identified. The parameters used were: possibility of one miscleavage, carbamidomethylation of the cysteines as fixed modification, deamidation of the N-terminal end of the protein, oxidation of the methionines, acetylation of the N-terminus end of the protein and phosphorylation of serines, threonines and tyrosines as variable modifications. Results were filtered, analyzed and grouped with the X!tandem pipeline v3.3.0 (<http://pappso.inra.fr/bioinfo/xtandempipeline>). Subsequently, peptides annotated as belonging to r-protein families were further checked by manual curation of TAIR10 using the PATMATCH v1.1 software.

2.8. Comparison of Ribo-seq and proteomics

Dataset 1 (sd01.xlsx) from Juntawong et al. [41] was downloaded from Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE50597). This dataset includes values for the number of ribosome footprints detected per coding region as a reads per kilobase per million (RPKM) value. The RPKM data from the Ribo-seq analysis (supporting dataset 1B) for the annotated r-proteins were selected and averaged over the two treatments used in this analysis to maximize the number of r-protein gene transcripts detected. This data was tabulated along with gene annotation and proteomic datasets. Scores were assigned according to the agreement between the number of proteomic methods that identified the r-protein and ≤5 RPKM value.

3. Results and discussion

3.1. R-protein re-annotation

We here present an updated annotation of the cytosolic r-protein-encoding genes of *Arabidopsis* (Table 1). The total number of r-protein

genes of *Arabidopsis* based on this re-annotation is 242 and includes the three RACK1 paralogs. The re-annotation combines and corrects information from several sources including, proteomic data presented here and reported previously [4,12,42,43], the RPG (Ribosomal Protein Gene database) [6], *Arabidopsis* cytoplasmic ribosomal protein gene family information from TAIR (<http://www.arabidopsis.org/browse/genefamily/athr.jsp>), as well as gene expression data from the Arabidopsis Transcriptome Genomic Express Database [1], and ribosome footprint data obtained by Juntawong et al. [41]. This study reports nine additional r-protein genes and makes one gene correction (RPS27aB from At2g47100 to At2g47110) (Table 1), based on sequence similarity to annotated r-protein genes and detected peptides in immunopurified ribosomes (Supplemental Tables S1 and S2). The newly identified r-protein genes are RPL18aD (At1g29965), RPL21G (At1g57860), RPL35aA (At1g07070), RPL41B (At1g79075), RPL41F (At3g12965), RPS2E (At1g58684), RPS2F (At1g58983), RPS25D (At4g34555) and RPS27C (At3g61111).

Our re-annotation included confirmation of pseudogenes and their exclusion from the final list. Of the 36 r-protein pseudogenes reported by Barakat et al. [8], TAIR10 annotates 13 (RPL3C, RPL4B, RPL4C, RPL5C, RPL13A, RPL21B, RPL21D, RPL21F, RPL37aA, RPS9A, RPS12B, RPS21A and RPS25C) as pseudogenes (Table 2), 16 as genes, and seven are not annotated. The genomic DNA sequences of the 13 pseudogenes (Table 2) and their corresponding expressed family members were compared to publicly available gene transcript (transcriptome) data [1]. Transcripts attributed to nine out of the 13 could be explained by the expression of functional paralogs (e.g. paralogs of RPL4B and RPS21A). For four of the 13 (RPL3C, RPL21B, RPL21D and RPS12B) there was evidence of transcript accumulation (Gbrowse, TAIR and Arabidopsis Transcriptome Genomic Express Database [1]) but translation would give rise to a non-functional protein due to mutations, confirming their annotation as pseudogenes (Table 2).

3.2. Peptides identified from immunopurified ribosomes using 1D and 2D LC-MS analysis

Arabidopsis ribosomes were obtained by translating ribosome affinity purification (TRAP) technology [10] from transgenic plants harboring a 35S::His₆FLAG-RPL18B transgene in the Col-0 accession genetic background. This method involves immunopurification of ribosomes from crude cell extracts that takes advantage of a FLAG-epitope tag on the solvent exposed amino-terminus of a 60S subunit proteins. TRAP was performed on leaf rosette tissue derived from plants with different metabolic status to maximize the potential for r-protein identification. Tissues prepared in duplicated biological experiments included excised rosette leaves incubated for 24 h in media supplemented with sucrose (6%) or without sucrose (control media) (1D and 2D proteomics) and rosette leaves of plants grown under a normal light–dark regime (control) or held in darkness for a prolonged period (sucrose-starved) (2D proteomics). The immunopurified ribosomes were digested with trypsin and the resulting peptide mixtures were analyzed both by 1D and

Table 1

Newly identified and re-annotated *Arabidopsis thaliana* cytosolic r-protein genes. Annotation is based on expression data and published annotations of r-proteins [1,6,8,12,23,41–43].

| Family | Locus name | Protein name | Other names | Sequence similarity | Number of ESTs |
|--------|------------|--------------|-------------|---------------------------------------|--|
| L18a | At1g29965 | RPL18aD | | >88% identical to RPL18aB and RPL18aC | >10 |
| L21 | At1g57860 | RPL21G | | 100% identical to RPL21E | >10 |
| L35a | At1g07070 | RPL35aA | | 100% identical to RPL35aE | >10 |
| L41 | At1g79075 | RPL41B | | 100% identical to whole family | >10 |
| | At3g12965 | RPL41F | | 100% identical to whole family | >10 |
| S2 | At1g58684 | RPS2E | | 100% identical to RPS2A and At1g58983 | >10, however 100% identical to At1g58983 |
| | At1g58983 | RPS2F | | 100% identical to RPS2A and At1g58684 | 0, however 100% identical to At1g58984 |
| S25 | At4g34555 | RPS25D | | 94% identical to RPS25B and E | >10 |
| S27 | At3g61111 | RPS27C | | ~80% identical to RPS27A, B and D | 2 |
| S27a | At2g47110 | RPS27aB | UBQ6 [45] | 97% identical to RPS27aC and UBQ5 | >10 |

Table 2*Arabidopsis thaliana* r-protein pseudogenes.

Annotation is based on TAIR, expression data and other published annotations or r-proteins [1,6,8,12,23,41–43].

| Family | Locus name | Protein name | Premature stop codon present | Longest protein encoded | Evidence for expression | ESTs | Reason for pseudogene annotation | Comments |
|--------|------------|--------------|------------------------------|-------------------------|------------------------------|---------------------------|--|---------------------------|
| L3 | At5g42445 | RPL3C | Yes | 113 | Specific expression detected | Yes, specific | Expressed but encode truncated protein | |
| L4 | At1g35200 | RPL4B | Yes | 58 | | No | No evidence for expression | |
| | At2g24730 | RPL4C | Yes | 84 | | No | No evidence for expression | |
| L5 | At5g40130 | RPL5C | Yes | 51 | | Yes, however not specific | No evidence for expression | |
| L13 | At3g48130 | RPL13A | Yes | 58 | | No | No evidence for expression | |
| L21 | At1g09486 | RPL21B | Yes | 97 | Specific expression detected | Yes, specific | Expressed but encode truncated protein | |
| | At1g31355 | RPL21D | Yes | 98 | Specific expression detected | Yes, specific | Expressed but encode truncated protein | |
| | At3g57820 | RPL21F | Yes | 47 | | Yes, however not specific | No evidence for expression | |
| L37a | At2g35180 | RPL37aA | Yes | 26 | | No | No evidence for expression | |
| S9 | At4g12160 | RPS9A | Yes | 59 | | Yes, however not specific | No evidence for expression | |
| S12 | At1g80800 | RPS12B | Yes | 33 | Specific expression detected | Yes, specific | Expressed but encode truncated protein | Previous At1g80750: RPL7A |
| S21 | At3g27450 | RPS21A | Yes | 98 | | No | No evidence for expression | |
| S25 | At3g30740 | RPS25C | Yes | 52 | | No | No evidence for expression | |

2D LC–MS procedures. In addition to conventional data-dependent acquisition (DDA) using LC–MS/MS identification, the novel acquisition procedure LC–MS^E was used. LC–MS^E entails data-independent acquisition (DIA) from alternating scans of low and high collision energy. This yields protein quantitation and identification information in a single run [13,14]. From each sample a tryptic digest was injected twice in the mass spectrometer for LC–MS^E analysis. About 50,000 search hits were obtained from a total of 24 1D MS^E runs and 37 LC–MS/MS (DDA) runs (1D and 2D). For the analysis, the peptide fragmentation information acquired by LC–MS/MS (DDA) and MS^E modes was matched to the TAIR *Arabidopsis* genome database using the ProteinLynx™ Global Server (PLGS) software package, followed by manual curation using the updated annotation as presented in Supplemental Table S1. The majority of the protein hits were observed in all samples/runs. However, there was inherent redundancy due to the matching of single peptide sequences to multiple highly-conserved proteins and the presence of derived peptides, as a result of in-source fragmentation during electrospray ionization in the MS.

Filtering of this redundancy to better assess the presence of individual r-proteins was performed at several levels. First, identical peptide sequences present in more than one sample/run were removed, resulting in a reduction in the number of peptide sequences from ca. 50,000. Secondly, for LC–MS/MS (DDA) and MS^E several search criteria were used (see **Materials and methods** section), resulting in a maximum false discovery rate of 2%, with additional filtering on peptide score for proteins identified with a single peptide. Finally, the DDA and MS^E outputs were merged in Excel. Altogether, 1541 non-redundant peptides were present after filtering (Supplemental Table S1). Filtering for loci that were matched with at least two or more proteotypic peptide sequences (with the exception for 31 paralogs encoding identical proteins) yielded 129 gene loci. These encode 129 r-proteins including three RACK1 paralogs. An average sequence coverage of 43% was reached. The 129 r-proteins represent 53% of the 242 re-annotated *Arabidopsis* r-proteins (Table 3). Although we only considered original peptides for peptide count in this manuscript, in-source decay and neutral loss fragments of ionized peptides that are detected by MS^E may provide additional confidence for the identification of a particular peptide sequence [13,14].

A 2D LC–MS analysis of the immunopurified ribosomes was also performed. Peptides were separated by Reversed Phase (RP) chromatography at high pH in the first dimension and low pH in the second dimension prior to on-line MS analysis. RP/RP analysis using nanoUPLC provides a high-resolution separation of complex peptide mixtures in two dimensions. The effectiveness of the 2D approach to detect less abundant proteins/peptides in the purified ribosomes was compared

to the 1D analysis. Three biological replicates were used to perform the immunopurification of polysomes in duplicate for four distinct sugar treatment conditions. These included the control and 6% sucrose treated rosette leaf tissue used for the 1D analysis, as well as rosette leaf tissue from control and a 4-day extended-night treated plants grown in soil. These correspond to energy-replete and energy-deficient stress conditions. Altogether a total of 24 tryptic digests were analyzed by 2D LC–MS. The mixture of peptides was eluted in seven fractions from the 1st RP column and analyzed by LC–MS. This resulted in a total of 168 2D LC–MS^E (DIA) data files and 59 additional 2D LC–MS/MS (DDA) files, which after processing and searching yielded about 91,000 search hits (Supplemental Table S2). Filtering on redundancy was performed at several levels, similar to the 1D LC–MS analysis. Removal of redundancy at the peptide level resulted in 1580 non-redundant peptide sequences. Filtering for loci that were matched with two or more original proteotypic peptide sequences (with the exception for paralogs with identical protein sequences) yielded 141 gene loci in the 2D dataset. These 141 genes encode 138 r-proteins and three RACK1 paralogs. An average sequence coverage of 45% was reached. The 141 r-proteins identified by the 2D proteomic analysis comprise 58% of the *Arabidopsis* r-proteins encoded by the 242 r-protein gene set, 12 more than the 1D analysis.

In the 2D LC–MS^E analysis we expected an increase in the number of detected peptide sequences, the number of identified peptides per protein and also the protein sequence coverage compared to the 1D analysis. Indeed, the number of detected peptide sequences increased by 3% (1541 vs. 1580). Twenty-nine r-proteins were uniquely identified in 2D LC–MS, whereas 17 were uniquely identified in the 1D LC–MS. In total 12 more r-proteins were detected using the 2D LC–MS method in comparison to the 1D LC–MS analysis. Most likely the 2D LC–MS characterized the peptide mixtures in greater detail partly because of better LC separation. The results of the 1D and 2D LC–MS^E experiments suggest that with the current experimental setup saturation level has nearly been reached in terms of r-protein identification with the LC–MS system used.

3.3. Peptide identification of sucrose cushion centrifugation purified ribosomes and 1D LC–MS analysis

To further increase the number of functional r-protein genes and proteins, we performed another 1D LC–MS/MS proteomic analysis based on an unrelated purification method using seedling tissue prepared with a different *Arabidopsis* accession. Ribosomes were obtained by ultracentrifugation of extracts prepared from 7-day-old seedlings and centrifuged through a 1.75 M sucrose cushion to obtain a pellet

highly enriched in ribosomes [39]. The *Ler* accession was used for this analysis. We confirmed that there are no sequence polymorphisms between Col-0 and *Ler* r-protein genes that would affect the encoded protein amino acid sequence of the identified peptides [44].

Five biological replicate analyses were performed on three independent *Ler* lines (see Materials and methods) using the density-based purification (“sedimented”) method. Combining the 15 samples, 17,808 MS spectra were detected, representing 2703 unique peptide sequences. The peptides were filtered (removal of low quality spectra and redundant sequences) and grouped with the X!tandem pipeline v3.3.0 [45] based on the sequence homology between the identified proteins. This identified 423 different proteins belonging to 318 different protein groups of which 72 corresponded to cytoplasmic r-protein families, which was further analyzed. The 730 peptide sequences corresponding to the 72 r-protein families were manually reannotated using PATMATCH v1.1 based on the TAIR10 genome. Of the 730 r-protein peptides, 313 were proteotypic for 106 proteins. Of these, 83 proteins were identified with at least two different proteotypic peptides (with an average of three proteotypic peptides), yielding a protein-coverage range from 14% and 89% and average of 45% (Table 4). This sedimented ribosome approach with seedlings identified 7 r-proteins not identified in the 1D or 2D proteomic analysis of rosette leaf tissue. Among the 730 peptides corresponding to r-proteins identified in the sedimented ribosomes, 146 were not detected either in the 1D or in the 2D analysis of TRAP ribosomes (Fig. 1).

3.4. Comparison with other ribosome proteome studies

A comparison between the results of this study and previous ribosome proteomic studies is summarized in Table 3. Here, we identified r-proteins with the average protein sequence coverage of 43% and 45% for the immunopurified ribosomes analyzed by TRAP1D LC-MS and TRAP 2D LC-MS, respectively and 45% for the sedimented ribosomes. R-proteins were identified with lower sequence coverages in previous studies [4,12,43]. Although some samples were run in data-dependent LC-MS/MS mode, most of the matched peptide sequences were identified by data-independent LC-MS^E. A conventional LC-MS/MS analysis starts with an MS scan followed by the selection of (in our setup) three precursor ions that are sequentially fragmented in the collision cell. Hence, a LC-MS/MS experiment is typically a serial process, which results in a competition to acquire as many MS/MS spectra for as many precursor ions as possible in a given period of time. The allotted MS/MS acquisition time must be limited in order to maximize the use of the instrument. In our study the MS/MS acquisition time was a limiting factor. In the latest generation of MS systems a higher MS/MS frequency enables increased coverage of peptide capture. In contrast, in the MS^E mode no precursor ions are selected and also single charged in-source decay and neutral loss fragments of ionized peptides can be detected, which results in the identification of more peptides per protein [13,14]. However, in this manuscript only original peptides were considered.

Previous ribosome proteomic studies identified 62, 75 and 87 gene-specific r-proteins [4,12,43]. In the study of Carroll et al. [12] peptides were assigned to 207 r-protein paralogs, but only 87 of these were identified by proteotypic peptides [12]. In this study we used two methods of ribosome purification, two different tissues (rosette, seedling), multiple treatments, and two genetic accessions to expand the gene and protein coverage. A total of 129 (TRAP 1D), 141 (TRAP 2D) and 83 (Sedimented) r-protein paralogs were identified in the three method-based assays (Table 3). In total, 31 of the 242 encoded r-protein genes encode proteins with 100% amino acid sequence identity to at least one other family member. This reduces the number of r-protein genes that can be discriminated at the protein level to 207. This includes seven r-protein families that encode completely identical proteins (*RPL11*, *RPL23*, *RPL38*, *RPL41*, *RPS18*, *RPS29* and *RPS30*) (Table 3). Thus, by using a combination of LC-MS^E and LC-MS/MS

methods and expanded biological sampling, this study recognized 143 of the 217 distinguishable r-protein sequences, (Table 3). Importantly, in previous studies ribosomes were isolated using sucrose density-gradient ultracentrifugation [4,12,43], whereas in the present study ribosomes were isolated by immunopurification (TRAP) [10] or a density-based purification procedure (Sedimented). Although the TRAP method, which relies on an epitope-tagged RPL18, might enrich for a ribosomal complex with a certain composition, the comparison of the TRAP and density-purified ribosomal proteome did not indicate a method-based bias. Another important difference is that previous studies separated r-proteins by SDS-PAGE or 2D gel electrophoresis and excised gel slices or protein spots for trypsin digestion [4,12,43] (or chymotrypsin and pepsin [12]) prior to MS analysis. This approach, which is biased towards proteins that can be detected by staining, differs from the direct trypsin digestion of immunopurified or sedimented ribosomes used in this study. Both the TRAP of ribosomal complexes as well as in-solution trypsin digestion of the protein mixtures without pre-fractionation are more convenient and rapid.

For the LC-MS analysis, tryptic peptides were used, similar to Gialvalisco et al., [43] and Chang et al. [4]. As some r-proteins have a relatively high arginine and lysine content (e.g., *RPL39* paralogs have 15–18 K/R of 44–51 aa), their tryptic digestion will result in very short peptides, which are undetectable by LC-MS. Carroll et al. [12] performed additionally chymotrypsin and pepsin digestion, which yielded peptides of r-proteins not identified in this study, including products of *RPL36A*, *RPL39A* and *RPL40* paralogs. Despite the intensive analysis of the *Arabidopsis* ribosomal proteome, the products of several annotated r-protein genes have escaped detection. These include *RPL29* paralogs, *RPL39B*, *RPL39C* and *RPL41* paralogs and some of the *RPP2* and *RPS15* paralogs. These r-proteins are small in size (i.e., *RPL41s* are only 25 aa, 3.4 kDa) and therefore difficult to detect. The possibility that they are present in sub-stoichiometric amounts in ribosomes in comparison with the identified r-proteins cannot be ruled out. Dedicated studies are needed to identify these remaining putative r-proteins.

In addition to the r-proteins the purification methods yield some non-r-proteins as previously reported [12,42,43]. Several non-r proteins identified in the TRAP ribosomes have functions in translation (i.e., Elongation Factor 1A/At1g07920 and isoleucyl-tRNA synthetase/At4g10320 or as chaperones (i.e., HSP70/At3g12580)). Other non-r-proteins may be co-purified with the epitope-tagged ribosomes due to their involvement in ribosome biogenesis, export, or turnover, or because they are encoded by abundant mRNAs undergoing translation. In a previous dataset we found a lower percentage of peptide coverage for non r-proteins and non-stoichiometric abundance in our ribosome preparations [42].

3.5. Translation of r-protein transcripts correlates with proteomic data

A method that quantitatively measures the position and number of ribosomes on gene transcripts termed ribosome profiling was recently applied to *Arabidopsis* [41,46]. The method involves digestion of ribosome-mRNA complexes with a nuclease and high-throughput sequencing of the ~30 nt fragment of mRNA that is protected by individual ribosomes. These are mapped back to the genome and counted to provide information on the number of ribosomes engaged in translation of a transcript. This so called Ribo-seq data provided the opportunity to evaluate ribosome association with r-protein mRNAs *in vivo* to the ribosome proteome obtained by mass-spectrometry (Table 3).

To perform the analysis, the reads per kilobase per million reads (RPKM) values from the seedling Ribo-seq data were obtained for all annotated r-protein genes and compared to the detection data of the corresponding r-protein in the proteomics datasets. The mapped ribosome footprint reads considered uniquely matched a gene with no more than two mismatches per ~30 nt fragment. Ribosome footprints that matched more than one r-protein family member were not tallied.

Table 3
Identification of *Arabidopsis thaliana* cytosolic r-protein genes and comparison to other published data.
Annotation is based on TAIR10, expression data and other published annotations or r-proteins. Proteomics columns: NSPA = No specific peptide assigned, Y = Detected, N = Not detected and N.D. = Not determined. Ribo-Seq RPKM values indicated, * = judged as expressed from reads mapped to genome sequence only. Comparison columns: the scores show the Identification of r-proteins in proteomics experimentation of this study compared to Ribo-seq. Both; indicates identification using both proteomics and Ribo-Seq, RNA only, indicates detection only in Ribo-Seq, Prot. only; indicates identification using proteomics only and None; indicate not detected in this study and using Ribo-Seq.

| Annotation | | | | Public proteomic data | | | Proteomic data | | | | | | Ribo-Seq | Agreement |
|------------|--------------|---------------------|---------------------------------------|-----------------------|-------------------------|----------------------|--------------------------------------|----------|--------------------------------------|----------|-----------------------|----------|------------------------|---------------------|
| Family | Protein Name | Locus Name (TAIR10) | Protein Similarity and Other Comments | Chang et al., 2005 | Giavalisco et al., 2005 | Carroll et al., 2008 | 1D TRAP LC-MS | | 2D TRAP LC-MS | | Sedimented | | Juntawong et al., 2014 | Ribo-Seq Proteomics |
| | | | | [4] | [43] | [12] | Max Sequence Coverage per Sample (%) | Detected | Max Sequence Coverage per Sample (%) | Detected | Sequence Coverage (%) | Detected | [41] | |
| RPL3 | RPL3A | At1g43170 | | Y | Y | Y | 50 | Y | 48 | Y | 45 | Y | 260 | Both |
| | RPL3B | At1g61580 | | NSPA | N | N | 24 | Y | 29 | Y | | N | 15 | Both |
| RPL4 | RPL4A | At3g09630 | | Y | Y | Y | 41 | Y | 34 | Y | 50 | Y | 155 | Both |
| | RPL4D | At5g02870 | | Y | Y | Y | 37 | Y | 30 | Y | 51 | Y | 179 | Both |
| RPL5 | RPL5A | At3g25520 | | Y | Y | Y | 37 | Y | 44 | Y | 54 | Y | 127 | Both |
| | RPL5B | At5g39740 | | NSPA | Y | Y | 36 | Y | 43 | Y | 54 | Y | 83 | Both |
| RPL6 | RPL6A | At1g18540 | | NSPA | Y | Y | 52 | Y | 51 | Y | 36 | Y | 163 | Both |
| | RPL6B | At1g74060 | | NSPA | Y | N | | N | | N | | N | 47 | RNA only |
| | RPL6C | At1g74050 | | NSPA | N | Y | | N | | N | | N | 47 | RNA only |
| RPL7a | RPL7aA | At2g47610 | | Y | N | Y | 26 | Y | 26 | Y | 53 | m | 223 | Both |
| | RPL7aB | At3g62870 | | Y | Y | Y | 26 | Y | 27 | Y | 46 | Y | 172 | Both |
| RPL7 | RPL7A | At1g80750 | | NSPA | N | N | | N | | N | | N | 23 | RNA only |
| | RPL7B | At2g01250 | | Y | N | Y | 49 | Y | 55 | Y | 50 | Y | 191 | Both |
| | RPL7C | At2g44120 | | Y | Y | Y | 52 | Y | 60 | Y | 49 | Y | 102 | Both |
| | RPL7D | At3g13580 | | Y | Y | Y | 42 | Y | 46 | Y | 37 | Y | 30 | Both |
| RPL8 | RPL8A | At2g18020 | | Y | Y | Y | 48 | Y | 47 | Y | 45 | Y | 270 | Both |
| | RPL8B | At3g51190 | | Y | N | N | | N | | N | | N | 1 | None |
| | RPL8C | At4g36130 | | Y | N | N | 47 | Y | 51 | Y | 41 | Y | 71 | Both |
| RPL9 | RPL9B | At1g33120 | 100% identical to RPL9C | NSPA | Y | N | 71 | Y | | N | 62 | Y | 9 | Both |
| | RPL9C | At1g33140 | 100% identical to RPL9B | NSPA | N | N | 71 | Y | | N | 62 | Y | 11 | Both |
| | RPL9D | At4g10450 | | Y | N | N | 48 | Y | 44 | Y | | N | 43 | Both |
| RPL10a | RPL10aA | At1g08360 | | Y | Y | Y | 37 | Y | 44 | Y | 47 | Y | 117 | Both |
| | RPL10aB | At2g27530 | | Y | N | Y | 39 | Y | 40 | Y | 43 | Y | 78 | Both |
| | RPL10aC | At5g22440 | | Y | Y | N | 34 | Y | 36 | Y | 33 | Y | 43 | Both |
| RPL10 | RPL10A | At1g14320 | | Y | Y | Y | 36 | Y | 44 | Y | 41 | Y | 213 | Both |
| | RPL10B | At1g26910 | | NSPA | N | N | 29 | Y | 36 | Y | | N | 8 | Both |
| | RPL10C | At1g66580 | | Y | N | Y | 36 | Y | 42 | Y | 35 | Y | 56 | Both |
| RPL11 | RPL11A | At2g42740 | Whole family 100% identical | NSPA | N | Y | 40 | Y | 51 | Y | | N | 10 | Both |
| | RPL11B | At3g58700 | Whole family 100% identical | NSPA | Y | Y | 40 | Y | 51 | Y | | N | 29 | Both |
| | RPL11C | At4g18730 | Whole family 100% identical | NSPA | N | Y | 40 | Y | 51 | Y | | N | 95 | Both |
| | RPL11D | At5g45775 | Whole family 100% identical | NSPA | N | Y | 42 | Y | 51 | Y | | N | 62 | Both |
| RPL12 | RPL12A | At2g37190 | | NSPA | Y | N | 61 | Y | 66 | Y | | N | 104 | Both |
| | RPL12B | At3g53430 | | NSPA | N | N | 61 | Y | 66 | Y | | N | 203 | Both |
| | RPL12C | At5g60670 | | NSPA | N | N | 65 | Y | 60 | Y | | N | 35 | Both |
| RPL13a | RPL13aA | At3g07110 | | NSPA | Y | Y | 34 | Y | 47 | Y | 41 | Y | 97 | Both |
| | RPL13aB | At3g24830 | | Y | N | Y | 34 | Y | 41 | Y | 42 | Y | 145 | Both |
| | RPL13aC | At4g13170 | | NSPA | N | N | | N | 37 | Y | | N | 46 | Both |
| | RPL13aD | At5g48760 | | NSPA | Y | Y | 32 | Y | 35 | Y | | N | 41 | Both |
| RPL13 | RPL13B | At3g49010 | | Y | N | Y | 50 | Y | 51 | Y | 46 | Y | 176 | Both |
| | RPL13C | At3g48960 | | NSPA | N | N | | N | 22 | Y | | N | 1 | Prot. only |
| | RPL13D | At5g23900 | | NSPA | Y | Y | 35 | Y | 49 | Y | 43 | Y | 62 | Both |
| RPL14 | RPL14A | At2g20450 | | NSPA | N | Y | 39 | Y | 42 | Y | 40 | Y | 44 | Both |
| | RPL14B | At4g27090 | | Y | Y | Y | 46 | Y | 54 | Y | 46 | Y | 229 | Both |
| RPL15 | RPL15A | At4g16720 | | Y | Y | N | | N | | N | | N | 277 | RNA only |
| | RPL15B | At4g17390 | | NSPA | N | N | | N | | N | | N | 190 | RNA only |
| RPL17 | RPL17A | At1g27400 | | NSPA | N | Y | 43 | Y | | N | | N | 96 | Both |
| | RPL17B | At1g67430 | | Y | Y | Y | 37 | Y | 33 | Y | 51 | Y | 292 | Both |
| RPL18a | RPL18aA | At1g29970 | | NSPA | N | N | | N | | N | | N | 9 | RNA only |
| | RPL18aB | At2g34480 | | Y | Y | Y | 38 | Y | 44 | Y | 46 | Y | 236 | Both |
| | RPL18aC | At3g14600 | | Y | Y | N | 40 | Y | 39 | Y | 24 | Y | 62 | Both |

Table 3 (continued)

| Annotation | | | | Public proteomic data | | | Proteomic data | | | | | | Ribo-Seq | Agreement |
|------------|--------------|---------------------|---------------------------------------|-----------------------|-------------------------|----------------------|--------------------------------------|----------|--------------------------------------|----------|-----------------------|----------|------------------------|---------------------|
| Family | Protein Name | Locus Name (TAIR10) | Protein Similarity and Other Comments | Chang et al., 2005 | Giavalisco et al., 2005 | Carroll et al., 2008 | 1D TRAP LC-MS | | 2D TRAP LC-MS | | Sedimented | | Juntawong et al., 2014 | Ribo-Seq Proteomics |
| | | | | [4] | [43] | [12] | Max Sequence Coverage per Sample (%) | Detected | Max Sequence Coverage per Sample (%) | Detected | Sequence Coverage (%) | Detected | [41] | |
| RPL18 | RPL18aD | At1g29965 | | N | N | N | | N | | N | | N | 2 | None |
| | RPL18A | At2g47570 | | NSPA | N | N | | N | | N | | N | 1 | None |
| | RPL18B | At3g05590 | | NSPA | N | Y | 44 | Y | 42 | Y | 38 | Y | 1326 | Both |
| RPL19 | RPL18C | At5g27850 | | NSPA | Y | Y | 40 | Y | 42 | Y | 46 | Y | 148 | Both |
| | RPL19A | At1g02780 | | NSPA | N | N | 32 | Y | 27 | Y | | N | 469 | Both |
| | RPL19B | At3g16780 | | NSPA | N | N | 19 | Y | 20 | Y | | N | 49 | Both |
| RPL21 | RPL19C | At4g02230 | | Y | N | N | 31 | Y | 28 | Y | 15 | Y | 50 | Both |
| | RPL21A | At1g09590 | 100% identical to RPL21C | Y | Y | Y | | N | | N | | N | 32 | RNA only |
| | RPL21C | At1g09690 | 100% identical to RPL21A | Y | N | Y | | N | | N | | N | 47 | RNA only |
| RPL22 | RPL21E | At1g57660 | 100% identical to RPL21G | Y | N | Y | | N | | N | | N | 8 | RNA only |
| | RPL21G | At1g57860 | 100% identical to RPL21E | N | N | N | | N | | N | | N | 6 | RNA only |
| | RPL22A | At1g02830 | | NSPA | N | N | | N | | N | | N | 1 | None |
| RPL23a | RPL22B | At3g05560 | | Y | Y | Y | 54 | Y | 49 | Y | | N | 134 | Both |
| | RPL22C | At5g27770 | | NSPA | N | Y | 50 | Y | 49 | Y | | N | 82 | Both |
| | RPL23aA | At2g39460 | | NSPA | Y | Y | | N | | N | | N | 272 | RNA only |
| RPL23 | RPL23aB | At3g55280 | | Y | N | Y | | N | | N | | N | 65 | RNA only |
| | RPL23A | At1g04480 | Whole family 100% identical | Y | N | Y | 61 | Y | | N | 55 | Y | 73 | Both |
| | RPL23B | At2g33370 | Whole family 100% identical | Y | N | Y | 61 | Y | | N | 55 | Y | 65 | Both |
| RPL24 | RPL23C | At3g04400 | Whole family 100% identical | Y | N | Y | 61 | Y | | N | 55 | Y | 445 | Both |
| | RPL24A | At2g36620 | | NSPA | N | N | | N | | N | | N | 98 | RNA only |
| | RPL24B | At3g53020 | | NSPA | N | Y | | N | | N | | N | 178 | RNA only |
| RPL26 | RPL24C | At2g44860 | | NSPA | N | N | | N | | N | | N | 25 | RNA only |
| | RPL26A | At3g49910 | | Y | Y | Y | 23 | Y | 39 | Y | 41 | Y | 293 | Both |
| | RPL26B | At5g67510 | | Y | N | Y | | N | | N | 17 | Y | 39 | Both |
| RPL27a | RPL27aA | At1g12960 | | NSPA | N | N | | N | | N | | N | 1 | None |
| | RPL27aB | At1g23290 | | NSPA | Y | N | | N | | N | | N | 281 | RNA only |
| | RPL27aC | At1g70600 | | Y | N | N | | N | | N | | N | 130 | RNA only |
| RPL27 | RPL27A | At2g32220 | | NSPA | N | N | 27 | Y | 31 | Y | | N | 10 | Both |
| | RPL27B | At3g22230 | | Y | N | Y | 35 | Y | 34 | Y | 33 | Y | 132 | Both |
| | RPL27C | At4g15000 | | Y | N | Y | 42 | Y | 50 | Y | 37 | Y | 274 | Both |
| RPL28 | RPL28A | At2g19730 | | NSPA | Y | Y | 48 | Y | 49 | Y | 43 | Y | 204 | Both |
| | RPL28C | At4g29410 | | NSPA | N | Y | 31 | Y | 43 | Y | 41 | Y | 70 | Both |
| RPL29 | RPL29A | At3g06700 | | NSPA | N | N | | N | | N | | N | 174 | RNA only |
| | RPL29B | At3g06680 | | NSPA | N | N | | N | | N | | N | 45 | RNA only |
| | RPL30A | At1g36240 | | NSPA | N | N | | N | | N | | N | 1 | None |
| RPL30 | RPL30B | At1g77940 | | NSPA | Y | N | 53 | Y | 53 | Y | | N | 162 | Both |
| | RPL30C | At3g18740 | | Y | N | N | 48 | Y | | N | | N | 120 | Both |
| | RPL31A | At2g19740 | | Y | N | N | | N | | N | | N | 44 | RNA only |
| RPL31 | RPL31B | At4g26230 | | NSPA | N | N | | N | 42 | Y | | N | 63 | Both |
| | RPL31C | At5g56710 | | Y | Y | Y | 38 | Y | | N | | N | 188 | Both |
| RPL32 | RPL32A | At4g18100 | | Y | N | Y | 50 | Y | 42 | Y | 53 | Y | 232 | Both |
| | RPL32B | At5g46430 | | NSPA | N | Y | 47 | Y | 47 | Y | 53 | Y | 115 | Both |
| RPL34 | RPL34A | At1g26880 | | NSPA | N | N | | N | 32 | Y | | N | 109 | Both |
| | RPL34B | At1g69620 | | NSPA | N | Y | | N | | N | 26 | Y | 150 | Both |
| | RPL34C | At3g28900 | | Y | N | N | 28 | Y | | N | | N | 50 | Both |
| RPL35a | RPL35aA | At1g07070 | | NSPA | N | Y | | N | 41 | Y | | N | 13 | Both |
| | RPL35aB | At1g41880 | | NSPA | N | Y | | N | 50 | Y | | N | 26 | Both |
| | RPL35aC | At1g74270 | | NSPA | N | Y | | N | | N | | N | 60 | RNA only |
| RPL35 | RPL35aD | At3g55750 | | NSPA | N | Y | | N | | N | | N | 21 | RNA only |
| | RPL35aE | At1g06980 | | NSPA | N | N | | N | | N | | N | 2 | None |
| | RPL35A | At3g09500 | | NSPA | N | Y | 34 | Y | 46 | Y | | N | 125 | Both |
| | RPL35B | At2g39390 | | NSPA | Y | Y | 32 | Y | 39 | Y | | N | 50 | Both |

(continued on next page)

Table 3 (continued)

| Annotation | | | | Public proteomic data | | | Proteomic data | | | | | | Ribo-Seq | Agreement |
|------------|--------------|---------------------|---------------------------------------|-----------------------|-------------------------|----------------------|--------------------------------------|----------|--------------------------------------|----------|-----------------------|----------|------------------------|---------------------|
| Family | Protein Name | Locus Name (TAIR10) | Protein Similarity and Other Comments | Chang et al., 2005 | Giavalisco et al., 2005 | Carroll et al., 2008 | 1D TRAP LC-MS | | 2D TRAP LC-MS | | Sedimented | | Juntawong et al., 2014 | Ribo-Seq Proteomics |
| | | | | [4] | [43] | [12] | Max Sequence Coverage per Sample (%) | Detected | Max Sequence Coverage per Sample (%) | Detected | Sequence Coverage (%) | Detected | [41] | |
| RPL36a | RPL35C | At3g55170 | | NSPA | N | Y | | N | 32 | Y | | N | 10 | Both |
| | RPL35D | At5g02610 | | Y | N | N | 31 | Y | | N | | N | 28 | Both |
| | RPL36aA | At3g23390 | | N | N | Y | | N | | N | | N | 124 | RNA only |
| | RPL36aB | At4g14320 | | N | N | Y | | N | | N | | N | 87 | RNA only |
| RPL36 | RPL36A | At2g37600 | | NSPA | N | N | | N | | N | | N | 27 | RNA only |
| | RPL36B | At3g53740 | | NSPA | Y | Y | | N | 46 | Y | | N | 99 | Both |
| RPL37a | RPL36C | At5g02450 | | NSPA | Y | Y | 38 | Y | 45 | Y | | N | 150 | Both |
| | RPL37aB | At3g10950 | | N | N | N | | N | | N | | N | 1 | None |
| | RPL37aC | At3g60245 | | N | N | Y | 43 | Y | 38 | Y | | N | 267 | Both |
| RPL37 | RPL37A | At1g15250 | | NSPA | N | N | | N | | N | | N | 17 | RNA only |
| | RPL37B | At1g52300 | | NSPA | Y | Y | | N | | N | | N | 112 | RNA only |
| RPL38 | RPL37C | At3g16080 | | NSPA | N | Y | | N | | N | | N | 51 | RNA only |
| | RPL38A | At2g43460 | Whole family 100% identical | Y | Y | Y | 54 | Y | | N | | N | 135 | Both |
| | RPL38B | At3g59540 | Whole family 100% identical | Y | N | Y | 54 | Y | | N | | N | 124 | Both |
| RPL39 | RPL39A | At2g25210 | | N | N | Y | | N | | N | | N | 69 | RNA only |
| | RPL39B | At3g02190 | | N | N | N | | N | | N | | N | 24 | RNA only |
| | RPL39C | At4g31985 | | N | N | N | | N | | N | | N | 146 | RNA only |
| RPL40 | RPL40A | At2g36170 | | N | N | Y | | N | | N | | N | 48 | RNA only |
| | RPL40B | At3g52590 | | N | N | Y | | N | | N | | N | 245 | RNA only |
| RPL41 | RPL41A | At1g56045 | Whole family 100% identical | N | N | N | | N | | N | | N | 56 | RNA only |
| | RPL41B | At1g79075 | Whole family 100% identical | N | N | N | | N | | N | | N | expressed* | RNA only |
| | RPL41C | At2g40205 | Whole family 100% identical | N | N | N | | N | | N | | N | 91 | RNA only |
| | RPL41D | At3g08520 | Whole family 100% identical | N | N | N | | N | | N | | N | 21 | RNA only |
| | RPL41E | At3g11120 | Whole family 100% identical | N | N | N | | N | | N | | N | 63 | RNA only |
| | RPL41F | At3g12965 | Whole family 100% identical | N | N | N | | N | | N | | N | expressed* | RNA only |
| | RPL41G | At3g56020 | Whole family 100% identical | N | N | N | | N | | N | | N | 54 | RNA only |
| RPP0 | RPP0A | At2g40010 | | NSPA | N | N | 51 | Y | 32 | Y | | N | 5 | Prot. only |
| | RPP0B | At3g09200 | | Y | Y | Y | 50 | Y | 48 | Y | 57 | Y | 252 | Both |
| | RPP0C | At3g11250 | | Y | Y | N | 54 | Y | 52 | Y | | N | 24 | Both |
| RPP1 | RPP1A | At1g01100 | | Y | N | N | 60 | Y | 52 | Y | 89 | Y | 163 | Both |
| | RPP1B | At4g00810 | | NSPA | N | Y | | N | | N | 51 | Y | 60 | Both |
| | RPP1C | At5g47700 | | NSPA | N | N | 36 | Y | 51 | Y | 75 | Y | 79 | Both |
| RPP2 | RPP1D | At5g24510 | | N | N | N | | N | | N | | N | 1 | None |
| | RPP2A | At2g27720 | | Y | Y | Y | 85 | Y | 66 | Y | 76 | Y | 93 | Both |
| | RPP2B | At2g27710 | | Y | Y | Y | 71 | Y | 71 | Y | 62 | Y | 155 | Both |
| RPP3 | RPP2C | At3g28500 | | Y | N | Y | | N | | N | | N | 2 | None |
| | RPP2D | At3g44590 | | NSPA | N | N | | N | | N | | N | 32 | RNA only |
| | RPP2E | At5g40040 | | NSPA | N | N | | N | | N | | N | 1 | None |
| RPSa | RPP3A | At4g25890 | | Y | N | N | | N | | N | 31 | Y | 64 | Both |
| | RPP3B | At5g57290 | | NSPA | N | Y | 72 | Y | 36 | Y | 47 | Y | 188 | Both |
| RPS2 | RPSaA | At1g72370 | | Y | Y | Y | 46 | Y | 45 | Y | 50 | Y | 201 | Both |
| | RPSaB | At3g04770 | | Y | N | N | 28 | Y | 41 | Y | | N | 6 | Both |
| | RPS2A | At1g58380 | 100% identical to RPS2E and RPS2F | NSPA | N | N | | N | 39 | Y | | N | 11 | Both |
| | RPS2B | At1g59359 | | NSPA | N | N | | N | | N | | N | 1 | None |
| | RPS2C | At2g41840 | | Y | Y | Y | 38 | Y | 49 | Y | 41 | Y | 175 | Both |
| | RPS2D | At3g57490 | | NSPA | N | Y | 24 | Y | 33 | Y | | N | 12 | Both |

Table 3 (continued)

| Annotation | | | | Public proteomic data | | | Proteomic data | | | | | | Ribo-Seq | Agreement |
|------------|--------------|---------------------|---------------------------------------|-----------------------|-------------------------|----------------------|--------------------------------------|----------|--------------------------------------|----------|-----------------------|----------|------------------------|---------------------|
| Family | Protein Name | Locus Name (TAIR10) | Protein Similarity and Other Comments | Chang et al., 2005 | Giavalisco et al., 2005 | Carroll et al., 2008 | 1D TRAP LC-MS | | 2D TRAP LC-MS | | Sedimented | | Juntawong et al., 2014 | Ribo-Seq Proteomics |
| | | | | [4] | [43] | [12] | Max Sequence Coverage per Sample (%) | Detected | Max Sequence Coverage per Sample (%) | Detected | Sequence Coverage (%) | Detected | [41] | |
| | RPS2E | At1g58684 | 100% identical to RPS2A and RPS2F | NSPA | N | N | | N | 39 | Y | | N | 1 | Prot. only |
| | RPS2F | At1g58983 | 100% identical to RPS2A and RPS2E | NSPA | N | N | | N | 39 | Y | | N | 1 | Prot. only |
| RPS3a | RPS3aA | At3g04840 | | Y | Y | Y | 51 | Y | 45 | Y | 42 | Y | 192 | Both |
| | RPS3aB | At4g34670 | | Y | Y | Y | 52 | Y | 42 | Y | 48 | Y | 149 | Both |
| RPS3 | RPS3A | At2g31610 | | NSPA | Y | Y | 56 | Y | 49 | Y | | N | 76 | Both |
| | RPS3B | At3g53870 | | Y | Y | Y | 61 | Y | 50 | Y | 42 | Y | 115 | Both |
| | RPS3C | At5g35530 | | NSPA | Y | Y | 54 | Y | 60 | Y | | N | 97 | Both |
| RPS4 | RPS4A | At2g17360 | | NSPA | N | Y | | N | | N | | N | 61 | RNA only |
| | RPS4B | At5g07090 | | NSPA | Y | Y | | N | | N | | N | 90 | RNA only |
| | RPS4D | At5g58420 | | NSPA | Y | Y | | N | 52 | Y | | N | 87 | Both |
| RPS5 | RPS5A | At2g37270 | | Y | Y | N | 48 | Y | 52 | Y | | N | 64 | Both |
| | RPS5B | At3g11940 | | NSPA | Y | N | 45 | Y | 53 | Y | | N | 216 | Both |
| RPS6 | RPS6A | At4g31700 | | Y | N | Y | 42 | Y | 42 | Y | 46 | Y | 229 | Both |
| | RPS6B | At5g10360 | | Y | N | Y | 41 | Y | 44 | Y | 41 | Y | 155 | Both |
| RPS7 | RPS7A | At1g48830 | | Y | Y | Y | 51 | Y | 58 | Y | 28 | Y | 104 | Both |
| | RPS7B | At3g02560 | | NSPA | Y | N | 45 | Y | 40 | Y | 25 | Y | 81 | Both |
| | RPS7C | At5g16130 | | Y | Y | Y | 48 | Y | 44 | Y | 44 | Y | 117 | Both |
| RPS8 | RPS8A | At5g20290 | | Y | Y | Y | 51 | Y | 52 | Y | 49 | Y | 418 | Both |
| | RPS8B | At5g59240 | | NSPA | Y | N | 29 | Y | 32 | Y | | N | 2 | Prot. only |
| RPS9 | RPS9B | At5g15200 | | Y | Y | Y | 40 | Y | 56 | Y | 39 | Y | 298 | Both |
| | RPS9C | At5g39850 | | NSPA | N | N | 27 | Y | 41 | Y | | N | 14 | Both |
| RPS10 | RPS10A | At4g25740 | | NSPA | Y | N | | N | 41 | Y | 37 | Y | 68 | Both |
| | RPS10B | At5g41520 | | Y | N | Y | 18 | Y | 33 | Y | 35 | Y | 145 | Both |
| | RPS10C | At5g52650 | | Y | N | Y | 36 | Y | 39 | Y | 36 | Y | 128 | Both |
| RPS11 | RPS11A | At3g48930 | | NSPA | Y | Y | 44 | Y | 55 | Y | 44 | Y | 107 | Both |
| | RPS11B | At4g30800 | | NSPA | N | N | 33 | Y | 37 | Y | | N | 5 | Both |
| | RPS11C | At5g23740 | | NSPA | N | Y | 44 | Y | 55 | Y | 36 | Y | 72 | Both |
| RPS12 | RPS12A | At1g15930 | | Y | Y | N | 73 | Y | 67 | Y | 48 | Y | 119 | Both |
| | RPS12C | At2g32060 | | Y | Y | Y | 47 | Y | 67 | Y | | N | 85 | Both |
| RPS13 | RPS13A | At3g60770 | | NSPA | Y | Y | | N | 52 | Y | | N | 137 | Both |
| | RPS13B | At4g00100 | | NSPA | N | Y | | N | | N | | N | 172 | RNA only |
| RPS14 | RPS14A | At2g36160 | | Y | N | Y | | N | 65 | Y | | N | 116 | Both |
| | RPS14B | At3g11510 | | Y | Y | Y | | N | 65 | Y | | N | 152 | Both |
| | RPS14C | At3g52580 | | NSPA | N | N | 58 | Y | 59 | Y | | N | 34 | Both |
| RPS15a | RPS15aA | At1g07770 | 100% identical to RPS15aF | Y | Y | N | 41 | Y | 40 | Y | | N | 93 | Both |
| | RPS15aB | At2g19720 | | NSPA | N | N | | N | | N | | N | 9 | RNA only |
| | RPS15aC | At2g39590 | | NSPA | N | N | | N | | N | | N | 1 | None |
| | RPS15aD | At3g46040 | | Y | N | Y | 34 | Y | 45 | Y | | N | 57 | Both |
| | RPS15aE | At4g29430 | possibly mitochondrial | Y | N | N | | N | | N | | N | 19 | RNA only |
| | RPS15aF | At5g59850 | 100% identical to RPS15aA | Y | N | N | 41 | Y | 40 | Y | | N | 78 | Both |
| RPS15 | RPS15A | At1g04270 | | Y | Y | N | | N | | N | 41 | Y | 185 | Both |
| | RPS15B | At5g09490 | | NSPA | N | N | | N | | N | | N | 1 | None |
| | RPS15C | At5g09500 | | NSPA | N | N | | N | | N | | N | 1 | None |
| | RPS15D | At5g09510 | | Y | N | N | | N | | N | 41 | Y | 43 | Both |
| | RPS15E | At5g43640 | | NSPA | N | N | | N | | N | | N | 1 | None |
| | RPS15F | At5g63070 | | NSPA | N | N | | N | | N | | N | 1 | None |
| RPS16 | RPS16A | At2g09990 | | Y | Y | Y | | N | | N | | N | 14 | RNA only |
| | RPS16B | At3g04230 | | Y | N | N | 37 | Y | 37 | Y | | N | 12 | Both |
| | RPS16C | At5g18380 | | NSPA | N | N | | N | | N | | N | 62 | RNA only |
| RPS17 | RPS17A | At2g04390 | | NSPA | N | Y | | N | | N | | N | 35 | RNA only |
| | RPS17B | At2g05220 | | NSPA | N | Y | | N | | N | 61 | Y | 34 | Both |
| | RPS17C | At3g10610 | | NSPA | Y | Y | | N | | N | | N | 44 | RNA only |
| | RPS17D | At5g04800 | | NSPA | N | Y | 36 | Y | | N | | N | 37 | Both |
| RPS18 | RPS18A | At1g22780 | Whole family 100% | Y | N | Y | 41 | Y | 49 | Y | 44 | Y | 100 | Both |

(continued on next page)

Table 3 (continued)

| Annotation | | | | Public proteomic data | | | Proteomic data | | | | | | Ribo-Seq | Agreement |
|------------|--------------|---------------------|---------------------------------------|-----------------------|-------------------------|----------------------|--------------------------------------|----------|--------------------------------------|----------|-----------------------|----------|------------------------|---------------------|
| Family | Protein Name | Locus Name (TAIR10) | Protein Similarity and Other Comments | Chang et al., 2005 | Giavalisco et al., 2005 | Carroll et al., 2008 | 1D TRAP LC-MS | | 2D TRAP LC-MS | | Sedimented | | Juntawong et al., 2014 | Ribo-Seq Proteomics |
| | | | | [4] | [43] | [12] | Max Sequence Coverage per Sample (%) | Detected | Max Sequence Coverage per Sample (%) | Detected | Sequence Coverage (%) | Detected | [41] | |
| | RPS18B | At1g34030 | identical Whole family 100% identical | Y | Y | Y | | N | 49 | Y | 44 | Y | 65 | Both |
| | RPS18C | At4g09800 | Whole family 100% identical | Y | N | Y | | N | 49 | Y | 44 | Y | 83 | Both |
| RPS19 | RPS19A | At3g02080 | | Y | Y | Y | 49 | Y | 62 | Y | | N | 122 | Both |
| | RPS19B | At5g15520 | | NSPA | N | N | 49 | Y | 45 | Y | | N | 10 | Both |
| | RPS19C | At5g61170 | | Y | Y | Y | 44 | Y | 52 | Y | 42 | Y | 148 | Both |
| RPS20 | RPS20A | At3g45030 | 100% identical to RPS20C | NSPA | N | N | 46 | Y | | N | 47 | Y | 30 | Both |
| | RPS20B | At3g47370 | | Y | Y | N | | N | 37 | Y | 46 | Y | 65 | Both |
| | RPS20C | At5g62300 | 100% identical to RPS20A | NSPA | N | N | 46 | Y | | N | 47 | Y | 59 | Both |
| RPS21 | RPS21B | At3g53890 | | Y | N | N | | N | | N | | N | 191 | RNA only |
| | RPS21C | At5g27700 | | NSPA | N | N | | N | | N | | N | 108 | RNA only |
| RPS23 | RPS23A | At3g09680 | | Y | N | N | | N | | N | | N | 1 | None |
| | RPS23B | At5g02960 | | Y | Y | Y | | N | 49 | Y | | N | 189 | Both |
| RPS24 | RPS24A | At3g04920 | | Y | N | Y | 28 | Y | 34 | Y | | N | 104 | Both |
| | RPS24B | At5g28060 | | Y | Y | Y | 32 | Y | 34 | Y | 26 | Y | 118 | Both |
| RPS25 | RPS25A | At2g16360 | | NSPA | N | N | | N | 45 | Y | | N | 1 | Prot. only |
| | RPS25B | At2g21580 | | Y | N | N | 41 | Y | 36 | Y | | N | 81 | Both |
| | RPS25D | At4g34555 | | NSPA | N | N | | N | 44 | Y | | N | 15 | Both |
| | RPS25E | At4g39200 | | NSPA | Y | N | 31 | Y | | N | 46 | Y | 73 | Both |
| RPS26 | RPS26A | At2g40510 | | NSPA | N | Y | | N | | N | | N | 19 | RNA only |
| | RPS26B | At2g40590 | | NSPA | N | Y | 50 | Y | 50 | Y | | N | 13 | Both |
| | RPS26C | At3g56340 | | NSPA | Y | Y | | N | 49 | Y | 49 | Y | 144 | Both |
| RPS27a | RPS27aA | At1g23410 | | N | N | N | | N | | N | | N | 8 | RNA only |
| | RPS27aB | At2g47110 | | N | N | N | | N | | N | | N | 98 | RNA only |
| | RPS27aC | At3g62250 | | N | N | Y | | N | 54 | Y | | N | 176 | Both |
| RPS27 | RPS27A | At2g45710 | | NSPA | Y | Y | | N | | N | | N | 15 | RNA only |
| | RPS27B | At3g61110 | | Y | N | Y | | N | | N | | N | 157 | RNA only |
| | RPS27C | At3g61111 | | N | N | N | | N | | N | | N | 1 | None |
| | RPS27D | At5g47930 | | NSPA | Y | N | 45 | Y | 36 | Y | | N | 160 | Both |
| RPS28 | RPS28A | At3g10090 | 100% identical to RPS28B | Y | Y | N | 25 | Y | 41 | Y | | N | 28 | Both |
| | RPS28B | At5g03850 | 100% identical to RPS28A | Y | N | N | 25 | Y | 41 | Y | | N | 165 | Both |
| | RPS28C | At5g64140 | | Y | N | Y | 30 | Y | | N | | N | 69 | Both |
| RPS29 | RPS29A | At3g43980 | Whole family 100% identical | N | N | Y | | N | | N | | N | 20 | RNA only |
| | RPS29B | At3g44010 | Whole family 100% identical | N | N | Y | | N | 70 | Y | | N | 75 | Both |
| | RPS29C | At4g33865 | Whole family 100% identical | N | N | Y | | N | 70 | Y | | N | 157 | Both |
| RPS30 | RPS30A | At2g19750 | Whole family 100% identical | N | N | Y | | N | 19 | Y | | N | 45 | Both |
| | RPS30B | At4g29390 | Whole family 100% identical | N | N | Y | | N | 19 | Y | | N | 89 | Both |
| | RPS30C | At5g56670 | Whole family 100% identical | N | N | Y | | N | 19 | Y | | N | 91 | Both |
| RACK1 | RACK1A | At1g18080 | | Y | Y | Y | 52 | Y | 57 | Y | | N | 252 | Both |
| | RACK1B | At1g48630 | | NSPA | N | Y | 37 | Y | 41 | Y | | N | 48 | Both |
| | RACK1C | At3g18130 | | NSPA | N | Y | 35 | Y | 49 | Y | | N | 32 | Both |

Table 4
Protein identification data of identified ribosomal proteins.

| | Immuno purified 1D LC-MS | Immuno purified 2D LC-MS | Density based purification |
|--|--------------------------|--------------------------|----------------------------|
| Number of genes identified ^a | 129 | 141 | 83 |
| Average of max. protein sequence coverage (%) of identified proteins | 43 | 45 | 45 |
| Average number of spectra | 50,000 | 91,000 | 17,808 |
| Non-redundant peptide sequences | 1541 | 1580 | 730 |

^a Identified with at least two proteotypic peptides.

Therefore, the RPKM reported in this analysis represents the gene-specific ribosome reads on each r-protein paralog. The RPKM values confirming translation of r-protein paralogs transcripts were compared to three independent ribosome proteome analyses [4,12,43] and the three proteomic datasets reported here. This identified 159 r-protein genes/proteins that were detected by both Ribo-seq and MS proteomics. Of the 80 r-protein paralogs that had not been detected by any proteomic study, ribosome footprints were confirmed for 58 of the corresponding r-protein genes. This may reflect a difference between seedlings used for the Ribo-seq analysis and rosette leaves used for the higher resolution 2D MS analyses. The 58 r-protein genes identified only in the Ribo-seq analysis could include some r-proteins that are synthesized but do not assemble into ribosomes; alternatively, these could indicate the simply greater identification capacity of the RNA detection method over the proteomics. The r-proteins detected only by the RNA based method (RNA only) were of typically lower molecular mass and have a higher frequency of arginine and lysine and consequently digested to small peptides not detected under the current settings. Six r-proteins were only identified by proteomics, which might reflect the different plant material used in the RNA based experiments.

Seventy-two r-protein families were detected in at least one of the three different proteomic methods performed here. For a majority of these identified r-proteins, the number of times the protein is detected in different proteomic reports (Table 3) correlated with the RPKM value in the Ribo-seq dataset (Table 3). For 11 r-protein families we failed to identify any protein by proteomics, whereas in the Ribo-seq the transcripts were detected with ≥ 5 RPKM that mapped the paralogs. Members of six of these families were identified in one or more of the other proteomic studies (RPL21, RPL23a, RPL36a, RPL39, RPL40 and RPS21) (Table 3). This could reflect use of different plant material as the source of ribosomes (i.e. cell cultures [4,12] and leaves [43]). Members of four r-protein families with detected ribosome footprints have never been successfully identified by proteomics in *Arabidopsis* ribosomes (RPL29, RPL39, RPL40 and RPL41). The reason for this may include, (i) the low molecular mass and high lysine/arginine content, which is problematic

for current MS detection methods; (ii) possible non-stoichiometric abundance in the ribosome; and/or (iii) absence from ribosomes.

This study also provides some additional resolution to the question about members of the six-member *RPS15a* gene family, which resolves into two phylogenetic subclasses in angiosperms, type I (*RPS15aA*, *RPS15aC*, *RPS15aD*, *RPS15aE*) and type II (*RPS15aB*, *RPS15aE*), all of which encode a protein evolutionarily related to eubacterial-like *rps8* that is a component of chloroplast and likely mitochondrial ribosomes [4]. Other yeast and animal model genomes encode Type I *RPS15a* proteins that are present in cytosolic ribosomes but Type II *RPS15a* genes are absent. *Arabidopsis* and other angiosperms lack a mitochondrially-encoded eubacteria-like *rps8* gene, and *in vitro* mitochondrion import assay using soybean mitochondria confirmed that *Arabidopsis* S15aE (named *RPS15aE*; At4g29430) can be imported into the mitochondria, whereas *RPS15aA* (At1g07770) is not imported [47]. This led to the hypothesis that nuclear-encoded *RPS15aE* functionally replaces the organellar *rps8* in *Arabidopsis*. Chang et al., [4] identified *RPS15aE* among the r-proteins obtained by density centrifugation and 2D-PAGE, although at lower abundance than Type I S15a proteins [4]. *RPS15aE* was not identified in immunopurified or sedimented ribosomes in this study, whereas Carroll et al. (2008) only identified this protein in mitochondrial ribosomes [12]. Altogether the data support the conclusion that Type II S15a proteins are not a component of cytosolic ribosomes.

4. Conclusions

The *Arabidopsis* genome encodes 242 putatively functional r-protein genes (including three *RACK1* genes) comprising 81 families of cytosolic r-proteins that could combine into a multitude of distinguishable ribosomes. In the present study using LC-MS^E, members of 70 r-protein gene families including *RACK1* were identified as products of 165 r-protein genes. Differential incorporation of the products of r-protein paralogs into the ribosome in certain combinations could result in distinct ribosome function. In addition to the variation in protein composition, differences in post-translational modifications including phosphorylation can also contribute to distinctions in ribosome function. But

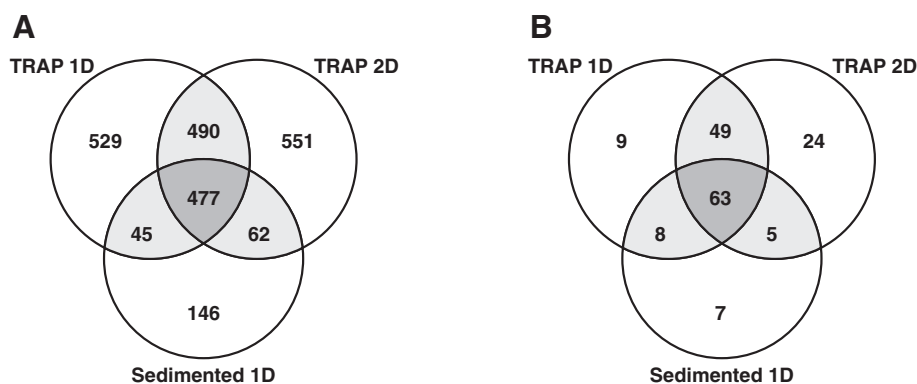


Fig. 1. Comparison of the peptides and proteins detected using the different ribosome purification and LC-MS detection methods. A. Venn diagram shows the detected peptide sequences in the immunopurified ribosomes analyzed by 1D LC-MS mode (TRAP 1D) and by 2D LC-MS mode (TRAP 2D) compared with the detected peptide sequences obtained with the density-purified ribosome 1D-LC-MS analysis (Sedimented). B. Venn diagram shows the detected r-proteins (including three *RACK1* proteins) in the immunopurified ribosome analyzed by 1D LC-MS mode (TRAP 1D) and by 2D LC-MS mode (TRAP 2D) compared to the detected r-proteins obtained with the density-purified ribosome 1D-LC-MS analysis (Sedimented). Proteins were considered to be detected only if two proteotypic peptides (or matching 100% identical proteins) for each protein were detected.

the consequences of such ribosome heterogeneity on global or mRNA specific rates of translation remain unknown. Over 20 r-protein mutant paralogs have been identified in genetic screens and the documented phenotypic deviation of these mutants from wild type shows their importance in developmental processes [37,48,49]. The molecular mechanism underlying altered development in these mutants is still unknown, but may include heterogeneity that has ramifications on mRNA translation and other processes [50,51]. This report provides a foundation for future research into the function of plant ribosomes and the relevance of their heterogeneity in higher plants.

Public availability of data: The raw data can be accessed from the PRIDE archive (<http://www.ebi.ac.uk/pride/archive/>) under the following accession numbers: PRD000415 (1D LC-MS^E), PXD002350 (2D LC-MS^E) and PXD002376 (Sedimented ribosomes). Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jprot.2015.07.004>.

Conflict of interest

The authors have declared no conflict of interest.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgments

We thank Jolanda Schuurmans for excellent technical assistance and Dr. Bas van Breukelen and Henk van den Toorn of the Biomolecular Mass Spectrometry and Proteomics Group at Utrecht University for their comments on the manuscript. This work was supported by grants from the Netherlands Genomics Initiative (NGI), Earth and Life Sciences Foundation subsidized by the Netherlands Organization for Scientific Research (NWO) and Centre for BioSystems Genomics (CBSG) and the US National Science Foundation (MCB-1021969). We thank Bio4Energy, a strategic research environment appointed by the Swedish government, for supporting this work. This work was partly supported by a French ANR grant (Blanc2011-SV6-01002) to Christian Meyer. Thomas Dobrenel is supported by the Kempe Foundation and by the French Ministère de la Recherche et de l'Enseignement Supérieur. Parts of this work have been supported by CBSG and the Dutch Proteomics Centre Hotel grants.

References

- [1] K. Yamada, J. Lim, J.M. Dale, H. Chen, P. Shinn, C.J. Palm, et al., Empirical analysis of transcriptional activity in the *Arabidopsis* genome, *Science* 302 (2003) 842–846.
- [2] N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science* 289 (2000) 905–920.
- [3] V. Ramakrishnan, B.T. Wimberly, D.E. Brodersen, W.M. Clemons, R.J. Morgan-Warren, A.P. Carter, et al., Structure of the 30S ribosomal subunit, *Nature* 407 (2000) 327–339.
- [4] I.F. Chang, K. Szick-Miranda, S. Pan, J. Bailey-Serres, Proteomic characterization of evolutionarily conserved and variable proteins of *Arabidopsis* cytosolic ribosomes, *Plant Physiol.* 137 (2005) 848–862.
- [5] T.M. Schmeing, V. Ramakrishnan, What recent ribosome structures have revealed about the mechanism of translation, *Nature* 461 (2009) 1234–1242.
- [6] A. Nakao, M. Yoshihama, N. Kenmochi, RPG: the Ribosomal Protein Gene database, *Nucleic Acids Res.* 32 (2004) D168–D170.
- [7] S. Klinge, F. Voigts-Hoffmann, M. Leibundgut, N. Ban, Atomic structures of the eukaryotic ribosome, *Trends Biochem. Sci.* 37 (2012) 189–198.
- [8] A. Barakat, K. Szick-Miranda, I.F. Chang, R. Guyot, G. Blanc, R. Cooke, et al., The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome, *Plant Physiol.* 127 (2001) 398–415.
- [9] K. Szick-Miranda, A.S. Zaniak, A.S. Zaniak, S. Abiddayo, K.L.C. Slater, Analysis of RPS15aE, an isoform of a plant-specific evolutionarily distinct ribosomal protein in *Arabidopsis thaliana*, reveals its potential role as a growth regulator, *Plant Mol. Biol. Rep.* 28 (2010) 239–252.
- [10] M.E. Zanetti, I.F. Chang, F. Gong, D.W. Galbraith, J. Bailey-Serres, Immunopurification of polyribosomal complexes of *Arabidopsis* for global analysis of gene expression, *Plant Physiol.* 138 (2005) 624–635.
- [11] Z. Zhang, P. Harrison, M. Gerstein, Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome, *Genome Res.* 12 (2002) 1466–1482.
- [12] A.J. Carroll, J.L. Heazlewood, J. Ito, A.H. Millar, Analysis of the *Arabidopsis* cytosolic ribosome proteome provides detailed insights into its components and their post-translational modification, *Mol. Cell. Proteomics* 7 (2008) 347–369.
- [13] G.-Z. Li, J.P.C. Vissers, J.C. Silva, D. Golick, M.V. Gorenstein, S.J. Geromanos, Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures, *Proteomics* 9 (2009) 1696–1719.
- [14] S.J. Geromanos, J.P.C. Vissers, J.C. Silva, C.A. Dorschel, G.-Z. Li, M.V. Gorenstein, et al., The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS, *Proteomics* 9 (2009) 1683–1695.
- [15] M. Schmid, T.S. Davison, S.R. Henz, U.J. Pape, M. Demar, M. Vingron, et al., A gene expression map of *Arabidopsis thaliana* development, *Nat. Genet.* 37 (2005) 501–506.
- [16] D.F. Louie, K.A. Resing, T.S. Lewis, N.G. Ahn, Mass spectrometric analysis of 40S ribosomal proteins from Rat-1 fibroblasts, *J. Biol. Chem.* 271 (1996) 28189–28198.
- [17] R.J. Arnold, B. Polevoda, J.P. Reilly, F. Sherman, The action of N-terminal acetyltransferases on yeast ribosomal proteins, *J. Biol. Chem.* 274 (1999) 37035–37040.
- [18] S.W. Lee, S.J. Berger, S. Martinovic, L. Pasa-Tolic, G.A. Anderson, Y. Shen, et al., Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR, *PNAS* 99 (2002) 5942–5947.
- [19] T.I. Odintsova, E.-C. Müller, A.V. Ivanov, T.A. Egorov, R. Bienert, S.N. Vladimirov, et al., Characterization and analysis of posttranslational modifications of the human large cytoplasmic ribosomal subunit proteins by mass spectrometry and Edman sequencing, *J. Protein Chem.* 22 (2003) 249–258.
- [20] Y. Yu, H. Ji, J.A. Doudna, J.A. Leary, Mass spectrometric analysis of the human 40S ribosomal subunit: native and HCV IRES-bound complexes, *Protein Sci.* 14 (2005) 1438–1446.
- [21] M. Yoshihama, T. Uechi, S. Asakawa, K. Kawasaki, S. Kato, S. Higa, et al., The human ribosomal protein genes: sequencing and comparative analysis of 73 genes, *Genome Res.* 12 (2002) 379–390.
- [22] A.L. Veuthey, G. Bittar, Phylogenetic relationships of fungi, plantae, and animalia inferred from homologous comparison of ribosomal proteins, *J. Mol. Evol.* 47 (1998) 81–92.
- [23] A.J. Carroll, The *Arabidopsis* cytosolic ribosomal proteome: from form to function, *Front. Plant Sci.* 4 (2013) 32.
- [24] K. Szick-Miranda, J. Bailey-Serres, Regulated heterogeneity in 12-kDa P-protein phosphorylation and composition of ribosomes in maize (*Zea mays* L.), *J. Biol. Chem.* 276 (2001) 10921–10928.
- [25] A.J. Williams, J. Werner-Fraczek, L.-F. Chang, J. Bailey-Serres, Regulated phosphorylation of 40S ribosomal protein S6 in root tips of maize, *Plant Physiol.* 132 (2003) 2086–2097.
- [26] J. Bailey-Serres, S. Vangala, K. Szick, C.H. Lee, Acidic phosphoprotein complex of the 60S ribosomal subunit of maize seedling roots. Components and changes in response to flooding, *Plant Physiol.* 114 (1997) 1293–1305.
- [27] M.V. Turkina, H. Klang Arstrand, A.V. Vener, Differential phosphorylation of ribosomal proteins in *Arabidopsis thaliana* plants during day and night, *PLoS One* 6 (2011) e29307.
- [28] D. Weijers, M. Franke-van Dijk, R.J. Vencken, A. Quint, P. Hooykaas, R. Offringa, An *Arabidopsis* Minute-like phenotype caused by a semi-dominant mutation in a ribosomal protein S5 gene, *Development* 128 (2001) 4289–4299.
- [29] I. Tzafirir, R. Pena-Muralla, A. Dickerman, M. Berg, R. Rogers, S. Hutchens, et al., Identification of genes required for embryo development in *Arabidopsis*, *Plant Physiol.* 135 (2004) 1206–1220.
- [30] A. Rosado, E.J. Sohn, G. Drakakaki, S. Pan, A. Swidergal, Y. Xiong, et al., Auxin-mediated ribosomal biogenesis regulates vacuolar trafficking in *Arabidopsis*, *Plant Cell* 22 (2010) 143–158.
- [31] T. Ito, G.T. Kim, K. Shinozaki, Disruption of an *Arabidopsis* cytoplasmic ribosomal protein S13-homologous gene by transposon-mediated mutagenesis causes aberrant growth and development, *Plant J.* 22 (2000) 257–264.
- [32] M. Van Lijsebettens, R. Vanderhaeghen, M. De Block, G. Bauw, R. Villarroel, M. Van Montagu, An S18 ribosomal protein gene copy at the *Arabidopsis* PFL locus affects plant development by its specific expression in meristems, *EMBO J.* 13 (1994) 3378–3388.
- [33] Y. Yao, Q. Ling, H. Wang, H. Huang, Ribosomal proteins promote leaf adaxial identity, *Development* 135 (2008) 1325–1334.
- [34] V. Pinon, J.P. Etchells, P. Rossignol, S.A. Collier, J.M. Arroyo, R.A. Martienssen, et al., Three *PIGGYBACK* genes that specifically influence leaf patterning encode ribosomal proteins, *Development* 135 (2008) 1315–1324.
- [35] M.L. Falcone Ferreyra, A. Pezza, J. Biarc, A.L. Burlingame, P. Casati, Plant L10 ribosomal proteins have different roles during development and translation under ultraviolet-B stress, *Plant Physiol.* 153 (2010) 1878–1894.
- [36] J.R. Warner, K.B. McIntosh, How common are extraribosomal functions of ribosomal proteins? *Mol. Cell* 34 (2009) 3–11.
- [37] G. Horiguchi, M. Van Lijsebettens, H. Candela, J.L. Micol, H. Tsukaya, Ribosomes and translation in plant developmental control, *Plant Sci.* 191–192 (2012) 24–34.
- [38] D. Deprost, L. Yao, R. Sormani, M. Moreau, G. Leterreux, M. Nicolai, et al., The *Arabidopsis* TOR kinase links plant growth, yield, stress resistance and mRNA translation, *EMBO Rep.* 8 (2007) 864–870.
- [39] J. Bailey-Serres, M. Freeling, Hypoxic stress-induced changes in ribosomes of maize seedling roots, *Plant Physiol.* 94 (1990) 1237–1243.
- [40] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227 (1970) 680–685.

- [41] P. Juntawong, T. Girke, J. Bazin, J. Bailey-Serres, Translational dynamics revealed by genome-wide profiling of ribosome footprints in *Arabidopsis*, PNAS 111 (2014) E203–E212.
- [42] M. Hummel, J.H.G. Cordewener, J.C.M. de Groot, S.C.M. Smeekens, A.H.P. America, J. Hanson, Dynamic protein composition of *Arabidopsis thaliana* cytosolic ribosomes in response to sucrose feeding as revealed by label free MS^E proteomics, Proteomics 12 (2012) 1024–1038.
- [43] P. Giavalisco, D. Wilson, T. Kreitler, H. Lehrach, J. Klose, J. Gobom, et al., High heterogeneity within the ribosomal proteins of the *Arabidopsis thaliana* 80S ribosome, Plant Mol. Biol. 57 (2005) 577–591.
- [44] J. Cao, K. Schneeberger, S. Ossowski, T. Günther, S. Bender, J. Fitz, et al., Whole-genome sequencing of multiple *Arabidopsis thaliana* populations, Nat. Genet. 43 (2011) 956–963.
- [45] R. Craig, R.C. Beavis, TANDEM: matching proteins with tandem mass spectra, Bioinformatics 20 (2004) 1466–1467.
- [46] M.-J. Liu, S.-H. Wu, J.-F. Wu, W.-D. Lin, Y.-C. Wu, T.-Y. Tsai, et al., Translational landscape of photomorphogenic *Arabidopsis*, Plant Cell 25 (2013) 3699–3710.
- [47] K.L. Adams, D.O. Daley, J. Whelan, J.D. Palmer, Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts, Plant Cell 14 (2002) 931–943.
- [48] B. Roy, A.G. von Arnim, Translational regulation of cytoplasmic mRNAs, The Arabidopsis Book 11 (2013) e0165.
- [49] M.E. Byrne, A role for the ribosome in development, Trends Plant Sci. 14 (2009) 512–519.
- [50] A. Rosado, N.V. Raikhel, Application of the gene dosage balance hypothesis to auxin-related ribosomal mutants in *Arabidopsis*, Plant Signal. Behav. 5 (2010) 450–452.
- [51] B.S. Tiruneh, B.-H. Kim, D.R. Gallie, B. Roy, A.G. von Arnim, The global translation profile in a ribosomal protein mutant resembles that of an eIF3 mutant, BMC Biol. 11 (2013) 123.