Gene 548 (2014) 81–90

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Bioinformatics analysis of plant orthologous introns: identification of an intronic tRNA-like sequence

Evgeny E. Akkuratov [a,b], Lorraine Walters [b,c], Arnab Saha-Mandal [b,1], Sushant Khandekar [d], Erin Crawford [e], Craig L. Zirbel [f], Scott Leisner [d], Ashwin Prakash [e,2], Larisa Fedorova [e], Alexei Fedorov [b,e,*]

[a] Faculty of Biology and Soil Science, St. Petersburg State University, St. Petersburg 199034, Russia
[b] Program in Bioinformatics and Proteomics/Genomics, University of Toledo, Health Science Campus, Toledo, OH 43614, USA
[c] Department of Bioengineering, University of Toledo, Main Campus, Toledo, OH 43606, USA
[d] Department of Biological Sciences, University of Toledo, Main Campus, Toledo, OH 43606, USA
[e] Department of Medicine, University of Toledo, Health Science Campus, Toledo, OH 43614, USA
[f] Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA

ARTICLE INFO

Article history:
Received 29 December 2013
Received in revised form 26 June 2014
Accepted 7 July 2014
Available online 8 July 2014

Keywords:
Genomics
Computational biology
MALAT1
mascRNA

ABSTRACT

Orthologous introns have identical positions relative to the coding sequence in orthologous genes of different species. By analyzing the complete genomes of five plants we generated a database of 40,512 orthologous intron groups of dicotyledonous plants, 28,519 orthologous intron groups of angiosperms, and 15,726 of land plants (moss and angiosperms). Multiple sequence alignments of each orthologous intron group were obtained using the Mafft algorithm. The number of conserved regions in plant introns appeared to be hundreds of times fewer than that in mammals or vertebrates. Approximately three quarters of conserved intronic regions among angiosperms and dicots, in particular, correspond to alternatively-spliced exonic sequences. We registered only a handful of conserved intronic ncRNAs of flowering plants. However, the most evolutionarily conserved intronic region, which is ubiquitous for all plants examined in this study, including moss, possessed multiple structural features of tRNAs, which caused us to classify it as a putative tRNA-like ncRNA. Intronic sequences encoding tRNA-like structures are not unique to plants. Bioinformatics examination of the presence of tRNA inside introns revealed an unusually long-term association of four glycine tRNAs inside the *Vac14* gene of fish, amniotes, and mammals.

1. Introduction

Introns are ubiquitous elements of eukaryotic genomes that perform several essential cellular functions (Fedorova and Fedorov, 2003; Mattick, 1994; Morello and Breviario, 2008). The relative abundance and the length of introns vary significantly in diverse branches of eukaryotes indicating heterogeneity in the importance of intron functions concomitant with phylogenetic divergence. This paper is focused on the comparison of intron roles in plants and animals.

Usually, functional regions of introns have evolutionarily conserved nucleotide sequences that have been preserved over millions of years (Rearick et al., 2011; Sironi et al., 2005). These functional intronic sequences may be one of the following types: 1) protein binding sites (enhancers or silencers for transcription or splicing); 2) alternatively-spliced intronic sequences that are incorporated into mRNAs for selective transcripts of the gene; and 3) non-coding RNA molecules including small ncRNAs (snoRNAs, microRNAs, endogenous siRNA, and piwiRNA) and long ncRNAs (analogous to lincRNA). Recently, our team generated a database of orthologous introns for five mammalian species (Rearick et al., 2011). Within this set of 63,000 groups of orthologous introns, thousands of evolutionarily conserved sequence segments were characterized and associated with various types of non-coding RNAs. In the current study, we took advantage of the availability of recently sequenced genomes of five plant species and performed the same bioinformatics analysis of orthologous introns within kingdom Plantae. We generated Exon–Intron Databases for moss (*Physcomitrella patens*; Pp), rice (*Oryza sativa* ssp. *japonica*; Os), poplar (*Populus trichocarpa*; Pt), grape (*Vitis vinifera*; Vv), and mouse ear cress (*Arabidopsis thaliana*; At). Then, we created databases of orthologous introns of dicots, angiosperms, and land plants. Sequences of each orthologous group were

Abbreviations: nts, nucleotides; CPIR-1, Conserved Plant Intronic Region.

* Corresponding author at: Program in Bioinformatics and Proteomics/Genomics, University of Toledo, Health Science Campus, Toledo, OH 43614, USA.
E-mail addresses: akkuratov.evgeny@gmail.com (E.E. Akkuratov), lorraine.walters@rockets.utoledo.edu (L. Walters), asahaman@ucalgary.ca (A. Saha-Mandal), sushant43606@yahoo.com (S. Khandekar), erin.crawford@utoledo.edu (E. Crawford), zirbel@bgsu.edu (C.L. Zirbel), scott.leisner@utoledo.edu (S. Leisner), ashwin.prakash@jhmi.edu (A. Prakash), lvfedorova3@gmail.com (L. Fedorova), Alexei.fedorov@utoledo.edu (A. Fedorov).
[1] Current address: Biochemistry and Molecular Biology Graduate Program, Alberta Children's Hospital Research Institute, University of Calgary, Canada.
[2] Current address: Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

http://dx.doi.org/10.1016/j.gene.2014.07.012
0378-1119/

aligned and conserved regions were characterized. This large-scale computational investigation revealed a prominent difference in the number, length and location of conserved intronic regions of plants and animals.

To compare the evolution of plant and animal intron sequences, different branches of vertebrates were used that evolved during approximately the same period of time as the plant taxa. Vertebrates were chosen because a number of their genomes have been completely sequenced and characterized in great detail. The entire group of our five land plant species (in which moss separated from the flowering plants approximately 420 million years ago, mya) may be compared with the *Osteichthyes* taxon of bony fish and tetrapods that has ~500 mya history. The ancestors of angiosperms diverged from gymnosperms around 245–202 mya, and the first angiosperms known to exist were from 140 mya (Moore et al., 2007). Hence, flowering plants may be compared to eutherian and marsupial mammals that diverged about 180–190 mya. Finally, dicotyledonous plants, which appeared in the fossil records as early as 110 mya (http://lifeofplant.blogspot.com/2011/04/eudicots.html), may be compared to the eutherian taxon of placental mammals, that originated about 100 mya (Venditti et al., 2011). However, we acknowledge that the divergence time estimates in plant phylogeny is highly heterogeneous and conditional on several assumptions of nucleotide substitution and molecular clock hypothesis. Some estimations of evolutionary time of plant separation have been discussed in Chernikova et al. (2011).

We acknowledge the availability of alternative public resources that present plant exon–intron datasets and in particular, the Common Introns Within Orthologous Genes (CIWOG) database described by Wilkerson et al. (2009). Another recent database built with the same objective is Plant Intron and Exon Comparison and Evolution (PIECE) (Wang et al., 2013). However, we used our own pipelines of programs for generation/verification of orthologous introns, which we have been working with for the past decade (Fedorov et al., 2005; Rearick et al., 2011). The primary reason for this is that all such analyses and databases come with key assumptions and parameter designations, and it was particularly important to have a database that could be compared directly to our extensive work with mammalian introns (Rearick et al., 2011). Comparison of our database with the others could also be useful in distinguishing borderline orthology assignments.

2. Materials and methods

2.1. Databases of plant orthologous introns

We defined orthologous introns as introns from orthologous genes that have the same position and phase relative to the coding sequence (Fedorov et al., 2005).

Genomic Exon–Intron Databases (EIDs) of five plant species were generated with our EID pipeline of programs described by Shepelev and Fedorov (2006). These plant EIDs are available from our web page (http://bpg.utoledo.edu/~afedorov/lab/eid.html). Using these Exon–Intron Databases we generated a database of orthologous introns of five plant species by the same algorithms and pipelines of programs as described previously for mammalian orthologous introns (Rearick et al., 2011). All program parameters and threshold settings used in the creation of plant orthologous introns were the same as those used for the creation of mammalian orthologous introns. All these programs are available from our web page (http://bpg.utoledo.edu/~afedorov/lab/prog.html). The corresponding protocol for the programs' execution is presented in Supplementary Fig. S1. Computation of plant intronic sequences for five species (*Arabidopsis*, poplar, grape, rice, and moss) produced 15,726 orthologous intron groups, while four species of flowering plants (*Arabidopsis*, poplar, grape, and rice) resulted in 28,519 groups. When the computation was confined to three species of dicotyledonous plants (*Arabidopsis*, poplar, and grape), 40,512 groups were obtained. Each group of orthologous introns was aligned with the Mafft program (version 6) (Katoh et al., 2002) and the entire set of multiple alignments of orthologous introns is available from our EID web site (http://bpg.utoledo.edu/~afedorov/lab/eid_plantOl.html). This web site contains 3pln_mafft.gz (25.8 MB file for Grape_Poplar_Arabidopsis alignments); 4pln_mafft.gz (22.7 MB file for Grape_Poplar_Arabidopsis_Rice alignments); and 5pln_mafft.gz (14.5 MB file for Grape_Poplar_Arabidopsis_Rice_Moss).

2.2. Computational analysis of RNA secondary structures

Sequences of the most evolutionarily Conserved Plant Intronic Region (so-called CPIR-1) for each of the five plant species were examined through the online *RNAfold* program from the *Vienna RNA* web server (http://rna.tbi.univie.ac.at/) (Gruber et al., 2008). The program predicts the minimum free energy of the secondary structure using the dynamic programming algorithm of Zuker and Stiegler (1981) and also calculates the equilibrium base pairing probabilities via McCaskill's partition function algorithm (McCaskill, 1990). The output consists of a graphical display of secondary structure, colored by base pairing probabilities.

For studying evolutionary conservation of RNA secondary structures of plant CPIR-1 sequences we used the *RNAalifold* online program, also available at the Vienna RNA web server, which predicts the consensus structure of a set of aligned RNA sequences (Hofacker et al., 2002). It averages energy contributions from all sequences in addition to applying the dynamic programming algorithm. The input was a CLUSTALW multiple alignment of CPIR-1 sequences of five plant species.

The *JAR3D* server (http://rna.bgsu.edu/jar3d) was used to identify possible recurrent 3D structural motifs from the sequences of conserved internal and hairpin loops in predicted secondary structures (Zirbel CL, Petrov AI, Roll J, Leontis NB, in preparation). JAR3D is a novel toolkit developed from the *WebFR3D* server (http://rna.bgsu.edu/FR3D) (Sarver et al., 2008).

Possible tRNA structures were examined using the online *tRNAscan-SE* web server under default parameters (http://lowelab.ucsc.edu/tRNAscan-SE/) (Schattner et al., 2005).

2.3. BLAST analysis of Conserved Plant Intronic Regions

Evolutionarily conserved intronic sequences were examined for their possible involvement in alternative splicing events or being non-coding RNAs using online BLASTN searching of their presence inside RNA databases. Specifically, intronic sequences were searched against the expressed sequence tags (ESTs, excluding human and mouse sequences) database at NCBI using the standard nucleotide BLASTN program (Altschul et al., 1990) with filters turned off. The EST sequences with fair matches (BLAST e-value < 0.001) against the conserved portion of intronic sequences were visually inspected and further analyzed for the presence of neighboring exons from the same gene. If the corresponding ESTs showed a reliable representation (BLAST e-value $< 10^{-6}$) of at least two properly spliced exons and the borders of conserved portion of the intronic sequences also possessed consensus of splicing junctions, they were regarded as alternatively spliced exons. In addition, the conserved intronic sequences were searched against the functional RNA database (fRNAdb) (http://www.ncrna.org/frnadb/) (Kin et al., 2007) and non-coding RNA database (NONCODE) (http://www.noncode.org/NONCODERv3/) (Bu et al., 2012) to detect possible functional RNAs.

2.4. tRNA phylogeny

The entire pool of *Arabidopsis* tRNA sequences (GtRNAdb-all-tRNAs.fa.gz) was downloaded from http://lowelab.ucsc.edu/GtRNAdb/download.html. This set has a total of 639 sequences (Chan and Lowe, 2009). Using *Perl*, these sequences were classified into groups of tRNAs bearing the same codon specificity, which gave 47 different groups. tRNAs with the highest cove-scores (Lowe and Eddy, 1997) from each group were extracted and, together with CPIR-1 sequence,

were aligned using the Clustal Omega (Goujon et al., 2010; Sievers et al., 2011) web resource. ClustalW2-phylogeny (Goujon et al., 2010; Larkin et al., 2007) was used for phylogenetic computations of the resulting alignment, using the neighbor joining clustering algorithm (Saitou and Nei, 1987). The sequence phylogeny thus obtained, with explanations, is presented in Fig. S2. The phylogenetic tree with specified branch lengths is also provided in an encrypted Newick file format "Akkuratov_Suppl_phylogenet.nwk.txt" in Supplementary materials that can be interactively examined by tree-viewing packages like TreeView (Page, 1996) and other compatible programs.

2.5. Mapping gene functions from the Gene Ontology database

The gene association files for *Arabidopsis* were downloaded from the Gene Ontology (GO) database (http://www.geneontology.org/GO.downloads.annotations.shtml) (Ashburner et al., 2000). The file "gene_association.tair.gz" provided a comprehensive source for *A. thaliana* GO annotations composed of gene associations made by The Arabidopsis Information Resource (TAIR) and The Institute for Genomic Research (TIGR) (Berardini et al., 2004). The GO functions for our entire set of 71 *Arabidopsis* genes (inside which we characterized evolutionarily conserved intronic regions) were mapped and further identified for functional enrichment. Statistical evaluation was performed using a Monte-Carlo simulation with in-house *Perl* scripts. For this purpose, 1000 random samples of 71 *Arabidopsis* genes (the same number as our gene samples containing intronic conserved regions) were created. For each of the random gene sample thus chosen, GO functions were subsequently analyzed and compared with that of our actual 71 *Arabidopsis* genes hosting intronic conserved regions.

2.6. Computation of intronic tRNAs in humans and mouse

GenBank feature-tables of each human chromosome (Build 37.3) were downloaded from the NCBI ftp site (ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/ARCHIVE/BUILD.37.3/) onto a local Linux workstation. In-house *Perl* programs encompassing coordinates of tRNAs and introns from the feature-tables were created. Application of these programs resulted in characterization of 53 cases of tRNAs inside human introns and exons. Then, cases with tRNAs inside pseudogenes and non-protein coding genes were discarded, resulting in a total of 24 tRNAs inside 14 experimentally confirmed protein-coding genes (Table S1). For each gene, the corresponding ortholog in mouse was identified by reciprocal best-match of protein sequences. Orthologous introns for these 14 human–mouse gene pairs were characterized and the presence of human tRNAs were checked in orthologous introns of mouse using the BLASTN program with default parameters with all filters on repetitive sequences turned off. This search revealed that identified human tRNAs were absent in orthologous mouse introns except for 4 glycine tRNAs, all of which were present inside the gene VAC14. The secondary structures of those tRNAs were obtained from tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) (Schattner et al., 2005).

2.7. Plant growth conditions and RNA analysis

A. thaliana Col-0 ecotype plants were grown hydroponically as described in Li et al. (2008). Plants grown under these conditions were labeled as controls and were harvested in the middle of the 16 h light period of the light/dark cycle. Some plants were placed in the dark for 12 h instead of 8 h, then harvested and labeled as Dark-treated plants. Some plants were placed in a cold room (4 °C) for 8 h under 100 μE of light, whereas others were placed in the cold for 8 h under the given light conditions and then moved back to the normal growth chamber for 24 h to recover from the cold stress. The above ground tissues from three plants for each treatment were individually harvested and RNA was isolated from each individual plant using the Qiagen RNeasy Plant Mini Kit. RNA quality was assessed by formaldehyde gel electrophoresis and quantified by Nanodrop spectrophotometry. A total of 5 μg of RNA was then loaded in a well of a formaldehyde 2% agarose gel.

To examine the expression of the putative CPIR-1 RNA in various tissues, plants were grown as above under control conditions and harvested at 35 days. By this time the plants had flowered. Plants were separated into rosette leaves, inflorescence, and roots. Each tissue from 6 plants was pooled together and RNA was isolated as above.

2.8. Northern blot experiments

A probe for use in Northern blots against the tRNA-like ncRNA in *A. thaliana* gene AT5G13240 was developed by designing PCR primers that flank the targeted region using Oligo primer analysis software v. 6.71 (Molecular Biology Insights, Inc., Cascade CO). The following primers were selected: forward — 5′ GAA AGG CTT TGA TCT ACT TG 3′ and reverse — 5′ TGT CCC AGC TTT CCT CCG AG 3′. Primer sequences were checked for specificity using NCBI/Primer-BLAST against the reference assembly for *A. thaliana*. Primers were validated by PCR using genomic DNA from *A. thaliana* ecotype Col-0. PCR conditions were as follows: 10 μl reactions containing 50 ng of gDNA, 0.05 μg of each primer, 1 × PCR buffer containing 3 mM $MgCl_2$ (cat# 1778, Idaho Technology, Inc., Salt Lake City, UT), 0.2 mM dNTPs, and 0.5 units of Taq polymerase (Promega, Madison, WI) cycled 35 times in a Rapidcycler 2 thermal cycler (Idaho Technology), 5 s at 94 °C, 10 s at 58 °C, 15 s at 72 °C, and slope = 9.9. A single product of the expected size was detected by electrophoresis using the DNA 1000 assay on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and the control reaction (containing no template) generated no PCR products, so primers and gDNA were sent to Lofstrand Labs (Gaithersburg MD) for probe creation and Northern blotting.

Total RNA extracts were sent in dry ice to Lofstrand Labs (Gaithersburg, Maryland, USA) to perform Northern blot experiments. The protocol for this procedure and experimental details are attached in Supplementary Fig. S3.

3. Results

3.1. Databases of plant orthologous introns

A database of orthologous introns of five plant species was generated using the same algorithms and pipelines of programs as described previously for mammalian orthologous introns (Rearick et al., 2011). Computation of plant intronic sequences for five species of land plants (*Arabidopsis*, poplar, grape, rice, and moss) produced 15,726 orthologous intron groups, while four species of angiosperms (*Arabidopsis*, poplar, grape, and rice) resulted in 28,519 groups. When the computation was confined to three species of dicotyledonous plants (*Arabidopsis*, poplar, and grape), 40,512 groups were obtained. Each group of orthologous introns was aligned with the Mafft program (Katoh et al., 2002) and the entire set of multiple alignments is available from our Exon–Intron Database web site.

An evolutionarily conserved intronic segment may represent alternatively spliced exon(s), non-coding RNA, functional regions involved in expression regulation (Fedorova and Fedorov, 2003), or so-called ultraconserved regions — 200 bp or longer segments which are interpreted to have functional properties like transcriptional enhancers and regulators of alternative splicing (Bejerano et al., 2004; Hudson et al., 2013). For computational finding of short ncRNAs (miRNA and endogenous siRNA), transcription factor binding sites, and alternative donor and acceptor splicing sites, we examined multiple alignments of plant orthologous introns for short regions (10–20 nucleotides) with strong sequence identity (≥80%). To identify alternatively skipped exons and snoRNAs, we searched for longer evolutionarily conserved regions (120 nucleotides) with less percentage of sequence identity (≥60%). Finally, for characterization of putative long ncRNAs, we

increased the length to 400 nucleotides and relaxed the strength of identity to 50%.

3.2. Conserved intronic regions in dicotyledonous plants

The number of conserved intronic sequences in dicots is three orders of magnitude fewer than that in placental mammals. Indeed, within 63,077 orthologous intron groups of five eutherian species (mouse, rat, dog, cow, and human) we previously characterized thousands of evolutionarily conserved intronic sequences including 9833 cases with long (>400 nts) conserved regions showing at least 50% identity (Rearick et al., 2011). In contrast, among 40,512 groups of orthologous introns of grape, poplar, and *Arabidopsis*, we found only two orthologous introns containing long (>400 nucleotides) evolutionarily conserved (≥50% identity) regions (Fig. 1). These two conserved intronic regions represent alternative exons described in the figure. We also

Fig. 1. Multiple sequence alignment of plant orthologous introns that have long evolutionarily conserved regions. Vv stands for *Vitis vinifera*; At — *Arabidopsis thaliana*; Pt — *Populus trichocarpa*. Identical nucleotides in the alignment are marked by stars beneath them. A. Sequences represent intron #6 of grape (exon/intron database id: 5548_NW_002238224); intron #6 of *Arabidopsis* (250A_NC_003070); and intron #7 of poplar (4186_NC_008468). Two regions highlighted in yellow correspond to two optionally-skipped exons present inside the reference sequence NM_001197976.1 of *Arabidopsis* and EST sequence EV021836.1 of *Brassica napus*. The first optional exon of *B. napus* contains a frameshift compared to the *Arabidopsis* counterpart (insertion of "A" at position 180 of *Arabidopsis* intron). Additionally, the first optional intron (shown in red) is absent in the following ESTs: EY747583.1, orange; JG606165.1, bleeding heart; EY779942.1, mandarin; EL387719.1, safflower; CK253964.1, potato; GW451284.1, coffee. Two optional exons and the intron between them are present as a single optional exon in bean (EG698013.1). Multiple indels inside optional exons result in frameshifts in different species and indicate that these alternative exons serve as nonsense-mediated decay (NMD) signals (Severing et al., 2009). B. Sequences represent intron #4 of grape (exon/intron database id: 11134_NW_002238109); intron #4 of *Arabidopsis* (8662D_NC_003074); and intron #3 of poplar (1455_NC_008467). Regions highlighted in yellow correspond to exonic sequences in a highly homologous (85% nucleotide identity) gene DEAD-box ATP-dependent RNA helicase 46 (NM_121465.1). No EST clones that have highlighted optional exons have been discovered.

examined the distribution of short ultra-conserved regions (20 invariable sequential nucleotides) within orthologous introns. Only seven dicot introns share such short ultra-conserved sequences (Fig. S4). In contrast, among 63,077 orthologous intron groups of 5 placental mammals, 3211 of them contain 20 nucleotide-long stretches of invariable nucleotides for all 5 species.

In order to estimate the impact of alternative splicing and ncRNAs on the conserved intronic regions in plants, we identified and examined the entire set of 25 groups of orthologous introns of *Arabidopsis*, grape and poplar, which contain conserved regions with at least 60% identical bases within a sequence window spanning over 120 nucleotides. This set of 25 orthologous groups is presented in Fig. S5. Sequences of these introns were compared with plant EST and ncRNA databases using the BLAST program (Altschul et al., 1990) in order to determine their possible functions. This examination, explained in Fig. S5, revealed the following: 1) Conserved regions within seventeen introns correspond to alternatively spliced exons. 2) Two correspond to ncRNAs (snoRNAs). 3) Three cases are likely *not* alternatively spliced exons because of numerous EST BLAST hits from corresponding genes with no matches against intronic regions. These three intronic sequences likely represent unknown ncRNAs or DNA/RNA functional regions (e.g. enhancers or silencers). 4) For the remaining three introns, we were unable to identify possible functions because of their poor representation in the EST database and no hits with ncRNA databases.

3.3. Conserved intronic regions in flowering plants

The number of conserved intronic regions even in a recent branch of plants (dicots) is relatively small. Therefore, in order to identify conserved regions in a broader group of plants, we applied computational algorithms with relaxed parameters for characterization of conserved regions. Specifically, we considered that an intron would contain an evolutionarily conserved region when it would have a stretch of ten bases containing at least nine invariable nucleotides for all studied species. Examination of orthologous introns of four flowering plants (including three dicots and one monocotyledonous plant — rice) revealed 52 groups that obeyed such sequence conservation criterion (invariant nts = 9; window size = 10 nts). Among these 52 cases, 18 correspond to internal intronic regions; 15 are short regions at the intron 5′ terminus; 18 are short regions at the intron 3′ terminus; and one case (case #51 in Fig. S6) contained a short conserved region at the 5′-terminus and in the middle of the same intron. Three cases #43, 48, and 50 from Fig. S6 with conserved 5′-intron termini contain the consensus for U12-spliceosomal introns RTATCCTTT, which are well known for their evolutionary conservation (Alioto, 2007). Among the aforementioned 34 cases of short conserved regions at the intron termini, 22 have additional evolutionarily conserved cryptic splicing sites within 10 nucleotides of the intron terminus (highlighted in yellow in Fig. S6). Examination of the plant EST database confirmed that cryptic sites likely have participated in splicing in 18 out of these 22 cases indicated in Fig. S6. The utilization of these cryptic sites for splicing results in short insertions in mRNAs that always cause a shift in the reading frame. Two of the most evolutionarily conserved cases with alternative donor and acceptor splicing sites are also described below (cases 5 and 6 for land plants in Fig. S7). This observation is consistent with previously published results of frequent small insertions/deletions in mRNAs due to alternative donor or acceptor splicing sites in close vicinity to the major site (Campbell et al., 2006; English et al., 2010). Finally, among 19 cases of short conserved sequences in the middle of introns of flowering plants, three cases are associated with alternative splicing.

3.4. Conserved intronic regions common for moss and angiosperms

Only seven out of 15,726 orthologous intron groups representing moss, rice, *Arabidopsis*, poplar, and grape have evolutionarily conserved regions characterized by the same computational filter — 9 invariant nucleotides within a 10-nucleotide long scanning window. The case with the longest and most conserved region is shown in Fig. 2, while the remaining six cases with considerable nucleotide conservation (invariant nts = 9; window size = 10 nts) are shown in Fig. S7. The conserved intronic sequence indicated in Fig. 2 does not match any functional ncRNA from public databases, as checked by online BLAST searching programs, nor does it correspond to alternative splicing events based on EST examination using NCBI online EST dataset. Online BLAST examination did not detect the presence of this conserved sequence in any of the genomes outside land plants including *Chlamydomonas reinhardtii*. We termed this region CPIR-1 (Conserved Plant Intronic Region 1) and investigated it in detail below. The first of the remaining six conserved sequences indicated in Fig. S7 (case 1) corresponds to a snoRNA (R104 of *Arabidopsis*). The next sequence in Fig. S7 (case 2) corresponds to an exonic sequence in a paralogous gene (this particular case is also described in Fig. 1B for three dicot species). Case 3 represents the known case of alternative splicing, in which short conserved region in the middle of intron 4 corresponds to alternative transcription initiation, represented by NM_001160830 mRNA of *Arabidopsis*. Case 4 represents short (10 nucleotides) conserved region in the middle of intron that might be transcription factor binding sites (e.g. enhancers or silencers), small ncRNAs, or alternative splicing. Finally, the last two sequences, case 5 and case 6 from Fig. S7 represent short (10 nucleotides) conserved regions at the 5′- and 3′-intron termini, respectively. These two terminal conserved intronic sequences contain cryptic splice sites within the fourth and seventh nucleotides from the intron ends (shown in red in Fig. S7). Plant EST examination confirmed that, indeed, for case 5 the cryptic site is used for alternative splicing in 31% of transcripts for a variety of flowering plant species. ESTs representing this cryptic splice site have a 4-nucleotide insertion that causes a shift in the reading frame. Among the 16 EST sequences representing the tetratricopeptide repeat-containing gene of the sequence for case 6, only a single EST (HO807763.1) utilizes the described cryptic acceptor splicing site. Activation of this cryptic site causes a 7-nucleotide-long insertion and thus, a shift in the reading frame. All in all, plants as diverse as mosses and angiosperms share only a few conserved intronic regions. In contrast, among vertebrates of the *Osteichthyes* taxon, the number of conserved intronic regions of similar sequence conservation is estimated to be in the hundreds.

Finally, no statistically significant sequence similarity was found between plant and mammalian intronic conserved regions except for a single case with a short 5′-terminal region of U12-type *Arabidopsis* intron (case #43 from Fig. S6, *Arabidopsis* INTRON_2_15034_NC_003075 that is identical to its human orthologous intron #4 of BRCC3 gene, UniGene: HS.558537). Evolutionarily conserved structures may be more prominent than conserved sequences in ncRNAs. However, the examination of RNA structure conservation requires manual individual approaches and has not been broadly conducted in this study.

Using the RepeatMasker program we demonstrated that neither evolutionarily conserved intronic sequence of *Arabidopsis* described above corresponds to a DNA repetitive element.

Functions of plant genes with intronic evolutionarily conserved sequences from Figs. S4–S6 have been identified via NCBI and Gene Ontology databases and described in Fig. S8. Statistically significant enrichment ($p < 0.001$) of genes involved in nucleic acid binding (17 genes with GO:0003676), RNA binding (15 genes with GO:0003723), nucleotide binding (14 genes with GO:0000166), positive regulation of transcription (10 genes with GO:0045893), mRNA splicing (7 genes with GO:0000398), RNA processing (7 genes with GO:006396) and RNA splicing (5 genes with GO:0008380) was confirmed using Monte-Carlo simulations.

3.5. Characterization of tRNA-like plant intronic putative ncRNA

The most evolutionarily Conserved Plant Intronic Region (named CPIR-1; Fig. 2) described above contains two blocks of conserved

Fig. 2. Multiple alignment of plant orthologous intron sequences within an RNA-polymerase III inhibitor gene containing the most evolutionarily conserved sequence (CPIR-1). Vv stands for *Vitis vinifera* (intron EID identifier: INTRON_6-640_NW_002238155); At — *Arabidopsis thaliana* (INTRON_6-16795_NC_003076); Pp — *Physcomitrella patens* (INTRON_5-5711_NW_001865287); Pt — *Populus trichocarpa* (INTRON_6-13465_NC_008476); Os — *Oryza sativa* (INTRON_6-9449_NC_008397). Two conserved regions are highlighted in blue and yellow. Three stable stem–loop structures inside the yellow region are shown in magenta, green, and red. These regions match the D-loop, anticodon loop, and T-loop motifs of tRNAs, respectively.

regions (highlighted in blue and yellow in Fig. 2) that are separated by a short 10–20 nucleotide-long linker. The sequence of CPIR-1 has been preserved inside the plant introns of an RNA-polymerase III inhibitor gene (GenBank accession number AT5G13240.1 from *A. thaliana*) for at least 420 million years (the divergence time between moss and flowering plants). This gene is represented by >500 copies in the EST database with BLAST e-values $<10^{-8}$. Therefore, the absence of hits of the CPIR-1 sequence with plant ESTs indicates that this conserved region is unlikely an alternative splicing exon but rather an unknown ncRNA. To decipher possible functions of this conserved region, we applied the online RNAfold program to find stable local 2D stem–loop structures inside the CPIR-1 sequence. Then we analyzed possible spatial conformations of these local RNA structures using the JAR3D web server (http://rna.bgsu.edu/jar3d) that compares the input sequence with all known RNA 3D motifs available in the RNA 3D Motif Atlas (http://rna.bgsu.edu/rna3dhub/motifs). This examination revealed that two regions (shown in magenta and red in Fig. 2) inside the conserved yellow block showed significant 3D-structural matches with the D-loop and T-loop of tRNAs, respectively. This result led us to analyze the folding of the yellow block alone, which revealed a secondary structure closely resembling that of tRNAs. Computer-predicted 2D structures of this RNA segment are shown in Fig. 3A–E for the five plants examined. In addition, we co-folded these five sequences together using the RNAalifold program from the Vienna RNA web server in order to characterize evolutionarily conserved RNA 2D structures. The output of this analysis is shown in Fig. 3G. This figure demonstrates that the evolutionarily conserved 2D-structure of the yellow CPIR-1 region is remarkably similar to tRNAs. Table S2 depicts characteristics common for all transfer RNAs and compares their presence in plant CPIR-1 sequences. Importantly, we observed eight compensatory substitutions in the putative amino acid and anticodon stems that maintain secondary structure of base paring but disrupt nucleotide sequence conservation. At the same time, one of the most important functional elements of tRNA, the anticodon loop, is not conserved and is significantly different among the five analyzed species. For this reason, the tRNAscan-SE online tool does not recognize the CPIR-1 sequence or score it as a tRNA. Thus, the CPIR-1 sequence is not expected to be a functional tRNA, but rather appears to be a tRNA-like ncRNA. BLAST comparison of the *Arabidopsis* CPIR-1 sequence against the entire set of 639 *Arabidopsis* tRNAs did not reveal any significant match. However, Clustal-Omega alignment of *Arabidopsis* CPIR-1 with *Arabidopsis* tRNAs revealed 56% identity of CPIR-1 to Proline tRNAs (Fig. S2). Thus, CPIR-1 might have originated from a Pro-tRNA sequence.

Intriguingly, this tRNA-like sequence is inside an intron of the RNA-polymerase III inhibitor gene. Since many tRNA molecules are transcribed by RNA-polymerase III using internal promoters inside tRNA sequences, we hypothesized that CPIR-1 may also be transcribed and this transcription may influence the expression of the RNA-polymerase III inhibitor gene. In order to test this hypothesis, we performed Northern blot analysis of total RNA extracts obtained from various *Arabidopsis* tissues and from whole plants grown under different propagation conditions (see Fig. 4). Because we did not know where CPIR would be found, we took as comprehensive an approach as possible. We examined organs from young seedlings and mature plants. When this work

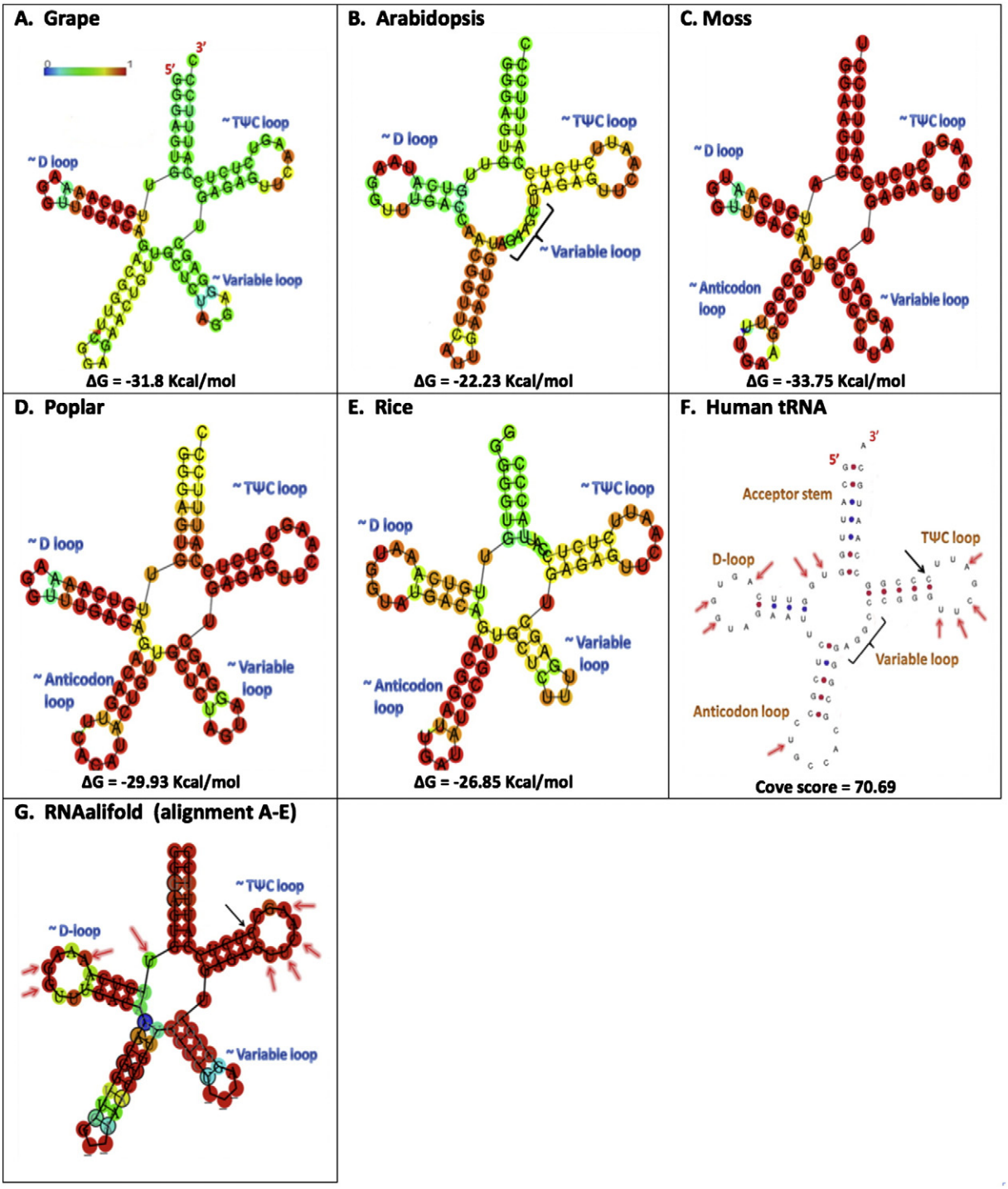

Fig. 3. Predicted structures of the evolutionarily conserved region of the CPIR-1 sequence shown in yellow in Fig. 2 for the five plant species: A) *Vitis vinifera*, B) *Arabidopsis thaliana*, C) *Physcomitrella patens*, D) *Populus trichocarpa*, and E) *Oryza sativa*. Structures A–E were obtained from the RNAfold server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) that scores RNA structures based on their thermodynamic stability values (ΔG) and colors them according to their base pairing probability, the scale of which varies from 0 to 1 as shown in the top left of panel A. The structures in panels A–E have tRNA-like features with respect to conserved loops that have been labeled with ~. Figure F shows the 2D structure of the human Gly-tRNA from intron 1 of the VAC14 gene. This structure was generated by tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/). A cove score (produced by tRNAscan-SE) greater than 20 is assigned to a sequence that the program predicts to be a real tRNA. The labels and arrows represent regions that are consonant with the features of real tRNAs described in the Results section. Panel G represents the consensus of secondary structure of the tRNA-like motif of the five plant sequences (A–E) obtained from RNAalifold with default parameters. Characteristics common to real tRNAs are shown by arrows.

was performed, we were unable to detect the presence of CPIR. We then reasoned that perhaps CPIR could be found in stressed plants, so we tested two common types of stress to which plants are exposed, namely reduced light and cold. As a probe, we used the entire sequence of the *Arabidopsis* intron containing CPIR-1 and the adjacent 5′- and 3′-exons. Thus, this probe should hybridize with RNA-pol III inhibitor mRNA and our putative tRNA-like ncRNA. The results in Fig. 4 clearly demonstrate that the RNA-pol III inhibitor gene is indeed expressed in each *Arabidopsis* tissue (roots, rosette leaves and inflorescence; panel A). Likewise, the RNA pol III transcript can be detected in plants propagated under a variety of growth conditions (dark treatment, cold treatment, cold treatment and then recovery from cold; panel B). However,

no hybridization signal with CPIR-1 in the low-molecular-weight region (20–200 nucleotides) was detected in any *Arabidopsis* tissue or in whole plants propagated under different growth conditions.

3.6. Intronic tRNAs of vertebrates

Since the association of tRNA with introns is not well reported in the literature, we performed large-scale bioinformatics examination of this issue in vertebrates. Computational analysis of GenBank feature-tables for the GRCh37.p5 release of the human genome demonstrated that 24 tRNAs are annotated inside the introns of 14 human protein-coding genes that have well-defined functions (the data are presented in Supplementary Table S2). If all 631 human tRNAs annotated in GenBank were randomly distributed inside the human genome, on average 157 tRNAs should be inside introns, which occupy ~25% of the genome. Hence, there is a strong avoidance of tRNAs to be within human introns. We also explored whether human intronic tRNAs are also present inside the orthologous introns of mouse and other species. This investigation showed that these tRNAs are absent in corresponding orthologous mouse introns for all genes except one, named *Vac14*. Reciprocally, several mouse genes have intronic tRNAs that are not present in the corresponding human orthologous genes. Further examination demonstrated that Vac14 has two Gly-tRNAs in intron 1 and two Gly-tRNAs in intron 9 in both human and mouse. These four intronic Gly-tRNAs exist in other mammalian species and in chicken, green anole, and zebrafish. However, *no* tRNA was found in Vac14 introns of more evolutionarily distant species such as jawless fish — lancelet *(Branchiostoma floridae*), sea squirt (*Ciona intestinalis*), or sea urchin (*Strongylocentrotus purpuratus*). All in all, a unique evolutionarily conserved association of tRNA with introns exists only in the *Vac14* gene of vertebrates and it has persisted for more than 500 million years (the time of separation of bony fish from tetrapods).

4. Discussion

Evolutionarily conserved intronic regions usually represent important functional elements such as non-coding RNAs, alternatively spliced exonic regions, and transcription regulatory elements (Fedorova and Fedorov, 2003). In this paper we demonstrated that the number of conserved regions in plant introns is two to three orders of magnitude fewer than that in vertebrates. Plant and vertebrate taxa for this comparison were chosen in such a way that they evolved during approximately the same period of time. However, the rate of nucleotide substitutions varies from species to species (Bromham, 2009; Gillooly et al., 2005). Recent deep-sequencing investigation demonstrated that *Arabidopsis* has a spontaneous mutation rate of 7×10^{-9} base per site per generation (Ossowski et al., 2010). On the other hand, Smith and Donoghue (2008) demonstrated that trees and shrubs were evolving approximately 2.7–10 times more slowly than related herbaceous plants. According to these authors, the majority of trees/shrubs have $0.5–1.5 \times 10^{-9}$ substitutions per site per year, while this parameter for herbs usually is in the range of $1–4 \times 10^{-9}$. Among the four flowering plants examined in this paper, two belong to trees/shrubs (poplar and grape) and two to herbs (*Arabidopsis* and rice). For comparison, the mutation rate in mammalian genomes is approximately 2.2×10^{-9} per base per year according to Kumar and Subramanian (2002). Taken together, even if we assume that plants have a slightly higher frequency of mutations in their genomes than mammals, this cannot explain the drastic difference in the number of conserved intronic regions between these two taxa. Therefore, our results indicate that either plants have much lower numbers of alternatively splicing isoforms and intronic ncRNAs, or that these sequences in plants are not under strong purifying selection pressure and, thus, evolve much faster than in vertebrates.

It is important to note that the average length of introns in the five analyzed plant species is considerably less than the average intron length of mammals and vertebrates. However, the length of introns does not correlate much with the presence of functional elements inside them. For example, all known functional snoRNA genes of humans (more than 200 genes) are located inside very short introns (mostly shorter than 4 kb) of highly expressed genes, which are functionally involved in ribosome biosynthesis or translation (Lestrade and Weber, 2006). The fugu fish represents another interesting example of the functional irrelevance of intron length. This species drastically shrunk its

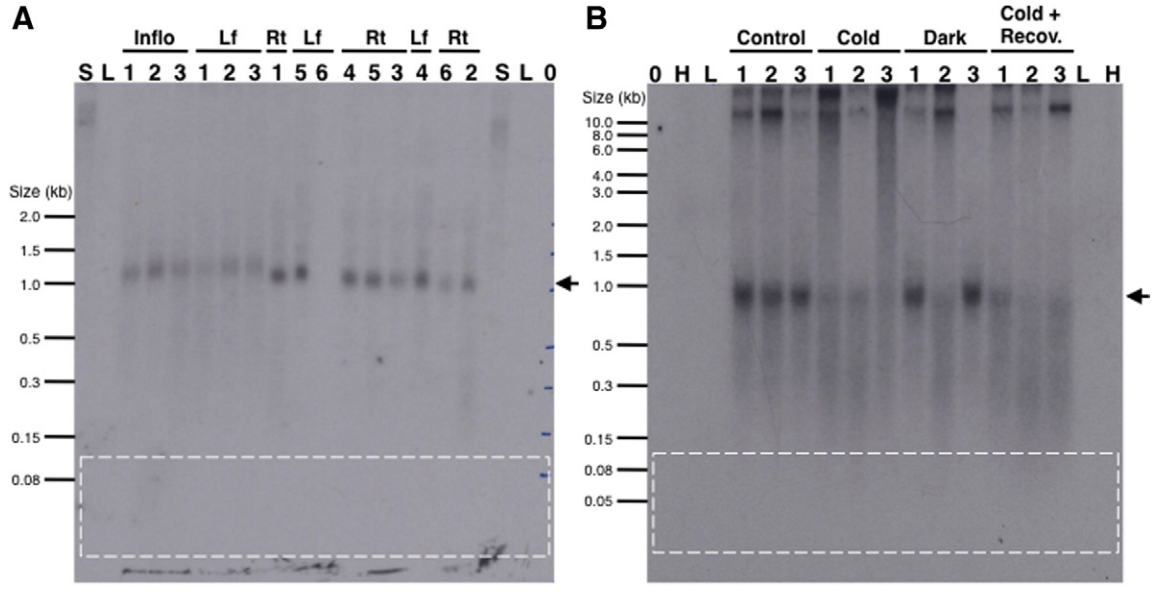

Fig. 4. Gel blot analysis of *Arabidopsis thaliana* total RNA from various plant organs (A), or under different growth conditions (B), probed with the exon6–intron7–exon7 region from *the RNA polymerase III inhibitor* gene. Size in kb of the markers is indicated on the left. Inflo, inflorescence tissue; Lf, rosette leaf tissue; Rt, root tissue; S or H, Invitrogen 0.5–10-kb RNA ladder; L, NEB Low Range ssRNA ladder; 0, no RNA control. Numbers for panel A indicate the plant pool of tissue from which the RNA was derived; organs with the same numbers are from the same pool of plants. Numbers for B indicate individual plants for each treatment. Treatments [e.g. cold, dark, and cold + recov. (cold treatment followed by a 24 h recovery period under control conditions)] are described in the Materials and methods section. Arrows indicate the size of normal transcripts and the dashed box indicates where the tRNA-like structure should be located if abundant.

genome by several times yet retained the vast majority of its introns, which are shortened relative to other species (Aparicio et al., 2002). A majority of conserved intronic regions in fugu are still present. For these reasons we did not normalize for intron length in comparing plants and animals and concentrated on the total number of conserved regions inside the introns of the plant species studied.

We demonstrated that at least three quarters of conserved intronic regions of dicots correspond to alternatively-spliced exonic sequences. In the majority of cases, these alternative exons have small insertions/deletions in different species causing shifts in the reading frame. This observation is in line with the data provided by Severing et al. (2009) regarding minimum involvement of evolutionarily conserved alternative splicing for the increase of proteome diversity in flowering plants. In addition, our results support the findings of Filichkin et al. (2010) that 78% of alternative splicing events in *Arabidopsis* create premature termination codons (PTC) and are likely involved in the regulation of expression via nonsense-mediated decay (NMD) and regulated unproductive splicing and translation (RUST) mechanisms.

Hundreds of thousands of small ncRNAs have been described for *Arabidopsis* (Mi et al., 2008), so it was a surprise to find only a few evolutionarily conserved intronic ncRNAs in this bioinformatics investigation. Nonetheless, we revealed the most evolutionarily conserved intronic sequence of higher plants (CPIR-1) and classified it as a putative tRNA-like ncRNA. This classification was based on eight compensatory mutations that preserve the predicted secondary structure of the putative "anticodon" and "acceptor" stems in the CPIR-1 sequence and due to the remarkable overall resemblance of multiple tRNA features including: conserved nucleotides, base-pairs, sizes of stems and loops, and sequence matches to known 3D structures of tRNA hairpins. We conjecture that this CPIR-1 sequence was originally a functional tRNA molecule that was located inside an intron. This location gave the tRNA an additional role in the regulation of the host gene. Eventually, the sequence became inactive as a tRNA due to mutations in the anticodon loop, yet the other function related to its intronic position was preserved. Unfortunately, in our first experimental attempt, we were unable to prove that CPIR-1 is a non-coding RNA. Interestingly, Wilusz, Sunwoo, and Spector have described MALAT-1, a long ncRNA (~7 kb) of mammals, which includes a 61-nt long tRNA-like small RNA, named mascRNA (Wilusz et al., 2009). Wilusz and co-authors demonstrated that mascRNA is produced via processing of the MALAT-1 nascent transcript and is rapidly degraded.

In mammals, some tRNA genes are present inside spliceosomal introns. For humans it happens with 3.8% of tRNAs. Usually, this tRNA–intron association is not evolutionarily maintained. However, there is an exception with four Gly-tRNAs inside the introns of the *Vac14* gene that have been present there for at least 500 million years. *Vac14* encodes a major scaffold protein within a complex that regulates phosphatidylinositol 3,5-biphosphate [PtdIns(3,5)P_2] levels across the animal kingdom (Davy and Robinson, 2003; Sbrissa et al., 2004). Recently discovered, PtdIns(3,5)P_2 is a ubiquitous eukaryote phosphoinositide of very low abundance. It has been proposed to have at least five independent functions including: recruitment of cytosolic proteins to define organelle specificity; functional regulation of endolysosomal membrane proteins; determination of physical properties and fusogenic potential of endolysosomal membranes; precursor for PI(3)P or PI(5)P; and modulation of endolysosomal pH (Dove et al., 2009). In mammals, PtdIns(3,5)P_2 supports and probably regulates retrograde membrane trafficking from lysosomal and late endosomal compartments to the Golgi complex. PtdIns(3,5)P_2 controls vesicle formation through highly specific activation of vesicle specific Ca^{2+} channels (Dong et al., 2010). *Vac14* is essential for mouse nervous system development (Zhang et al., 2007). Therefore, the exceptionally long-term association of Gly-tRNAs with Vac14 introns suggests that this tRNA–intron association might be fruitful and provide an additional source of regulation for the biologically important PtdIns(3,5)P_2 pathway.

5. Conclusions

We have shown that the number of conserved regions in plant introns is hundreds of times fewer than that in animal introns. A majority of plant intronic conserved regions with characterized functions represent alternatively-spliced sequences. Only a handful of conserved intronic ncRNAs of flowering plants have been registered. We found one conserved sequence with conserved secondary structure and apparent similarity to tRNA, but with enough differences to rule out its possibility of being a real functional tRNA.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2014.07.012.

Authors' contributions

EA, LW, ASM, and AP performed computational experiments and analyzed the data. SK, EC, and SL performed wet lab experiments with *Arabidopsis* RNA and DNA. CLZ performed 2D and 3D computational analysis of potential ncRNAs. AF and LF designed and supervised the study and paper writing.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by National Science Foundation Career award "Investigation of intron cellular roles" (grant number MCB-0643542) to AF; by USDA-ARS Specific Cooperative Agreement (58-3607-1-193) to SL; and National Institutes of Health (grant number 1R01GM085328-01A1) to CLZ.

References

Alioto, T.S., 2007. U12DB: a database of orthologous U12-type spliceosomal introns. Nucleic Acids Res. 35, D110–D115.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., et al., 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 297, 1301–1310.
Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.
Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., et al., 2004. Ultraconserved elements in the human genome. Science 304, 1321–1325.
Berardini, T.Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., et al., 2004. Functional annotation of the *Arabidopsis* genome using controlled vocabularies. Plant Physiol. 135, 745–755.
Bromham, L., 2009. Why do species vary in their rate of molecular evolution? Biol. Lett. 5, 401–404.
Bu, D., Yu, K., Sun, S., Xie, C., Skogerbo, G., et al., 2012. NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res. 40, D210–D215.
Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M., Buell, C.R., 2006. Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. BMC Genomics 7, 327.
Chan, P.P., Lowe, T.M., 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. 37, D93–D97.
Chernikova, D., Motamedi, S., Csuros, M., Koonin, E.V., Rogozin, I.B., 2011. A late origin of the extant eukaryotic diversity: divergence time estimates using rare genomic changes. Biol. Direct 6, 26.
Davy, B.E., Robinson, M.L., 2003. Congenital hydrocephalus in hy3 mice is caused by a frameshift mutation in Hydin, a large novel gene. Hum. Mol. Genet. 12, 1163–1170.
Dong, X.P., Shen, D., Wang, X., Dawson, T., Li, X., et al., 2010. PI(3,5)P_2 controls membrane trafficking by direct activation of mucolipin Ca^{2+} release channels in the endolysosome. Nat. Commun. 1, 38.
Dove, S.K., Dong, K., Kobayashi, T., Williams, F.K., Michell, R.H., 2009. Phosphatidylinositol 3,5-bisphosphate and Fab1p/PIKfyve underPPIn endo-lysosome function. Biochem. J. 419, 1–13.
English, A.C., Patel, K.S., Loraine, A.E., 2010. Prevalence of alternative splicing choices in *Arabidopsis thaliana*. BMC Plant Biol. 10, 102.
Fedorov, A., Stombaugh, J., Harr, M.W., Yu, S., Nasalean, L., et al., 2005. Computer identification of snoRNA genes using a mammalian orthologous intron database. Nucleic Acids Res. 33, 4578–4583.
Fedorova, L., Fedorov, A., 2003. Introns in gene evolution. Genetica 118, 123–131.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., et al., 2010. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Res. 20, 45–58.
Gillooly, J.F., Allen, A.P., West, G.B., Brown, J.H., 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. Proc. Natl. Acad. Sci. U. S. A. 102, 140–145.
Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., et al., 2010. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 38, W695–W699.
Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuboock, R., Hofacker, I.L., 2008. The Vienna RNA Websuite. Nucleic Acids Res. 36, W70–W74.
Hofacker, I.L., Fekete, M., Stadler, P.F., 2002. Secondary structure prediction for aligned RNA sequences. J. Mol. Biol. 319, 1059–1066.
Hudson, R.S., Yi, M., Volfovsky, N., Prueitt, R.L., Esposito, D., et al., 2013. Transcription signatures encoded by ultraconserved genomic regions in human prostate cancer. Mol. Cancer 12, 13.
Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.
Kin, T., Yamada, K., Terai, G., Okida, H., Yoshinari, Y., et al., 2007. fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. Nucleic Acids Res. 35, D145–D148.
Kumar, S., Subramanian, S., 2002. Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. U. S. A. 99, 803–808.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., et al., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Lestrade, L., Weber, M.J., 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic Acids Res. 34, D158–D162.
Li, J., Leisner, S.M., Frantz, J., 2008. Alleviation of copper toxicity in *Arabidopsis thaliana* by silicon addition to hydroponic solutions. J. Am. Soc. Hortic. Sci. 133, 670–677.
Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.
Mattick, J.S., 1994. Introns: evolution and function. Curr. Opin. Genet. Dev. 4, 823–831.
McCaskill, J.S., 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers 29, 1105–1119.
Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., et al., 2008. Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5′ terminal nucleotide. Cell 133, 116–127.
Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. U. S. A. 104, 19363–19368.
Morello, L., Breviario, D., 2008. Plant spliceosomal introns: not only cut and paste. Curr. Genomics 9, 227–238.
Ossowski, S., Schneeberger, K., Lucas-Lledo, J.I., Warthmann, N., Clark, R.M., et al., 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327, 92–94.
Page, R.D., 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.
Rearick, D., Prakash, A., McSweeny, A., Shepard, S.S., Fedorova, L., et al., 2011. Critical association of ncRNA with introns. Nucleic Acids Res. 39, 2357–2366.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Sarver, M., Zirbel, C.L., Stombaugh, J., Mokdad, A., Leontis, N.B., 2008. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. J. Math. Biol. 56, 215–252.
Sbrissa, D., Ikonomov, O.C., Strakova, J., Dondapati, R., Mlak, K., et al., 2004. A mammalian ortholog of *Saccharomyces cerevisiae* Vac14 that associates with and up-regulates PIKfyve phosphoinositide 5-kinase activity. Mol. Cell. Biol. 24, 10437–10447.
Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, W686–W689.
Severing, E.I., van Dijk, A.D., Stiekema, W.J., van Ham, R.C., 2009. Comparative analysis indicates that alternative splicing in plants has a limited role in functional expansion of the proteome. BMC Genomics 10, 154.
Shepelev, V., Fedorov, A., 2006. Advances in the Exon–Intron Database (EID). Brief. Bioinform. 7, 178–185.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.
Sironi, M., Menozzi, G., Comi, G.P., Cagliani, R., Bresolin, N., et al., 2005. Analysis of intronic conserved elements indicates that functional complexity might represent a major source of negative selection on non-coding sequences. Hum. Mol. Genet. 14, 2533–2546.
Smith, S.A., Donoghue, M.J., 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322, 86–89.
Venditti, C., Meade, A., Pagel, M., 2011. Multiple routes to mammalian diversity. Nature 479, 393–396.
Wang, Y., You, F.M., Lazo, G.R., Luo, M.C., Thilmony, R., et al., 2013. PIECE: a database for plant gene structure comparison and evolution. Nucleic Acids Res. 41, D1159–D1166.
Wilkerson, M.D., Ru, Y., Brendel, V.P., 2009. Common introns within orthologous genes: software and application to plants. Brief. Bioinform. 10, 631–644.
Wilusz, J.E., Sunwoo, H., Spector, D.L., 2009. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 23, 1494–1504.
Zhang, Y., Zolov, S.N., Chow, C.Y., Slutsky, S.G., Richardson, S.C., et al., 2007. Loss of Vac14, a regulator of the signaling lipid phosphatidylinositol 3,5-bisphosphate, results in neurodegeneration in mice. Proc. Natl. Acad. Sci. U. S. A. 104, 17518–17523.
Zuker, M., Stiegler, P., 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 9, 133–148.