Exhaustive identification of conserved upstream open reading frames with potential translational regulatory functions from animal genomes

Hiro Takahashi[1,2#]*, Shido Miyaki[2#], Hitoshi Onouchi[3#], Taichiro Motomura[1], Nobuo Idesako[2], Anna Takahashi[4], Shuichi Fukuyoshi[5], Toshinori Endo[6], Kenji Satou[7] , Satoshi Naito[3,8], and Motoyuki Itoh[9]*

[1]Graduate School of Medical Sciences, Kanazawa University, Kanazawa 920-1192, Japan

[2]Graduate School of Horticulture, Chiba University, Matsudo 271-8510, Japan

[3]Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

[4]Faculty of Information Technologies and Control, Belarusian State University of Informatics and Radio Electronics, Minsk 220013, Belarus

[5]Institute of Medical, Pharmaceutical and Health Sciences, Kanazawa University, Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan

[6]Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan

[7]Faculty of Biological Science and Technology, Institute of Science and Engineering, Kanazawa University, Kanazawa 920-1192, Japan

[8]Graduate School of Life Science, Hokkaido University, Sapporo 060-0810, Japan

[9]Graduate School of Pharmaceutical Science, Chiba University, Chuo-ku, Chiba 260-8675, Japan

*Correspondence. Tel: +81-76-234-4484; Fax: +81-76-234-4484; Email: takahasi@p.kanazawa-u.ac.jp

Correspondence may also be addressed to Motoyuki Itoh. Email: mito@chiba-u.jp

[#]Joint first authors.

Key words: upstream open reading frame; translational regulation; bioinformatics; nascent peptide

Abstract

Background: Upstream open reading frames (uORFs) are located in the 5′-untranslated regions of many eukaryotic mRNAs, and some peptides encoded in these regions play important regulatory roles in controlling main ORF (mORF) translation. To comprehensively identify uORFs encoding functional peptides, genome-wide searches for uORFs with conserved peptide sequences (CPuORFs) have been conducted in various organisms using comparative genomic approaches. However, in animals, CPuORFs have been identified only by comparing uORF sequences between a limited number of closely related species, and it is unclear how many previously identified CPuORFs encode regulatory peptides.

Results: Here, we conducted exhaustive genome-wide searches for animal CPuORFs conserved in various taxonomic ranges, using the ESUCA pipeline, which we recently developed for efficient comprehensive identification of CPuORFs. ESUCA can efficiently compare uORF sequences between an unlimited number of species using BLAST and automatically determine the taxonomic ranges of sequence conservation for each CPuORF. By applying ESUCA to human, chicken, zebrafish, and fruit fly genomes, 1,430 (1,339 novel and 91 known) CPuORFs were identified. We examined the effects of 14 human CPuORFs on mORF translation using a transient expression assay. Through this analysis, we identified six novel regulatory CPuORFs that repressed mORF translation in a sequence-dependent manner, all of which were conserved beyond Amniota.

Conclusions: We discovered a much higher number of animal CPuORFs than previously identified. Furthermore, our results suggest that human CPuORFs conserved beyond Amniota are more likely to encode regulatory peptides than those conserved in narrower taxonomic ranges.

Background

The human genome contains many regions encoding potential functional small peptides outside of the well-annotated protein-coding regions (Ingolia, et al., 2014). Some upstream open reading frames (uORFs), which are located in the 5′-untranslated regions (5′-UTRs) of mRNAs, have been shown to encode such functional small peptides. Most uORF-encoded peptides play regulatory roles in controlling the translation of protein-coding main ORFs (mORFs) (Cruz-Vera, et al., 2011; Ito and Chiba, 2013; Morris and Geballe, 2000; Somers, et al., 2013). During the translation of these regulatory uORFs, nascent peptides interact inside the ribosomal exit tunnel to cause ribosome stalling (Bhushan, et al., 2010). Ribosome stalling on a uORF results in translational repression of the downstream mORF because stalled ribosomes block scanning of subsequent pre-initiation complexes and prevent them from reaching the start codon of the mORF (Wang and Sachs, 1997). In some genes, uORF peptides are involved in translational regulation in response to metabolites (Ito and Chiba, 2013).

To comprehensively identify uORFs encoding functional peptides, genome-wide searches for uORFs with conserved peptide sequences (CPuORFs) have been conducted using comparative genomic approaches in plants (Hayden and Jorgensen, 2007; Takahashi, et al., 2019; Takahashi, et al., 2012; Tran, et al., 2008; van der Horst, et al., 2018; Vaughn, et al., 2012). To date, 157 CPuORF families have been identified by comparing 5′-UTR sequences between plant species. Of these, 101 families were identified in our previous studies by applying our original methods, BAIUCAS (Takahashi, et al., 2012) and ESUCA (an advanced version of BAIUCAS) (Takahashi, et al., 2019) to genomes of *Arabidopsis*, rice, tomato, poplar, and grape.

ESUCA has many unique functions (Takahashi, et al., 2019), such as efficient comparison of uORF sequences between an unlimited number of species using BLAST, automatic determination of taxonomic ranges of CPuORF sequence conservation, systematic calculation of K_a/K_s ratios of CPuORF sequences, and wide compatibility with any eukaryotic genome whose sequence database is registered in ENSEMBL (Zerbino, et al., 2018). More importantly, to distinguish between ‘spurious’ CPuORFs conserved because they encode parts of mORF-encoded proteins and ‘true’ CPuORFs conserved because of the functions of their encoded small peptides, ESUCA assesses whether a transcript containing a fusion of a uORF and an mORF is a major or minor form among homologous transcripts (Takahashi, et al., 2019). By using these functions, ESUCA can efficiently identify CPuORFs likely to encode functional small peptides. In fact, our recent study demonstrated that poplar CPuORFs encoding regulatory peptides were efficiently identified by selecting ones conserved across diverse eudicots using ESUCA (Takahashi, et al., 2019).

To date, only a few studies on genome-wide identification of animal CPuORFs have reported. In these previous studies, uORF sequences were compared between a limited number of closely related species, such as human and mouse or several species in dipteran, leading to identification of 204 and 198 CPuORFs in human and mouse, respectively (Crowe, et al., 2006), and 44 CPuORFs in fruit fly (Hayden and Bosco, 2008). Additionally, the relationships between taxonomic ranges of CPuORF conservation and the likelihood of having a regulatory function have not been studied in animals.

Accordingly, in this study, we applied ESUCA to genomes of fruit fly, zebrafish, chiken, and human to exhaustively identify animal CPuORFs and to determine the taxonomic range of their sequence conservation. Using ESUCA, we identified 1,430 animal (1,339 novel and 91 known) CPuORFs belonging to 1,337 CPuORF families. We examined the effects of 15 CPuORFs conserved in various taxonomic ranges on mORF translation, using a transient expression assay. Through this analysis, we identified six novel regulatory CPuORFs that repress mORF translation in a sequence-dependent manner. All of the six regulatory CPuORFs are conserved beyond Amniota, suggesting that human CPuORFs conserved beyond Amniota are more likely to encode functional peptides than those conserved in narrower taxonomic ranges.

Materials and methods

Extraction of CPuORFs using ESUCA

ESUCA was developed as an advanced version of BAIUCAS (Takahashi et al., 2012) in our previous study (Takahashi et al., 2019). ESUCA consists of six steps, and some of these steps are divided into substeps, as shown in Fig. 1A and 1B. To identify animal CPuORFs using ESUCA, the following eight-step procedures were conducted, including the six ESUCA steps: 0) data preparation for ESUCA, 1) uORF extraction from the 5′-UTR,

2) calculation of uORF-mORF fusion ratios, 3) uORF-tBLASTn against transcript sequence databases, 4) mORF-tBLASTn against downstream sequence datasets for each uORF, 5) calculation of K_a/K_s ratios, 6) determination of the taxonomic range of uORF sequence conservation, and 7) manual validation after ESUCA. See the Supplementary Materials and Methods for details.

Determination of the taxonomic range of uORF sequence conservation for animal CPuORFs

To apply ESUCA to animal genomes, we defined 19 animal taxonomic categories, as shown in Fig. 1C. See the Supplementary Materials and Methods for details.

Plasmids and reporter assays

DNA fragments containing a control CPuORF (Con) or the frameshift mutant version (fs) of the 15 selected genes were subcloned into pSV40:UTR:Fluc. Reporter assays were conducted using a Dual-Luciferase Reporter Assay system (Promega, Madison, WI, USA). See the Supplementary Materials and Methods for details.

Results

Genome-wide search for animal CPuORFs using ESUCA

Prior to ESUCA application (Fig. 1A and 1B), we counted the number of protein-coding genes for four species, i.e., fruit fly, zebrafish, chiken, and human. As shown in Supplementary Table S1, 13,938, 25,206, 14,697, and 19,956 genes were extracted for fruit fly, zebrafish, chiken, and human, respectively. After step 1 of ESUCA, we calculated the numbers of uORFs and protein-coding genes with any uORF for each species. As shown in Supplementary Table S1, 17,035 (7,066), 39,616 (14,453), 8,929 (3,535), and 44,085 (12,321) uORFs (genes) were extracted from fruit fly, zebrafish, chicken, and human genomes, respectively. In this analysis, when multiple uORFs from a gene shared the same stop or start codon, they were counted as one. Potential candidate CPuORFs were narrowed down by selection at step 2.5 of ESUCA in a step-by-step manner, as shown in Supplementary Table S1. The numbers of BLAST hits (expressed sequence tag [EST], transcriptome shotgun assembly [TSA], assembled EST/TSA, and RefSeq RNA sequences) extracted at step 3.5 are also shown in

Supplementary Table S1. After the final step of ESUCA, 49, 195, 235, and 1,453 candidate CPuORFs were extracted from fruit fly, zebrafish, chiken, and human, respectively. We conducted manual validation for the extracted candidate CPuORFs as described in our previous study (Takahashi, et al., 2019) and in the Supplementary Materials and Methods. We selected CPuORFs conserved in at least two orders other than the order to which the original species belongs. In total, 1,430 animal CPuORFs (35 for fruit fly, 151 for zebrafish, 206 for chicken, and 1,038 for human) were identified (Fig. 1D). Of these, 1,339 CPuORFs were newly identified in the current study. Detailed information on the identified CPuORFs is shown in Supplementary Table S2. The identified CPuORF-containing genes were classified into 1,124 ortholog groups on the basis of similarities of mORF-encoded amino acid sequences, using OrthoFinder ver. 1.1.4 (Emms and Kelly, 2015). CPuORFs with similar amino acid sequences from the same ortholog groups were categorized as the same CPuORF families (homology groups [HGs]; Supplementary Materials and Methods). The identified 1,430 CPuORFs were classified into 1,337 HGs. We assigned HG numbers to 1,337 HGs in an order based on numbers of orders in which any CPuORF belonging to each HG was extracted, the taxonomic range of the sequence conservation of each HG, and gene ID numbers. When multiple CPuORF families were identified in the same ortholog groups, the same HG number with a different subnumber was assigned to each of the families.

Sequence-dependent effects of CPuORFs on mORF translation

To address the relationship between taxonomic ranges of CPuORF conservation and likelihood of having regulatory function, we selected 15 human CPuORFs conserved in various taxonomic ranges, including a previously identified sequence-dependent regulatory CPuORF, the *PTP4A1* CPuORF (Hardy, et al., 2019), as a positive control, and examined their sequence-dependent effects on the expression of the downstream reporter gene using transient expression assays (Fig. 2). Other uORFs overlapping any of the selected CPuORFs were eliminated by introducing mutations that changed the ATG codons of the overlapping uORFs to other codons but did not alter the amino acid sequences of the CPuORFs (Supplementary Fig. S5). The resulting modified CPuORFs were used as Con CPuORFs (Fig. 2B). To assess the importance of amino acid sequences for the effects of these CPuORFs on mORF translation, fs mutations were introduced into the Con CPuORFs such that the amino acid sequences of their conserved regions could be altered (see Supplementary Materials and Methods and Supplementary Fig. S1 for details). In seven of the 15 CPuORFs, the introduced frameshift mutations significantly upregulated the expression of the reporter gene, indicating that these CPuORFs repressed mORF translation in a sequence-dependent manner (Fig. 2C). All of the seven CPuORFs with the sequence-dependent repressive effects were conserved beyond Amniota (Fig. 2A). In contrast, any of four CPuORFs conserved only among Amniota did not show significant sequence-dependent effects (Fig. 2C). These results suggest that human CPuORFs conserved beyond Amniota are more likely to encode regulatory peptides than those conserved in narrower taxonomic ranges.

Discussion

In the current study, by applying ESUCA to four animal genomes, we identified 1,430 CPuORFs belonging to 1,337 HGs. Taxonomic ranges of sequence conservation of these CPuORFs largely vary, demonstrating that ESUCA can identify CPuORFs conserved in various taxonomic ranges (Supplementary Table S3). Moreover, seven of 11 human CPuORFs conserved beyond Amniota exhibited sequence-dependent repressive effects on mORF translation, whereas all four CPuORFs conserved only among Amniota showed no significant sequence-dependent effects. This result suggest that human CPuORFs conserved beyond Amniota are more likely to encode regulatory peptides than those conserved in narrower taxonomic ranges. Of the 1,038 CPuORFs extracted from the human genome, 78 are conserved beyond Amniota (Supplementary Table S3). Therefore, these 78 CPuORFs are promising candidates of regulatory CPuORFs encoding peptides that control mORF translation. CPuORFs encoding functional peptides may also be found among the remaining human CPuORFs conserved in narrower taxonomic ranges because the K_a/K_s ratios suggest that purifying selection acted on the amino acid sequences of these CPuORFs.

In this study, we identified six novel human regulatory CPuORFs (in the *MKKS*, *SLC6A8*, *FAM13B*, *MIEF1*, *KAT6A*, and *LRRC8B* genes) with sequence-dependent repressive effects on mORF translation. Of these, the *MKKS* CPuORF is a translational regulator that represses the production of a protein involved in McKusick-Kaufman syndrome (Akimoto, et al., 2013); however, the amino acid sequence dependence of the

CPuORF function was not reported. Interestingly, the *MIEF1* CPuORF-encoded peptide is a functional peptide localized in the mitochondria (Samandi, et al., 2017). Thus, the *MIEF1* CPuORF may have dual functions.

Chemical screening recently identified a compound that causes nascent peptide-mediated ribosome stalling in the mORF of the human *PCSK9* gene, resulting in specific translational inhibition of *PCSK9* and a reduction in total plasma cholesterol levels (Lintner, et al., 2017). Nascent peptide-mediated ribosome stalling in some of the previously identified regulatory CPuORFs is promoted by metabolites, such as polyamine, arginine, and sucrose (Ito and Chiba, 2013; Yamashita, et al., 2017). Therefore, compounds that promote nascent peptide-mediated ribosome stalling in CPuORFs could be identified by chemical screening through a method similar to that used for the screening of the stall-inducing compound for *PCSK9*. The data from the current study may be useful for selection of CPuORFs as potential targets for pharmaceutical drugs and for identification of regulatory CPuORFs.

Funding: This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant nos. JP16K07387 to H.O., JP18H03330 to H.T. and H.O., JP18H02568 to M.I.); the Ministry of Education, Culture, Sports, Science and Technology (MEXT) KAKENHI (grant nos. JP17H05658 to S.N., JP26114703 to H.T, JP17H05659 to H.T); and the Naito Foundation (to H.O.).

Competing interests: none declared.

References

Akimoto, C., *et al.* (2013) Translational repression of the McKusick-Kaufman syndrome transcript by unique upstream open reading frames encoding mitochondrial proteins with alternative polyadenylation sites, *Biochim Biophys Acta*, **1830**, 2728-2738.

Bhushan, S., *et al.* (2010) Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide, *Mol. Cell*, **40**, 138-146.

Crowe, M.L., Wang, X.Q. and Rothnagel, J.A. (2006) Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides, *BMC Genomics*, **7**, 16.

Cruz-Vera, L.R., *et al.* (2011) Nascent polypeptide sequences that influence ribosome function, *Curr. Opin. Microbiol.*, **14**, 160-166.

Emms, D.M. and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy, *Genome Biol.*, **16**, 157.

Hardy, S., *et al.* (2019) Magnesium-sensitive upstream ORF controls PRL phosphatase expression to mediate energy metabolism, *Proc Natl Acad Sci U S A*, **116**, 2925-2934.

Hayden, C.A. and Bosco, G. (2008) Comparative genomic analysis of novel conserved peptide upstream open reading frames in *Drosophila melanogaster* and other dipteran species, *BMC Genomics*, **9**, 61.

Hayden, C.A. and Jorgensen, R.A. (2007) Identification of novel conserved peptide uORF homology groups in *Arabidopsis* and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes, *BMC Biol.*, **5**, 32.

Ingolia, N.T., *et al.* (2014) Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes, *Cell Rep.*, **8**, 1365-1379.

Ito, K. and Chiba, S. (2013) Arrest peptides: cis-acting modulators of translation, *Annu. Rev. Biochem.*, **82**, 171-202.

Lintner, N.G., *et al.* (2017) Selective stalling of human translation through small-molecule engagement of the ribosome nascent chain, *PLoS Biol*, **15**, e2001882.

Morris, D.R. and Geballe, A.P. (2000) Upstream open reading frames as regulators of mRNA translation, *Mol. Cell Biol.*, **20**, 8635-8642.

Samandi, S., *et al.* (2017) Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins, *Elife*, **6**.

Somers, J., Poyry, T. and Willis, A.E. (2013) A perspective on mammalian upstream open reading frame function, *Int. J. Biochem. Cell Biol.*, **45**, 1690-1700.

Takahashi, H., *et al.* (2019) ESUCA: a pipeline for genome-wide identification of upstream open reading frames with evolutionarily conserved sequences and determination of the taxonomic range of their conservation, *bioRxiv*, 524090.

Takahashi, H., *et al.* (2012) BAIUCAS: a novel BLAST-based algorithm for the identification of upstream open reading frames with conserved amino acid sequences and its application to the *Arabidopsis thaliana* genome, *Bioinformatics*, **28**, 2231-2241.

Tran, M.K., Schultz, C.J. and Baumann, U. (2008) Conserved upstream open reading frames in higher plants, *BMC Genomics*, **9**, 361.

van der Horst, S., *et al.* (2018) Novel pipeline identifies new upstream ORFs and non-AUG initiating main ORFs with conserved amino acid sequences in the 5' leader of mRNAs in *Arabidopsis thaliana*, *RNA*, **25**, 292-304.

Vaughn, J.N., *et al.* (2012) Known and novel post-transcriptional regulatory sequences are conserved across plant families, *RNA*, **18**, 368-384.

Wang, Z. and Sachs, M.S. (1997) Ribosome stalling is responsible for arginine-specific translational attenuation in *Neurospora crassa*, *Mol. Cell Biol.*, **17**, 4904-4913.

Yamashita, Y., *et al.* (2017) Sucrose sensing through nascent peptide-meditated ribosome stalling at the stop codon of Arabidopsis *bZIP11* uORF2, *FEBS Lett.*, **591**, 1266-1277.

Zerbino, D.R., *et al.* (2018) Ensembl 2018, *Nucleic Acids Res.*, **46**, D754-D761.

Figure Legends

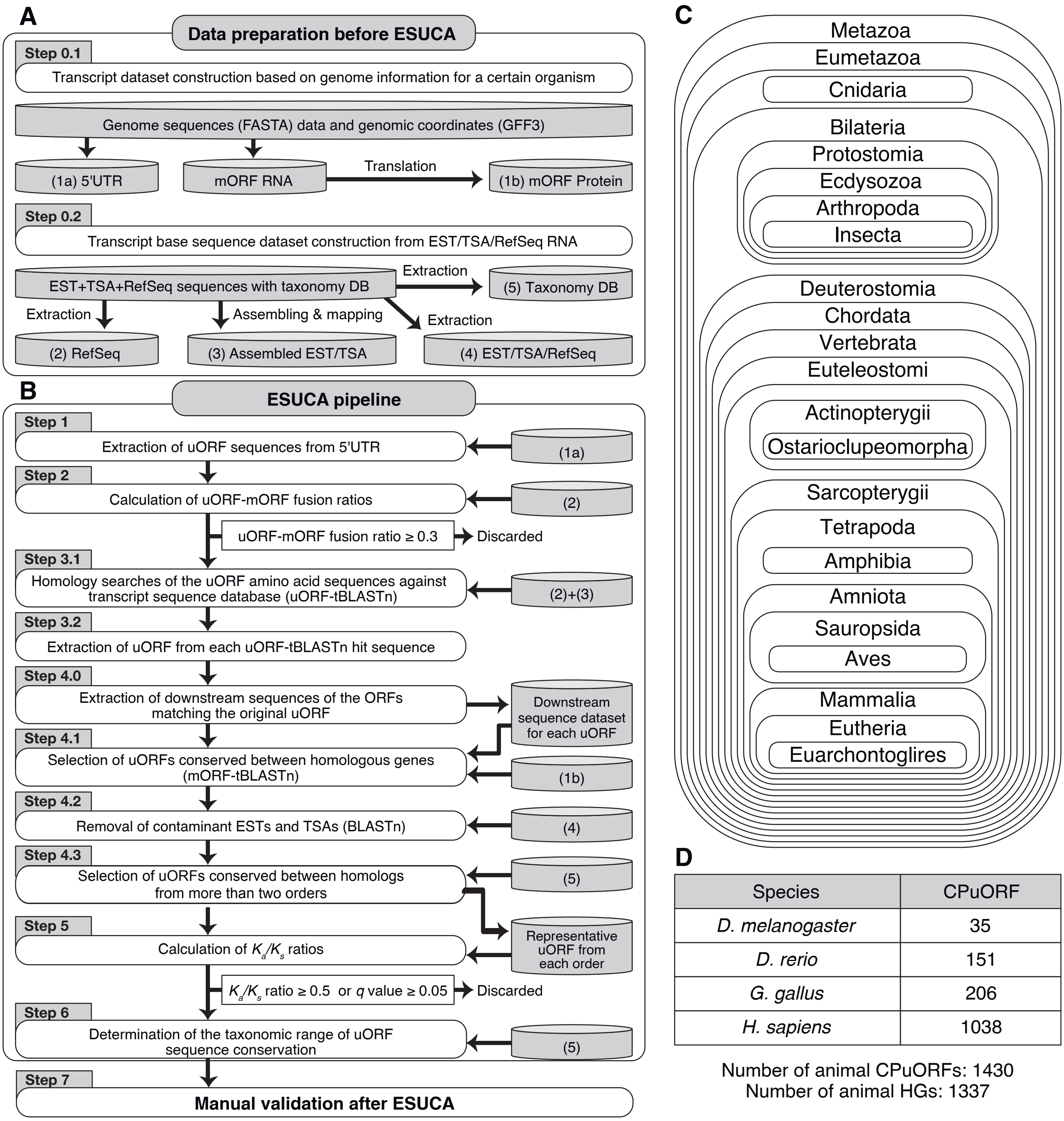
Figure 1. Identification of animal CPuORFs using ESUCA. (A) Data preparation. (B) Outline of the ESUCA pipeline. Numbers with parenthesis indicate datasets labeled with the same numbers in A. (C) Defined animal taxonomic categories. (D) Numbers of identified CPuORFs.

Figure 2. Taxonomic conservation and experimental validation of 15 selected human CPuORFs. (A) Taxonomic ranges of conservation of CPuORFs examined in transient assays. Filled cells in each taxonomic category indicate the presence of uORF-tBLASTn and mORF-tBLASTn hits for CPuORFs of the indicated genes. (B) Reporter constructs used for transient assays. The hatched box in the frameshift (fs) mutant CPuORF indicates the frame-shifted region. Dotted boxes represent the first five nucleotides of the mORFs associated with the 15 human CPuORFs. (C) Relative luciferase activities of control (white) or frameshift (gray) CPuORF reporter plasmids. Means ± SDs of three biological replicates are shown. *$p < 0.05$.

Supplementary Figure Legend

Figure S1. Nucleotide sequences of the 5′-UTRs and amino acid sequences of the CPuORFs analyzed in this study. (AO) The 5′-UTRs of *PTP4A1* (A), *MKKS* (B), *SLC6A8* (C), *FAM13B* (D), *MIEF1* (E), *EIF5* (F), *MAPK6* (G), *MEIS2* (H), *KAT6A* (I), *SLC35A4* (J), *LRRC8B* (K), *CDH11* (L), *PNRC2* (M), *BACH2* (N), and *FGF9* (O). The nucleotide sequences of the CPuORFs are shown in bold. The deduced amino acid sequences of the control (Con) and frameshift (fs) CPuORFs are indicated. The nucleotide sequences of other uORFs are underlined with a bold line. Dotted underlines indicate the nucleotide sequences of other uORFs overlapping the CPuORFs, whose initiation codons were altered to other codons by introducing nucleotide substitutions that did not change the amino acid sequences of the CPuORFs. The replaced nucleotides are shown as white letters in a black background. The nucleotides that were deleted and inserted in the frameshift mutants are shaded. The main coding sequences that were contained in the reporter constructs are boxed. The shaded boxes indicate the nucleotides changed to avoid the appearance of in-frame termination codons. Cloning sites were added at either end of the nucleotide sequences in controls to be subcloned into plasmid pGL4.10 with an SV40 promoter (pSV40:5'UTR::luc2) are underlined (Fig. 2B).

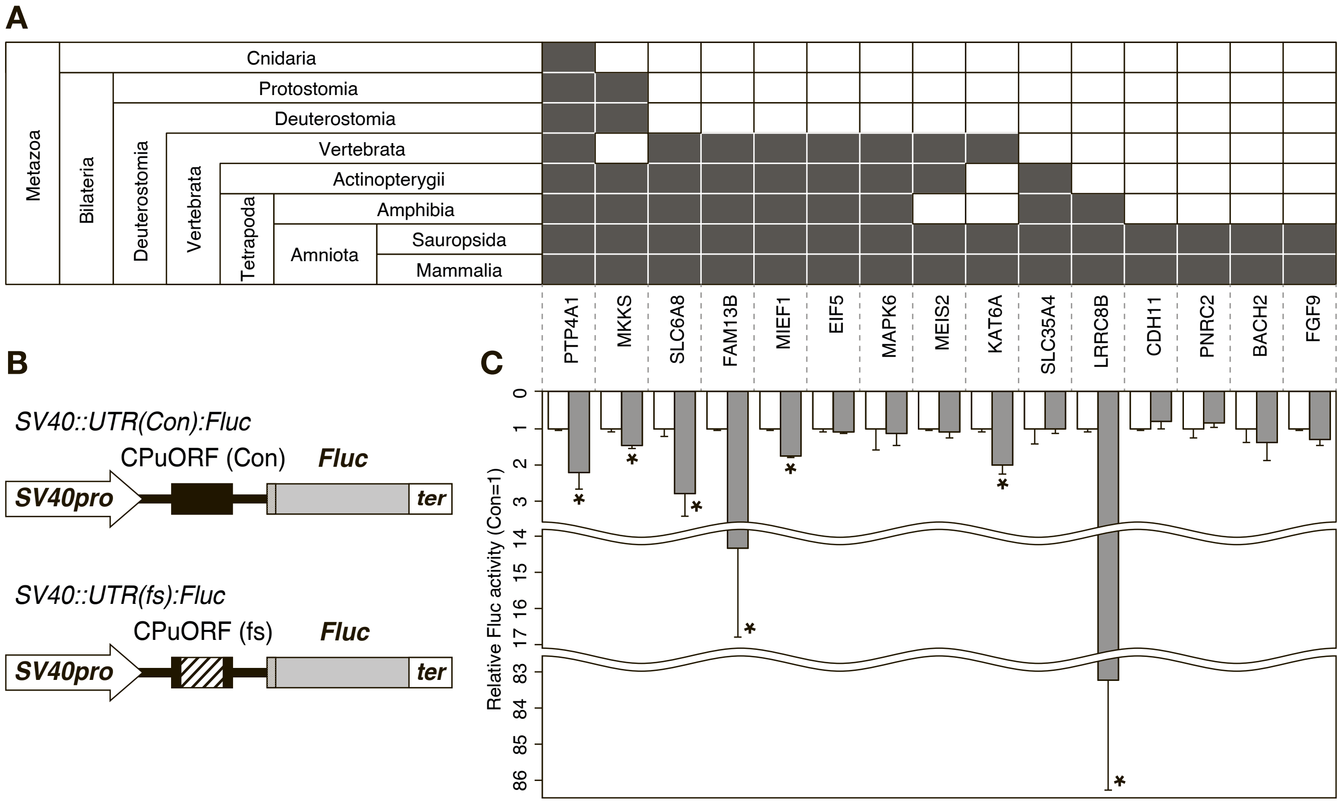
Figure 1


A
Data preparation before ESUCA
Step 0.1
Transcript dataset construction based on genome information for a certain organism
Genome sequences (FASTA) data and genomic coordinates (GFF3)
(1a) 5'UTR
mORF RNA
Translation
(1b) mORF Protein
Step 0.2
Transcript base sequence dataset construction from EST/TSA/RefSeq RNA
EST+TSA+RefSeq sequences with taxonomy DB
Extraction
(5) Taxonomy DB
Extraction
Assembling & mapping
Extraction
(2) RefSeq
(3) Assembled EST/TSA
(4) EST/TSA/RefSeq
B
ESUCA pipeline
Step 1
Extraction of uORF sequences from 5'UTR
(1a)
Step 2
Calculation of uORF-mORF fusion ratios
(2)
uORF-mORF fusion ratio ≥ 0.3
Discarded
Step 3.1
Homology searches of the uORF amino acid sequences against transcript sequence database (uORF-tBLASTn)
(2)+(3)
Step 3.2
Extraction of uORF from each uORF-tBLASTn hit sequence
Step 4.0
Extraction of downstream sequences of the ORFs matching the original uORF
Downstream sequence dataset for each uORF
Step 4.1
Selection of uORFs conserved between homologous genes (mORF-tBLASTn)
(1b)
Step 4.2
Removal of contaminant ESTs and TSAs (BLASTn)
(4)
Step 4.3
Selection of uORFs conserved between homologs from more than two orders
(5)
Step 5
Calculation of K_a/K_s ratios
Representative uORF from each order
K_a/K_s ratio ≥ 0.5 or q value ≥ 0.05
Discarded
Step 6
Determination of the taxonomic range of uORF sequence conservation
(5)
Step 7
Manual validation after ESUCA
C
Metazoa
Eumetazoa
Cnidaria
Bilateria
Protostomia
Ecdysozoa
Arthropoda
Insecta
Deuterostomia
Chordata
Vertebrata
Euteleostomi
Actinopterygii
Ostarioclupeomorpha
Sarcopterygii
Tetrapoda
Amphibia
Amniota
Sauropsida
Aves
Mammalia
Eutheria
Euarchontoglires
D

Species	CPuORF
D. melanogaster	35
D. rerio	151
G. gallus	206
H. sapiens	1038

Number of animal CPuORFs: 1430
Number of animal HGs: 1337

Figure 2



	PTP4A1	MKKS	SLC6A8	FAM13B	MIEF1	EIF5	MAPK6	MEIS2	KAT6A	SLC35A4	LRRC8B	CDH11	PNRC2	BACH2	FGF9
Cnidaria	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
Protostomia	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□
Deuterostomia	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□
Vertebrata	■	□	■	■	■	■	■	■	■	□	□	□	□	□	□
Actinopterygii	■	■	■	■	■	■	■	■	□	■	□	□	□	□	□
Amphibia	■	■	■	■	■	■	■	□	□	■	■	□	□	□	□
Sauropsida	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Mammalia	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■