Published in final edited form as:

Gene. 2011 July 15; 481(1): 1–6. doi:10.1016/j.gene.2011.04.002.

Tandem termination signal in plant mRNAs

Alex V. Kochetov^{a,b,*}, Oxana A. Volkova^{a,b}, Alexander Poliakov^{c,d}, Inna Dubchak^{c,d}, and Igor B. Rogozin^{a,e}

Oxana A. Volkova: ov@bionet.nsc.ru; Alexander Poliakov: AVPoliakov@lbi.gov; Inna Dubchak: ildubchak@lbi.gov; Igor B. Rogozin: rogozin@ncbi.nlm.nih.gov

^a Institute of Cytology & Genetics, Lavrentieva ave. 10, Novosibirsk, 630090, Russia

^b Novosibirsk State University, Novosibirsk, 630090, Russia

^c Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

^d DOE Joint Genome Institute, Walnut Creek, California 94598, USA

^e National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

Abstract

It was proposed that if some mRNA characteristics resulted in a low efficiency of termination signal, an additional closely located stop codon (tandem stop codons) could be used to prevent the harmful readthrough. However, the role of tandem terminators in higher eukaryotes was not verified and remains hypothetical. In this work the sequence features of *Arabidopsis thaliana* and *Oryza sativa* mRNAs were analyzed. It was found that plant mRNAs with UGA terminator were characterized by a higher frequency of nonsense codons in the first triplet position of 3'-UTR that could result from a weak natural selection for "reserve" stop signal. Interestingly, the presence of tandem stop codons positively correlated with a specific amino acid composition in the C-terminal position of the encoded proteins. In particular, C-terminal glycine positively correlated with significantly higher frequencies of reserve terminators at the beginning positions of 3'-UTR in UGA-containing mRNAs. This finding coincides with some earlier observations concerning the role of glycine and its codons in inefficient termination of translation and recoding (*e.g.*, 2A oligopeptide).

Keywords

mRNA; *Arabidopsis thaliana*; *Oryza sativa*; stop codon; tandem terminators; readthrough

1. Introduction

It is known that efficiency of translation termination can vary and specific variants of the terminator context can provoke readthrough resulting in a synthesis of C-end extended variants of encoded proteins (*e.g.*, Bjornsson et al., 1996; Cassan et al., 2001; Namy et al., 2001; Harrel et al., 2002; Tork et al., 2004; Cridge et al., 2006; Lao et al., 2009). This

© 2011 Elsevier B.V. All rights reserved.

*Correspondent author: A.V.K., ak@bionet.nsc.ru; fax: +7-383-3331278; phone: +7-383-3634994.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

mechanism is used by some viral and cellular mRNAs for additional synthesis of functional protein isoforms (Skuzeski et al., 1991; Robinson and Cooley, 1997; Steneberg et al., 1998; Steneberg and Samakovlis, 2001; Namy et al., 2002; Namy et al., 2003; Dreher and Miller, 2006; Lao et al., 2009) or for compensation of a negative effect of nonsense mutations (*e.g.*, Kaler et al., 2009). For instance, the stop codon of the yeast PDE2 gene, encoding a cAMP phosphodiesterase, is readthrough with an efficiency of between 2 and 8%, depending upon the genetic background. This lengthens the Pde2p protein by 20 amino acids causing physiologically relevant alterations in yeast cellular cAMP levels (Namy et al., 2002). Despite the cases of functional readthrough were intensively investigated the role of the stop codon nucleotide contexts and C-end amino acids remains doubtful and accurate prediction of the readthrough mRNAs is not possible. For example, some of the identified stop codons which are readthrough do not have obvious suboptimal 5'-or 3'-contexts (Namy et al., 2003; Cridge et al., 2006). Probably, the readthrough efficiency depends on various mRNA and protein features (*e.g.*, Singh et al., 2010) as well as the current state of translation factors that hampers its prediction.

This type of programmed stop codon readthrough potentially makes an important contribution to the spectrum of gene regulatory strategies employed by the cell. It can be used to expand the range of polypeptides encoded by a core set of genes. Recent computational evaluations predicted a number of candidate genes producing C-end extended protein isoforms in yeast (Namy et al., 2003; Williams et al., 2004) and *Drosophila* (Sato et al., 2003; Lin et al., 2007) genomes. In plants, although candidate readthrough genes have been identified using computational approaches in rice (Liu and Xue, 2004) all of the experimentally verified examples, to date, involve readthrough-mediated decoding of viral RNA genomes (Dreher and Miller, 2006; Lao et al., 2009).

Indeed, C-terminal protein segment can play an important role in a protein subcellular targeting (ER-retention signal, peroxisome and vacuole targeting) and posttranslational modifications (*e.g.*, prenylation motif) (Austin et al., 2007). However, this also means that an occasional readthrough can be harmful: erroneous synthesis of C-end extended variants of some proteins could result in improper functional characteristics and negatively influences related cellular processes. It was considered earlier that if some mRNAs contain a suboptimal translation termination signal and are prone to readthrough, closely located downstream in-frame nonsense codons can be used as an additional (reserve) stop signal to terminate translation (*e.g.*, Major et al., 2002; Williams et al., 2004; Liang et al., 2005; Adachi and Cavalcanti, 2009). The presence of closely located additional stop codons downstream of ORFs reduces the number of extra amino acids, which could influence protein folding: the addition of fewer extra amino acids increases the likelihood that the protein will preserve its three-dimensional structure and function. These factors suggest that the existence of tandem stop codons would confer a selective advantage. However, this assumption was not investigated in detail and the functional significance of tandem stop codons is still under discussion (Major et al., 2002).

We hypothesized that if some reserve stop codons are functional, the nonsense codons could occur significantly more frequently at the very beginning of 3'-UTR and in the same reading frame as CDS. In this work the frequencies of nonsense codons within the proximal segments of 3'-UTRs of plant mRNAs were analyzed. It was found that *A. thaliana* and *O. sativa* mRNAs with UGA stop codons were characterized by a higher frequency of nonsense codons in the first triplet position of 3'-UTR. Interestingly, higher frequency of the "reserve" terminator positively correlated with an occurrence of Gly in the C-terminal position of the encoded proteins. Probably, specific amino acid composition of C-terminal protein positions can influence translation termination efficiency.

2. Methods

GenBank entries (RefSeq subset) were obtained using the following search fields and terms: “Organism”, *Arabidopsis thaliana*; “Molecule”, mRNA; Gene Location: “Genomic DNA/RNA”; Limits: “exclude STSs, exclude working draft, exclude TPA, exclude patents”. Nucleotide sequences of CDS with 15-nucleotide-long 5'-UTR segment and 100-nucleotide long 3'-UTR segment were isolated with the aid of software *ReadSeq* (<http://iubio.bio.indiana.edu/soft/molbio/readseq/java/>). The samples were purified from identical and highly similar (>95%) nucleotide sequences with the aid of software *CleanUp* (Grillo et al. 1996). 17623 mRNAs of *Arabidopsis thaliana* and 14925 mRNAs of *Oryza sativa* were sampled.

Statistical differences in the positional nucleotide or triplet frequencies between different mRNA subsamples were evaluated with Mann-Whitney *U*-test. Expected frequencies of nonsense codons were calculated taking into account the bias in the frequencies of corresponding dinucleotides (for example, an expected UGA frequency was calculated as $U \cdot G \cdot A \cdot R(UG) \cdot R(GA) \cdot R(UNG)$, where U, G, A are the nucleotide frequencies in a proximal 100-nucleotide long 3'-UTR fragment in a corresponding mRNA subsample, and $R(XY)$ was calculated as a ratio of observed to expected frequencies of a dinucleotide XY) (Karlin et al, 1994). Orthologs in *Arabidopsis thaliana* and *Oryza sativa* genomes were identified in pairwise genomic alignments constructed by algorithms that implement an efficient combination of global and local alignment methods based on Shuffle-LAGAN global chaining algorithm (Dubchak et al., 2009).

3. Results and Discussion

3.1. Comparison of mRNA samples containing different stop codons

The average frequencies of different stop codons in the mRNA sample analyzed are shown in Table 1. The samples of mRNAs with different terminators were also compared for the following parameters: the average content of G+C nucleotides in 3rd position of codons in protein coding sequences (CDS_GC3rd), and the average content of G+C nucleotides in 3'-UTR (3'UTR_GC; Table 1).

The frequencies of terminators UAA, UGA and UAG were considerably different (*A. thaliana*: UGA>UAA>UAG; *O. sativa*: UGA>UAG>UAA). *A. thaliana* mRNA samples containing different terminators were characterized by similar G+C content whereas corresponding samples of *O. sativa* mRNAs were slightly different. In general, the usage of particular stop codon in plant genes was not strongly influenced by a general base composition of corresponding genome regions (*e.g.*, isochores in mammalian genomes) (Beutler et al., 1989; Zheng and Zhang, 2008; Sabbia et al., 2009; Burns et al., 2009; Mukhopadhyay and Ghosh, 2010).

3.2. Nonsense triplet distribution at the beginning of 3'-UTR

The average frequencies of nonsense codons in the region spanning fifteen proximal triplets from the 3'-UTR beginning are shown in Table 2 (frame 0 corresponds to the CDS reading frame – *e.g.*, UAA.aaa.aaa.; frame 1 is shifted for one position: UAAa.aaa.aaa.; frame 2 is shifted for two positions, respectively). One may see that plant mRNA samples with different terminators were characterized by similar mean frequencies of nonsense codons in this 3'-UTR segment. The expected frequencies of nonsense triplets were calculated on the basis of average nucleotide content and dinucleotide preferences in the proximal 3'-UTR fragment (Karlin et al., 1994).

The distributions of the average nonsense codons frequencies in three reading frames in the first 15 triplets of *A. thaliana* 3'-UTRs were evaluated (Fig. 1A–C). It was found that the only position where nonsense codon frequency was significantly higher than the average value was position +1 (frame 0) downstream of UGA stop codon ($p < 0.0005$).

O. sativa mRNAs with different terminators were also characterized by some increase in nonsense triplet frequency at the first position downstream of UGA stop codon in the frame 0 ($p < 0.005$) (Fig. 1D–F). Unlike the *A. thaliana* case, we observed high frequencies of nonsense triplets at the first position in the frame +1. This increase was specific for UAA- and UAG-containing mRNA samples. Functional significance of these nonsense triplets located out of the CDS reading frame is not clear.

We additionally analyzed the presence of nonsense triplets in the first nine 3'-UTR positions after conserved UGA stop codons in orthologous *A. thaliana* and *O. sativa* mRNAs (Table 3). A significant excess of nonsense triplets in the frame 0 was also found ($P = 0.00001$ according to the Fisher exact test). However, the frequency of cases when nonsense triplets were found in both orthologs is relatively low (4%). It is likely that the observed significant tendency to have the in-frame UGA/UAA/UAG triplets after UGA stop codons additionally depends on some other features of mRNA or corresponding protein. We also analyzed the frequency of nonsense triplets after variable UGA stop codons (UGA > UAG/UAA and UAG/UAA > UGA) in the set of orthologous genes. For UGA stop codons in *A. thaliana* and UAG/UAA stop codons in *O. sativa* a significant increase in the frequency of nonsense triplets in *A. thaliana* in the frame 0 compared to frames 1 and 2 was found ($P = 0.04$ according to the Fisher exact test) (Table 3). A similar tendency was observed for UAG/UAA stop codons in *A. thaliana* and UGA stop codons in *O. sativa*, however this increase was not significant (results not shown). These results further support the hypothesis that nonsense triplets in the frame 0 are associated with UGA stop codons.

3.3. Interdependence between the type of C-terminal amino acid and tandem terminator

It is known that the nucleotide context of stop codon and amino acid composition in C-terminal positions of the encoded protein could influence termination efficiency (e.g., Bjornsson et al., 1996; Cassan et al., 2001; Namy et al., 2001; Harrel et al., 2002; Tork et al., 2004; Cridge et al., 2006; Lao et al., 2009). It may be hypothesized that the subsample of mRNAs with tandem stop codons contains a fraction whose characteristics provoke the readthrough that resulted in a selection of additional stop codon in evolution. In the case of tandem terminators, 3'-nucleotide context of stop codon is fixed (UGA.UAA, UGA.UGA, or UGA.UAG) and 5'-context could be more functionally important. In particular, stop codon readthrough could result from a specific amino acid composition of C-end of a corresponding protein if it is poorly compatible with efficient termination of translation. The difference in average frequencies of amino acids in C-terminal positions were calculated for proteins encoded by mRNAs with either single or tandem terminators (UAA-, UGA- and UAG containing mRNAs were analyzed separately; Table 4). It was found that tandem terminator correlated with specific amino acid preferences in C-terminal protein position. In particular, *A. thaliana* proteins encoded by UGA-containing mRNAs with double stop codons were characterized by a significantly higher frequency of Gly ($P < 10^{-6}$) (Table 4). Similar analysis of *O. sativa* proteins encoded by UGA-containing mRNAs with tandem terminator showed that C-terminal Gly was the only overrepresented amino acid (albeit the difference was less significant in comparison with *A. thaliana*, $P < 0.005$, Z score = -2.65).

Thus, the most significant difference between the proteins encoded by mRNAs with single and tandem stop codons was the difference in C-terminal Gly. If Gly negatively influences the recognition of stop codon UGA and provokes the readthrough this could be the reason of fixation of the reserve stop codon in evolution. To test it further, mRNA samples encoding

proteins with either C-terminal Gly or other amino acids were isolated and the distributions of in-frame nonsense codons at the beginning of 3'-UTRs of these mRNAs were compared. It was found that the presence of Gly in C-terminal protein position specifically correlated with a higher occurrence of in-frame nonsense codons downstream of terminator UGA (Fig. 2, positions 1 ($p < 0.000001$), 2 ($p < 0.005$), 4 ($p < 0.02$)). UGA-containing *O. sativa* mRNAs encoding proteins with Gly in C-terminal position were also characterized by an increased frequency of "tandem" terminators (data not shown)).

4. Conclusion

It is widely considered that translation termination efficiency can depend on the specific nucleotide context of a stop codon and C-end amino acids of the encoded proteins (*e.g.*, Bjornsson et al., 1996; Cassan et al., 2001; Namy et al., 2001; Harrel et al., 2002; Tork et al., 2004; Cridge et al., 2006; Lao et al., 2009). Low termination efficiency could result in a readthrough and a synthesis of C-end extended protein isoforms (similarly, leaky scanning and reinitiation mechanisms can result in an additional synthesis of functional N-end truncated protein isoforms) (Kochetov et al., 2005; Kochetov, 2008; Kochetov et al., 2008; Volkova and Kochetov, 2010; Bazykin and Kochetov, 2011). It was demonstrated that in a few cases these C-end extended protein isoforms were functional, and the readthrough was used to increase the mRNA coding potential (Skuzeski et al., 1991; Robinson and Cooley, 1997; Steneberg et al., 1998; Steneberg and Samakovlis, 2001; Namy et al., 2002; Namy et al., 2002; Dreher and Miller, 2006; Kaler et al., 2009). Rather than being a translation "error", stop-codon readthrough can have important effects on other cellular processes such as mRNA degradation and, in some cases, can confer a beneficial phenotype to the cell [von der Haar and Tuite, 2007; Yamaguchi-Kabata et al., 2008; it was predicted that 149 *Drosophila* mRNAs encode C-end extended protein isoforms due to readthrough (Lin et al., 2007)]. However, it is also quite possible that an occasional readthrough and production of C-end extended protein isoforms can be harmful (Inoue et al., 2007; Tse et al., 2010) since C-terminal part of proteins is often functionally essential (Austin et al., 2007).

It was suggested earlier that closely located in-frame stop codons could be used as a "reserve" termination signal for the ribosomes skipping the suboptimal 5'-proximal termination site (Major et al., 2002; Williams et al., 2004; Liang et al., 2005; Adachi and Cavalcanti, 2009). However, the role of tandem terminators in plants was not verified and remains hypothetical. In this work we analyzed the samples of *A. thaliana* and *O. sativa* mRNAs for the presence of functional "tandem" terminators. mRNA samples with stop codons UAA, UGA and UAG were analyzed separately because these terminators could differ in a readthrough capability and context dependency.

The frequency of nonsense codons at the 3'-UTR beginning is likely to result from (i) occasional nucleotide combinations whose frequencies are determined by average nucleotide content and dinucleotide preferences (*i.e.*, the expected values in Table 2); (ii) negative selection against in-frame nonsense triplets if the readthrough is functional and C-extended protein isoform is synthesized; (iii) natural selection for in-frame nonsense-triplets if the readthrough is harmful and "reserve" terminator is needed. We hypothesized that if "reserve" stop codons were selected in evolution as an additional compensatory signal against the harmful translational readthrough, their average frequency should be higher at the very beginning of 3'-UTR and in the CDS reading frame. To test this assumption we analyzed *Arabidopsis thaliana* mRNA sample for the frequency of nonsense codons located in different reading frames at the proximal 3'-UTR segments. It was found that mRNA sample with UGA stop codon was characterized by a significantly higher frequency of nonsense codons in the first triplet position of 3'-UTR in comparison with both the expected values and the average nonsense frequencies in other reading frames (Fig. 1A-C). Analysis

of *O. sativa* mRNAs also showed that in the reading frame of CDS the highest frequency of nonsense codons occurred at the first position of 3'-UTR downstream of terminator UGA (Fig. 1D-F). Increase in in-frame nonsense triplets frequency was found in plant mRNAs with conservative terminator UGA (Table 3). Note that the UGA nonsense codon is widely considered to be the most prone to readthrough. It was quite likely that this statistical deviation reflects a weak natural selection for the “reserve” stop codon and a group of mRNAs (probably rather small) has a functional tandem terminator. We assumed that an appearance of tandem terminator resulted from some specific features of mRNA (or related protein) not compatible with efficient termination of translation.

Comparison of amino acid composition in C-terminal position of plant proteins encoded by mRNAs with tandem *versus* single stop codons revealed the statistically significant differences (Table 4). In particular, proteins encoded by UGA-containing mRNAs with double stop codons were characterized by a significantly higher frequency of Gly. It may be hypothesized that the “tandem” stop codons could correlate with a specific C-terminal amino acid composition poorly compatible with efficient termination. Indeed, the reverse analysis showed that the presence of Gly in C-terminal protein position specifically correlated with a higher occurrence of nonsense codons located at the very beginning of 3'-UTR in *A. thaliana* mRNAs with terminator UGA (positions 1, 2, 4 in the CDS frame; Fig. 2).

Interestingly, C-terminal Gly was demonstrated to be involved in a specific translation termination control in eu- and prokaryotes. For instance, 2A oligopeptides are autonomous elements containing a D(V/I)EXNPGP motif at the C terminus. Protein synthesis from an open reading frame containing an internal 2A coding sequence yields two separate polypeptides, corresponding to sequences up to and including 2A and those downstream. Ribosomes pause at the end of the 2A coding sequence, over the Gly and Pro codons, and the nascent chain up to and including this Gly is released. Translation-terminating release factors eRF1 and eRF3 play key roles in the reaction. This is an example of recoding where 2A promotes unconventional termination after decoding of the glycine codon and continued translation beginning with the 3' adjacent proline codon (Doronina et al., 2008; Bidou et al., 2010). It was considered that specific features of Gly residue make an important contribution to this mechanism (Atkins et al., 2007). The glycine codons GGA or GGG, on the 5' side of stop codons UAG and UGA, were also associated with uniquely low termination efficiency in *Escherichia coli* (Zhang et al., 1998). It may be assumed that C-terminal Gly also negatively influences translation termination in UGA-containing *Arabidopsis thaliana* and *Oryza sativa* mRNAs. Interestingly, frequencies of Gly in three C-terminal protein positions are non-uniform (Table 5): Gly is less frequent in C-terminal position of proteins translated from UGA- and UAG-containing mRNAs in comparison with both upstream positions in the same sample and with proteins encoded by UAA-containing mRNAs. It is quite likely that a combination of C-terminal Gly and UGA / UAG stop codons can be unfavorable and should be avoided in gene engineering constructs designed for plant transgenesis. However, a number of plant mRNAs with UGA stop codon encode proteins with C-terminal Gly. In these cases some characteristics of mRNAs and corresponding proteins could prevent the harmful consequences of poor termination and readthrough. First, if a readthrough produces functional (or at least harmless) variant of a protein – thus, no need in a tandem terminator existed. Moreover, some C-extended protein isoforms produced by these mechanisms were found to be functional (e.g., Steneberg and Samakovlis, 2001; Namy et al., 2003; Dreher and Miller, 2006; Lao et al., 2009). There can also be some other features influencing the (potential) negative role of C-terminal Gly in termination – for example, certain variants of nucleotide context.

It is likely that (predicted) negative role of C-terminal Gly in poor termination on UGA stop codon is mediated by some side effects of interactions between release factors, mRNA, ribosome, Gly-tRNAs, etc. – inside the terminating complex. If some mutation in a stop codon context decreases the translation termination efficiency and makes mRNA prone to a harmful readthrough, various ways to neutralize it are open. Compensatory mutations can appear in different sites (*e.g.*, changing stop codon itself (from UGA to UAA), or enhancing the termination efficiency through the replacement of nucleotides in other important positions). Indeed, one possible way is an appearance of a closely located downstream stop codon which can serve as a “reserve” terminator for the ribosomes reading through the suboptimal 5'-proximal site. It is likely that the higher frequency of nonsense codons in the first triplet position of 3'-UTRs of plant mRNAs with UGA terminator reflects this evolution trend. It may be assumed that “tandem” stop codons in plant mRNAs may play a functional role, but the number of genes where this mechanism is utilized is hardly large (*ca.* 100 – 120, similar to *Drosophila* (Lin et al., 2007)). In particular, the tandem stop codons could compensate for the specific amino acid composition of C-terminal part of some polypeptides incompatible with efficient translation termination but needed to support some important protein functions. In our opinion, such mRNAs can represent a promising model for the investigation of plant translation termination signal.

Acknowledgments

We are grateful to Prof. Lev L. Kisselev for discussions. This work was supported by Russian Ministry of Science & Education (2.1.1/10551 (2.1.1/6382); 02.740.11.0705; 02.740.11.0277; P128), the Program of Russian Academy of Sciences (Molecular and Cellular Biology) and intramural funds of the DHHS (NIH, National Library of Medicine). The authors are also grateful to SD RAS Complex Integration Program for partial support. The work conducted by the U.S. Department of Energy Joint Genome Institute (I.D. and A.P.) was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

1. Adachi M, Cavalcanti AR. Tandem stop codons in ciliates that reassign stop codons. *J Mol Evol.* 2009; 68:424–431. [PubMed: 19294453]
2. Atkins JF, Wills NM, Loughran G, Wu CY, Parsawar K, Ryan MD, Wang CH, Nelson CC. A case for “StopGo”: reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA.* 2007; 13:803–810. [PubMed: 17456564]
3. Austin RS, Provart NJ, Cutler SR. C-terminal motif prediction in eukaryotic proteomes using comparative genomics and statistical over-representation across protein families. *BMC Genomics.* 2007; 8:191. [PubMed: 17594486]
4. Beutler E, Gelbart T, Han JH, Koziol JA, Beutler B. Evolution of the genome and genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage. *Proc Natl Acad Sci USA.* 1989; 86:192–196. [PubMed: 2463621]
5. Bazykin GA, Kochetov AV. Alternative translation start sites are conserved in eukaryotic genomes. *Nucleic Acids Res.* 2011; 39:567–577. [PubMed: 20864444]
6. Bjornsson A, Mottagui-Tabar S, Isaksson LA. Structure of the C-terminal end of the nascent peptide influences translation termination. *EMBO J.* 1996; 15:1696–1704. [PubMed: 8612594]
7. Bidou L, Rousset JP, Namy O. Translational errors: from yeast to new therapeutic targets. *FEMS Yeast Res.* 2010; 10:1070–1082. [PubMed: 20955199]
8. Burns CC, Campagnoli R, Shaw J, Vincent A, Jorba J, Kew O. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *J Virol.* 2009; 83:9957–9969. [PubMed: 19605476]
9. Cassan M, Rousset JP. UAG readthrough in mammalian cells: effect of upstream and downstream stop codon contexts reveal different signals. *BMC Mol Biol.* 2001; 2:3. [PubMed: 11242562]

10. Cridge AG, Major LL, Mahagaonkar AA, Poole ES, Isaksson LA, Tate WP. Comparison of characteristics and function of translation termination signals between and within prokaryotic and eukaryotic organisms. *Nucleic Acids Res.* 2006; 34:1959–1973. [PubMed: 16614446]
11. Doronina VA, Wu C, de Felipe P, Sachs MS, Ryan MD, Brown JD. Site-specific release of nascent chains from ribosomes at a sense codon. *Mol Cell Biol.* 2008; 28:4227–4239. [PubMed: 18458056]
12. Dreher TW, Miller WA. Translational control in positive strand RNA plant viruses. *Virology.* 2006; 344:185–197. [PubMed: 16364749]
13. Dubchak I, Poliakov A, Kislyuk A, Brudno M. Multiple whole-genome alignments without a reference organism. *Genome Res.* 2009; 19:682–689. [PubMed: 19176791]
14. Grillo G, Attimonelli M, Liuni S, Pesole G. CLEANUP: a fast computer program for removing redundancies from nucleotide sequence databases. *Comp Appl Biosci.* 1996; 12:1–8. [PubMed: 8670613]
15. Harrell L, Melcher U, Atkins JF. Predominance of six different hexanucleotide recoding signals 3' of read-through stop codons. *Nucleic Acids Res.* 2002; 30:2011–2017. [PubMed: 11972340]
16. Inoue K, Ohyama T, Sakuragi Y, Yamamoto R, Inoue NA, Yu LH, Goto Y, Wegner M, Lupski JR. Translation of SOX10 3' untranslated region causes a complex severe neurocristopathy by generation of a deleterious functional domain. *Hum Mol Genet.* 2007; 16:3037–3046. [PubMed: 17855451]
17. Kaler SG, Tang J, Donsante A, Kaneski CR. Translational read-through of a nonsense mutation in ATP7A impacts treatment outcome in Menkes disease. *Ann Neurol.* 2009; 65:108–113. [PubMed: 19194885]
18. Karlin S, Ladunga I, Blaisdell BB. Heterogeneity of genomes: measures and values. *Proc Natl Acad Sci USA.* 1994; 91:12837–12841. [PubMed: 7809131]
19. Kochetov AV. Alternative translation and hidden coding potential of eukaryotic mRNAs. *BioEssays.* 2008; 30:683–691. [PubMed: 18536038]
20. Kochetov AV, Ahmad S, Ivanisenko V, Volkova OA, Kolchanov NA, Sarai A. uORFs, reinitiation and alternative translation start sites in human mRNAs. *FEBS Lett.* 2008; 582:1293–1297. [PubMed: 18358843]
21. Kochetov AV, Sarai A, Rogozin IB, Shumny VK, Kolchanov NA. The role of alternative translation start sites in the generation of human protein diversity. *Mol Genet Genomics.* 2005; 273:491–496. [PubMed: 15959805]
22. Lao NT, Maloney AP, Atkins JF, Kavanagh TA. Versatile dual reporter gene systems for investigating stop codon readthrough in plants. *PLoS One.* 2009; 4:e7354. [PubMed: 19816579]
23. Liang H, Cavalcanti ARO, Landweber LF. Conservation of tandem stop codons in yeasts. *Genome Biol.* 2005; 6:R31. [PubMed: 15833118]
24. Lin MF, Carlson JW, Crosby MA, Matthews BB, Yu C, Park S, Wan KH, Schroeder AJ, Gramates LS, St Pierre SE, Roark M, Wiley KL Jr, Kulathinal RJ, Zhang P, Myrick KV, Antone JV, Celniker SE, Gelbart WM, Kellis M. Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res.* 2007; 17:1823–1836. [PubMed: 17989253]
25. Liu Q, Xue Q. Computational identification and sequence analysis of stop codon readthrough genes in *Oryza sativa*. *Biosystems.* 2004; 77:33–39. [PubMed: 15527943]
26. Major LL, Edgar TD, Yee, Yip P, Isaksson LA, Tate WP. Tandem termination signals: myth or reality? *FEBS Lett.* 2002; 514:84–93. [PubMed: 11904187]
27. Mukhopadhyay P, Ghosh TC. Relationship between gene compactness and base composition in rice and human genome. *J Biomol Struct Dyn.* 2010; 27:477–488. [PubMed: 19916569]
28. Namy O, Hatin I, Rousset JP. Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Rep.* 2001; 2:787–793. [PubMed: 11520858]
29. Namy O, Duchateau-Nguyen G, Rousset JP. Translational readthrough of the PDE2 stop codon modulates cAMP levels in *Saccharomyces cerevisiae*. *Mol Microbiol.* 2002; 43:641–652. [PubMed: 11929521]
30. Namy O, Duchateau-Nguyen G, Hatin I, Hermann-Le Denmat S, Termier M, Rousset JP. Identification of stop codon readthrough genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 2003; 31:2289–2296. [PubMed: 12711673]

31. Robinson DN, Cooley L. Examination of the function of two kelch proteins generated by stop codon suppression. *Development*. 1997; 124:1405–1417. [PubMed: 9118811]
32. Sabbia V, Romero H, Musto H, Naya H. Composition profile of the human genome at the chromosome level. *J Biomol Struct Dyn*. 2009; 27:361–370. [PubMed: 19795918]
33. Sato M, Umeki H, Saito R, Kanai A, Tomita M. Computational analysis of stop codon readthrough in *D. melanogaster*. *Bioinformatics*. 2003; 19:1371–1380. [PubMed: 12874049]
34. Singh H, Andrabi M, Kahali B, Ghosh TC, Miziguchi K, Kochetov AV, Ahmad S. On nucleotide solvent accessibility in RNA structure. *Gene*. 2010; 463:41–48. [PubMed: 20470873]
35. Skuzeski JM, Nichols LM, Gesteland RF, Atkins JF. The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. *J Mol Biol*. 1991; 218:365–373. [PubMed: 2010914]
36. Steneberg P, Englund C, Kronhamn J, Weaver TA, Samakovlis C. Translational readthrough in the *hdc* mRNA generates a novel branching inhibitor in the *Drosophila* trachea. *Genes Dev*. 1998; 12:956–967. [PubMed: 9531534]
37. Steneberg P, Samakovlis C. A novel stop codon readthrough mechanism produces functional Headcase protein in *Drosophila* trachea. *EMBO Rep*. 2001; 2:593–597. [PubMed: 11463742]
38. Tork S, Hatin I, Rousset JP, Fabret C. The major 5'-determinant in stop codon read-through involves two adjacent adenines. *Nucleic Acids Res*. 2004; 32:415–421. [PubMed: 14736996]
39. Tse H, Cai JJ, Tsoi HW, Lam EP, Yuen KY. Natural selection retains overrepresented out-of-frame stop codons against frameshift peptides in prokaryotes. *BMC Genomics*. 2010; 11:491. [PubMed: 20828396]
40. Williams I, Richardson J, Starkey A, Stansfield I. Genome-wide prediction of stop codon readthrough during translation in the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 2004; 32:6605–6616. [PubMed: 15602002]
41. Volkova OA, Kochetov AV. Interrelations between the nucleotide context of human start AUG codon, N-end amino acids of the encoded protein and initiation of translation. *J Biomol Struct Dynam*. 2010; 27:611–618.
42. von der Haar T, Tuite MF. Regulated translational bypass of stop codons in yeast. *Trends Microbiol*. 2007; 15:78–86. [PubMed: 17187982]
43. Zheng WX, Zhan CT. Biological implications of isochore boundaries in the human genome. *J Biomol Struct Dyn*. 2008; 25:327–336. [PubMed: 18092827]
44. Yamaguchi-Kabata Y, Shimada MK, Hayakawa Y, Minoshima S, Chakraborty R, Gojobori T, Imanishi T. Distribution and effects of nonsense polymorphisms in human genes. *PLoS One*. 2008; 3:e3393. [PubMed: 18852891]
45. Zhang S, Ryden-Aulin M, Isaksson LA. Functional interaction between tRNA^{2Gly2} at the ribosomal P-site and RF1 during termination at UAG. *J Mol Biol*. 1998; 284:1243–1246. [PubMed: 9878344]

Fig. 1.

Average frequencies of nonsense triplets in fifteen proximal 3'-UTR positions of *Arabidopsis thaliana* and *Oryza sativa* mRNAs downstream of different terminator codons (frame 0 corresponds to CDS reading frame (UAA.nnn.nnn), frames 1 and 2 are shifted on one and two nucleotides, respectively).

Fig. 2.

The difference in average frequencies of in-frame nonsense codons within 15 proximal triplets of 3'-UTR of *A. thaliana* mRNAs between the subsamples encoding proteins with either C-terminal Gly or other residues (Gly-containing *versus* non-Gly-containing; subsamples with terminators UAA, UGA and UAG were analyzed separately).

Table 1

The average frequencies of terminator codons in *Arabidopsis* mRNAs and the average G+C content in 3'UTR (3'UTR_GC) and in 3rd codon positions of CDSs (CDS_GC3rd) in mRNA samples with different terminator codons.

	UAA	UGA	UAG
<i>Arabidopsis thaliana</i>			
Frequency	0.35	0.45	0.20
CDS_GC3 rd	0.43	0.43	0.43
3'UTR_GC	0.32	0.32	0.32
<i>Oryza sativa</i>			
Frequency	0.23	0.49	0.28
CDS_GC3 rd	0.62	0.66	0.66
3'UTR_GC	0.39	0.40	0.40

Table 2

Average frequencies of nonsense-codons within the fifteen 5'-proximal triplets of 3'-UTR in the samples with different terminators

	UAA	UGA	UAG
<i>A. thaliana</i>			
Frame 0	0.058	0.059	0.057
Frame 1	0.057	0.055	0.057
Frame 2	0.055	0.055	0.055
Expected	0.062	0.062	0.062
<i>O. sativa</i>			
Frame 0	0.051	0.050	0.048
Frame 1	0.054	0.047	0.051
Frame 2	0.050	0.047	0.050
Expected	0.055	0.051	0.052

Table 3

Nonsense triplets in orthologous *A. thaliana* and *O. sativa* 3'-UTRs after UGA stop codons.

	Nonsense triplet in both species	Nonsense triplet in <i>A. thaliana</i>	Nonsense triplet in <i>O. sativa</i>	No triplet species nonsense in both
Conserved UGA stop codon				
Frame 0 (positions +1:+9)	5%	18%	17%	60%
Frame 1 (positions+2:+10)	2%	15%	12%	71%
Frame 2 (positions+3:+11)	2%	18%	14%	66%
UGA stop codon in <i>A. thaliana</i> , UAG/UAA stop codon in <i>O. sativa</i>				
Frame 0 (positions +1:+9)	2%	16%	11%	71%
Frame 1 (positions+2:+10)	2%	12%	13%	73%
Frame 2 (positions+3:+11)	2%	12%	14%	72%

Table 4

The difference (Fd-Fs) in average amino acid frequencies in the C-terminal position of proteins encoded by the samples of *A. thaliana* mRNAs with double (Fd) or single (Fs) stop codons (statistical significance was evaluated with Mann-Whitney *U*-test; significantly different values were bolded). The samples with terminators UAA, UGA and UAG were analyzed separately.

	UAA	Z score	UGA	Z score	UAG	Z score
Ala	-0.002	0.162	-0.005	0.451	-0.022	1.272
Cys	-0.003	0.290	0.007	-1.093	-0.009	0.899
Asp	0.005	-0.481	0.004	-0.405	0.011	-0.726
Glu	0.015	-1.300	0.002	-0.182	0.024	-1.574
Phe	-0.015	1.276	-0.012	1.241	-0.044	2.785
Gly	-0.007	0.637	0.039	-4.990	0.014	-1.268
His	-0.013	1.407	0.008	-1.097	0.003	-0.294
Ile	-0.004	0.394	-0.007	0.767	-0.006	0.399
Lys	0.023	-1.759	-0.001	0.069	-0.027	1.602
Leu	-0.010	0.617	-0.027	2.092	-0.029	1.388
Met	0.016	-1.915	-0.001	0.092	0.009	-0.915
Asn	0.016	-1.281	0.004	-0.405	0.013	-0.832
Pro	-0.015	1.368	-0.023	2.554	-0.010	0.790
Gln	0.015	-1.565	0.007	-0.899	0.012	-0.994
Arg	0.013	-1.062	0.028	-2.525	-0.006	0.353
Ser	-0.021	1.314	-0.010	0.773	-0.005	0.225
Thr	-0.023	1.934	-0.011	1.176	-0.017	1.285
Val	0.014	-1.044	-0.003	0.284	0.054	-3.248
Trp	0.002	-0.378	-0.001	0.124	0.014	-1.943
Tyr	-0.006	0.640	0.001	-0.123	0.020	-1.497

Table 5

Frequencies of Gly in three C-terminal positions of proteins encoded by *A. thaliana* and *O. sativa* mRNAs with different terminator codons.

Pos.	UAA	UGA	UAG
<i>A. thaliana</i>			
-3	0.052	0.053	0.053
-2	0.064	0.055	0.050
-1	0.048	0.035	0.026
<i>O. sativa</i>			
-3	0.057	0.062	0.060
-2	0.068	0.066	0.065
-1	0.053	0.047	0.047