Title
Stop Codon Reassignment in the Wild

Authors
ivanova, Natalia
Schwientek, Patrick
Tripp, H. James
et al.

Publication Date
2014-03-19

Stop Codon Reassignment in the Wild

Natalia Ivanova[1*], Patrick Schwientek[1*], H. James Tripp[1], Christian Rinke[1], Marcel Huntemann[1], Axel Visel[1], Tanja Woyke[1], Nikos Kyrpides[1], Edward Rubin[1]

[*]*Equal First Authors*

[1] LBNL - Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, USA

**To whom correspondence should be addressed*: Natalia Ivanova: nnivanova@lbl.gov

Or Patrick Schwientek: PSchwientek@lbl.gov

March 21, 2014

ACKNOWLEDGMENTS:

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Stop Codon Reassignment in the Wild

Natalia N. Ivanova*, **Patrick Schwientek***, H. James Tripp, Christian Rinke,
Amrita Pati, Marcel Huntemann, Axel Visel, Tanja Woyke, Nikos C. Kyrpides, Edward M. Rubin

JGI JOINT GENOME INSTITUTE DEPARTMENT OF ENERGY

* Equal first authors

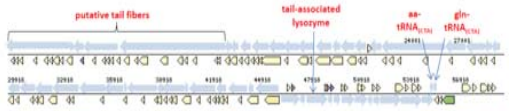
Introduction

Since the discovery of the genetic code and protein translation mechanisms (*1*), a limited number of variations of the standard assignment between unique base triplets (codons) and their encoded amino acids and translational stop signals have been found in bacteria and phages (*2-3*). Given the apparent ubiquity of the canonical genetic code, the design of genomically recoded organisms with non-canonical codes has been suggested as a means to prevent horizontal gene transfer between laboratory and environmental organisms (*4*). It is also predicted that genomically recoded organisms are immune to infection by viruses, under the assumption that phages and their hosts must share a common genetic code (*5*). This paradigm is supported by the observation of increased resistance of genomically recoded bacteria to phages with a canonical code (*4*). Despite these assumptions and accompanying lines of evidence, it remains unclear whether differential and non-canonical codon usage represents an absolute barrier to phage infection and genetic exchange between organisms.

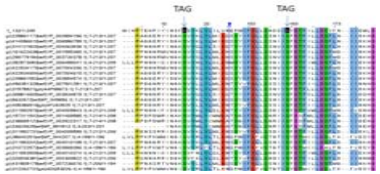
Our knowledge of the diversity of genetic codes and their use by viruses and their hosts is primarily derived from the analysis of cultivated organisms. Advances in single-cell sequencing and metagenome assembly technologies have enabled the reconstruction of genomes of uncultivated bacterial and archaeal lineages (*6*). These initial findings suggest that large scale systematic studies of uncultivated microorganisms and viruses may reveal the extent and modes of divergence from the canonical genetic code operating in nature.

To explore alternative genetic codes, we carried out a systematic analysis of stop codon reassignments from the canonical TAG *amber*, TGA *opal*, and TAA *ochre* codons in assembled metagenomes from environmental and host-associated samples, single-cell genomes of uncultivated bacteria and archaea, and a collection of phage sequences.

METHODS

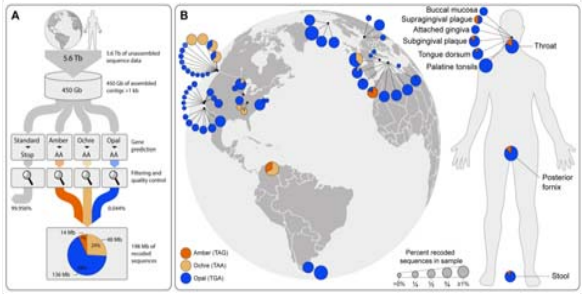


When stop codons have been reassigned to code for an amino acid, gene finding is poor assuming use of the standard genetic code (pale yellow gene calls). Gene calling and annotation greatly improve on the assumption that one of the three stop codons has been reassigned to code for some amino acid (grey gene calls). Also, the presence of tRNAs with CTA anti-codons matching TAG *(amber)* stop codons support the premise of *amber* codon reassignment.



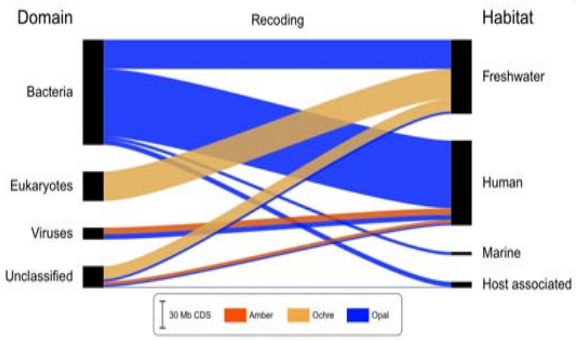
Once stop codon reassignment has been inferred by improvement in gene calling, the specific reassignment is confirmed by protein alignment of homologs. In the example shown above, the TAG stop codon (*amber*) has been reassigned to code for the amino acid serine, as evidenced by the alignment of internal *amber* stop codons to conserved serine residues in homologs of the recoded gene.

RESULTS



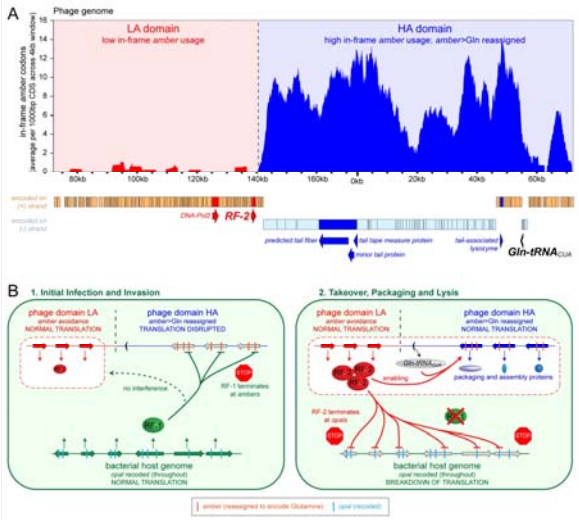
A. 5.6 Tb of raw reads from metagenomic samples taken from around the world and from several human body sites were assembled into 450 Gb of contigs >1 kb in length. Not surprisingly, the vast majority of contigs (99.956%) showed no improvement in gene finding on the assumption of stop codon reassignment. However, in 0.044% of the contigs (198 Mb), gene finding did improve on the assumption of stop codon reassignment. *Opal* reassignment predominated (69%), followed by *ochre* (24%), followed by *amber* (7%).

B. Varying ratios of reassigned to total contigs were observed in samples from terrestrial and aquatic environments and from human mouth and gut microbiomes. The greatest reassignment ratio was in a groundwater sample from a sulfidic aquifer where 10.4% of all the assembled contigs displayed evidence that one of the three stop codons had been reassigned (data not shown). High ratios of contig recoding were also detected in the human oral microbiome.



We observed distinct patterns of stop codon reassignment in the three domains of life, with bacteria showing only *opal* reassignments, *ochre* reassignments restricted to eukaryotes, and archaea devoid of codon reassignments. Among phages, we observed both *amber* and *opal* reassignments. Metagenomes of human body sites showed a high rate of reassignments compared to most other sampling sites. The majority of the remaining stop codon reassignments were found in freshwater environments (44%) representing 13% (56.0 Gb) of all examined metagenomes. In contrast marine samples contributed only 4% of recoded sequence, although they represent 48% (211.6 Gb) of the total data set (*15*). This suggests that codon reassignments are more abundant in freshwater samples compared to marine.

RESULTS (cont'd)



A. Genome organization of an *amber* recoded phage infecting unknown hosts, possibly some that are *opal* or *ochre* recoded. The phage carries release factor 2 (RF-2), which is not needed for translating *amber* recoded genes, and a Gln-tRNA$_{(CUA)}$ that is needed for translating *amber* recoded genes. *Amber* recoding is predominant only in one section of the genome (HA, high *amber*).

B. Hypothetical scenario of infection of a host that is not *amber* recoded by a phage that is *amber* recoded. 1) Initially, the phage can express genes in the low *amber* (LA) region of its genome, including RF-2. The host, being *opal* or *ochre* recoded, would have retained only RF-1. 2) The phage RF-2 interferes with host translation, allowing expression of phage Gln-tRNA$_{(CUA)}$, which allows expression of high *amber* (HA) genes for packaging and reassembly.

REFERENCES

1. M. Nirenberg *et al.*, RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc. Natl. Acad. Sci. U. S. A.* **53**, 1161 (1965).
2. R. D. Knight, S. J. Freeland, L. F. Landweber, Rewiring the keyboard: evolvability of the genetic code. *Nature Rev. Genet.* **2**, 49 (2001).
3. C. Rinke *et al.*, Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431 (2013).
4. M. J. Lajoie *et al.*, Probing the limits of genetic recoding in essential genes. *Science* **342**, 361 (2013).
5. L. A. Shackelton, E. C. Holmes, The role of alternative genetic codes in viral evolution and emergence. *J. Theor. Biol.* **254**, 128 (2008).
6. T. Woyke *et al.*, Assembling the marine metagenome, one cell at a time. *PLoS One* **4**, e5299 (2009).

ACKNOWLEDGMENTS

The work conducted by the U.S. Department of Energy Joint Genome Institute was supported in part by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231.

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.