the ancestral neurotoxic venom activity that has been retained and is utilised in so many of their counterparts? Dowell *et al.* [2] suggest that this loss of neurotoxicity is likely adaptive, and that dietary variation and/or predator–prey interactions might be responsible for driving the observed genotypic variation. This is not an unreasonable assumption given prior reports of correlations between diet and venom composition [1,19] and evidence of prey (and some predator) species developing strong resistance to viperid venoms [20]. Future comparative research incorporating both natural history information on prey composition and experimental evidence of venom toxicity to different prey items would likely reveal the adaptive basis for such divergent venom phenotypes.

In summary, venoms are complex cocktails, and their composition and therefore bioactivity is underpinned by seemingly complex and variable interactions between genes, their expression, their translation and their post-translational modification. Evidence that the loss of genes also has a strong influence on shaping venom phenotypes further reinforces the value of using animal venom systems to understand adaptation in the natural world.

REFERENCES

1. Casewell, N.R., Wüster, W., Vonk, F.J., Harrison, R.A., and Fry, B.G. (2013). Complex cocktails: the evolutionary novelty of venoms. Trends Ecol. Evol. *28*, 219–229.
2. Dowell, N., Giorgianni, M., Kassner, V., Selegue, J., Sanchez, E., and Carroll, S. (2016). The deep origin and recent loss of venom toxin genes in rattlesnakes. Curr. Biol. *26*, 2434–2445.
3. Fry, B.G. (2005). From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res. *15*, 403–420.
4. Fry, B.G., Vidal, N., Norman, J.A., Vonk, F.J., Scheib, H., Ramjan, S.F.R., Kuruppu, S., Fung, K., Hedges, S.B., Richardson, M.K., *et al*. (2006). Early evolution of the venom system in lizards and snakes. Nature *439*, 584–588.
5. Kasturiratne, A., Wickremasinghe, A.R., de Silva, N., Gunawardena, N.K., Pathmeswaran, A., Premaratna, R., Savioli, L., Lalloo, D.G., and de Silva, H.J. (2008). The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. PLoS Med. *5*, e218.
6. Chippaux, J.P., Williams, V., and White, J. (1991). Snake venom variability: methods of study, results and interpretation. Toxicon *29*, 1279–1303.
7. Casewell, N.R., Wagstaff, S.C., Wüster, W., Cook, D.A.N., Bolton, F.M.S., King, S.I., Pla, D., Sanz, L., Calvete, J.J., and Harrison, R.A. (2014). Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. Proc. Natl. Acad. Sci. USA *111*, 9205–9210.
8. Rokyta, D.R., Margres, M.J., and Calvin, K. (2015). Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. G3 (Bethesda) *5*, 2375–2382.
9. Durban, J., Pérez, A., Sanz, L., Gómez, A., Bonilla, F., Rodríguez, S., Chacón, D., Sasa, M., Angulo, Y., Gutiérrez, J.M., *et al*. (2013). Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. BMC Genomics *14*, 234.
10. Hargreaves, A.D., Swain, M.T., Hegarty, M.J., Logan, D.W., and Mulley, J.F. (2014). Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. Genome Biol. Evol. *6*, 2088–2095.
11. Junqueira-de-Azevedo, I.L.M., Bastos, C.M.V., Ho, P.L., Luna, M.S., Yamanouye, N., and Casewell, N.R. (2015). Venom-related transcripts from *Bothrops jararaca* tissues provide novel molecular insights into the production and evolution of snake venom. Mol. Biol. Evol. *32*, 754–766.
12. Reyes-Velasco, J., Card, D.C., Andrew, A.L., Shaney, K.J., Adams, R.H., Schield, D.R., Casewell, N.R., Mackessy, S.P., and Castoe, T.A. (2015). Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. Mol. Biol. Evol. *32*, 173–183.
13. Nei, M., Gu, X., and Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. USA *94*, 7799–7806.
14. Vonk, F.J., Casewell, N.R., Henkel, C.V., Heimberg, A.M., Jansen, H.J., McCleary, R.J.R., Kerkkamp, H.M., Vos, R.A., Guerreiro, I., Calvete, J.J., *et al*. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc. Natl. Acad. Sci. USA *110*, 20651–20656.
15. Lynch, V.J. (2007). Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. BMC Evol. Biol. *7*, 2.
16. Casewell, N.R., Wagstaff, S.C., Harrison, R.A., and Wüster, W. (2011). Gene tree parsimony of multilocus snake venom protein families reveals species tree conflict as a result of multiple parallel gene loss. Mol. Biol. Evol. *28*, 1157–1172.
17. Cocca, E., Ratnayake-Lecamwasam, M., Parker, S.K., Camardella, L., Ciaramella, M., di Prisco, G., and Detrich, H.W., III. (1995). Genomic remnants of alpha-globin genes in the hemoglobinless antarctic icefishes. Proc. Natl. Acad. Sci. USA *92*, 1817–1821.
18. Gilad, Y., Wiebe, V., Przeworski, M., Lancet, D., and Pääbo, S. (2004). Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. PLoS Biol. *2*, e5.
19. Daltry, J.C., Wüster, W., and Thorpe, R.S. (1996). Diet and snake venom evolution. Nature *379*, 537–540.
20. Holding, M.L., Drabeck, D.H., Jansa, S.A., and Gibbs, H.L. (2016). Venom resistance as a model for understanding the molecular basis of complex coevolutionary adaptations. Integr. Comp. Biol., in press. http://dx.doi.org/10.1093/icb/icw082.

Genomics: Evolution of the Genetic Code

Patrick J. Keeling
Canadian Institute for Advanced Research, Botany Department, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC, V6T 1Z4, Canada
Correspondence: pkeeling@mail.ubc.ca
http://dx.doi.org/10.1016/j.cub.2016.08.005

The genetic code is not *quite* universal. The rare variations that we know of reveal selective pressures on the code and on the translation machinery. New data suggest the code changes through ambiguous intermediates and that termination is context dependent.

We often hear in the scientific and even popular press that the 'genetic code' of some organism has been 'cracked' by genomics, which is of course total nonsense. The genetic code is not the nucleotide sequence of the genome, but

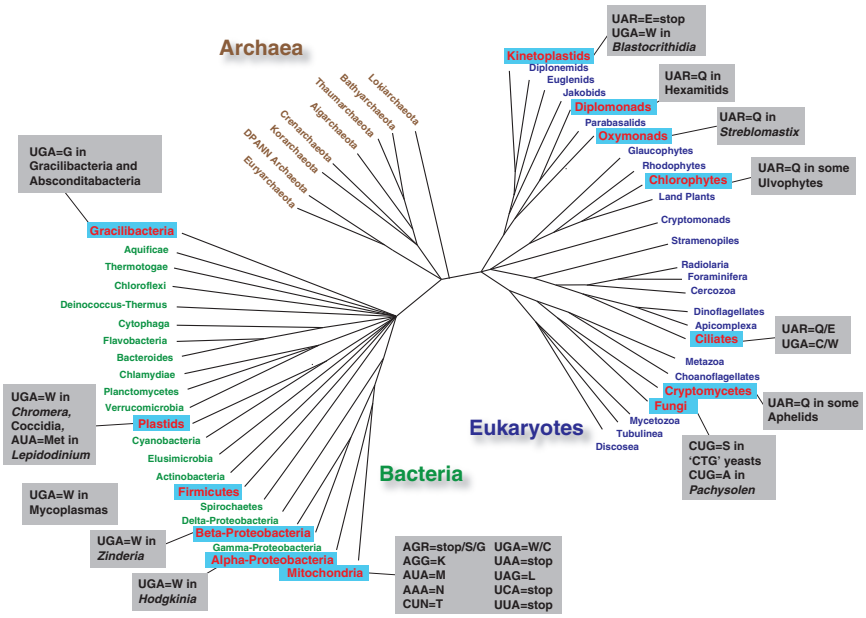

Figure 1. Variation in the genetic code.
Schematic tree of life showing known variations in the genetic code within the three domains of life — Archaea, eukaryotes, and Bacteria (including mitochondria and plastids). Alternative start codons are not included, and are relatively common, and ambiguous codons are listed by their non-canonical codon use only. Of particular note is the strong bias in changes between bacterial (UGA=W) and nuclear genomes (UAR=Q).

rather the set of rules by which those nucleotides are translated into amino acids and by extension genes into proteins. The code is actually closer to a cipher than a code, and individual species do not have a unique genetic code to be cracked; indeed one of the interesting characteristics of the code is that nearly all life shares exactly the same one, once called the 'universal genetic code' [1]. Understanding how this code originated and how it affects the molecular biology and evolution of life today are challenging problems, in part because it is so highly conserved — without variation to observe it is difficult to dissect the functional implications of different aspects of a character. But the universal code is not *quite* universal — it is perhaps the most conserved feature of molecular biology, but even still some variants do exist (for which reason I prefer the term 'canonical' over 'universal'). The nature of these non-canonical codes, as well as their distribution on the tree of life, give us a few insights into how translation may have originated, deep differences between translation systems in use today, and what forces keep the code so conserved. In a recent issue of *Current Biology*, Zahonova *et al.* [2] describe the latest such variant, discovered in a trypanosomatid (protist) symbiont of an insect by fossicking through symbiont sequences 'contaminating' the insect's transcriptome. This non-canonical code is not only new and different, but also special because it may provide a glimpse into the intermediate stages of one of the rarest of changes in biology.

The genetic code is a concept that captured the imaginations and drove much of the research of the first molecular biologists; almost immediately after the description of the double helix some very elegant models for how 4 'letters' of DNA could encode 20 letters of protein were proposed [3]. It is tempting to think that a system so central to life should be elegant, but of course that's not how evolution works; the genetic code was not designed by clever scientists, but rather built through a series of contingencies. The 'frozen accident', as it was described by Crick [4], that ultimately emerged is certainly non-random, but is more of a mishmash than an elegant plan, which led to new ideas about how the code may have evolved in a series of steps from simpler codes with fewer amino acids. So the code was not always thus, but once it was established before the last universal common ancestor of all extant life (LUCA) it has remained under very powerful selective constraints that kept the code frozen in *nearly* all genomes that subsequently diversified.

The phylogenetic distribution of the few known non-canonical codes (Figure 1) reveals some of these constraints. For one, variations in prokaryotic genomes are rare except in the proteobacteria-derived genomes of mitochondria, where the universal code is almost never used. This contrasts with almost no known changes in the cyanobacteria-derived plastids, which may reflect deeper differences in the way mitochondrial genomes evolve, perhaps rooted in a high rate of substitution. In nuclear genomes, variants are also rare, but as protist genome diversity is explored the number of cases is expanding. One interesting pattern to emerge is that the frequency of different kinds of changes in nuclear genomes is different from those of mitochondria and bacteria (the most obvious being the prevalence of UAR=Q versus UGA=W, respectively). Once again, this likely reflects a fundamental difference in their genomes, in this case the underlying translation systems (amino acyl tRNA syntheses and termination factors).

Despite their differences, all genetic systems share one significant bias: the vast majority of changes to the code we presently know of involve termination or stop codons being reassigned to encode an amino acid. This may be a true reflection of natural diversity of the code, because stop codons are by definition rare (only one of three possibilities appearing per gene, whereas even rare amino acids are typically found many times) and the fidelity of termination is potentially less critical than other possible changes. These characteristics may render such changes more statistically probable, less likely to be deleterious, or both. However, most non-canonical genetic codes are inferred from DNA sequence alone, or occasionally DNA sequences and corresponding tRNAs. Because the code governs the translation of nucleotide to amino acid sequences, a code can in principle only be confirmed when both the gene and protein

sequences are known. Stop codons stick out as a sore thumb in DNA sequence because they disrupt the coding region in a way that is difficult to overlook or misinterpret, so these codes can be strongly inferred from DNA sequence. In contrast, a code where two amino acid-encoding codons were altered would appear in DNA sequence to be slightly divergent, but not obviously translated by a different genetic code. But it is therefore possible, even likely, that non-canonical genetic codes involving switches between amino acids have been observed but escaped our detection.

Indeed, it was the stand-out nature of stop codons that first alerted Zahonova *et al.* to the presence of the new non-canonical codes in *Blastocrithidia*. But digging deeper they encountered something more interesting. The most frequent nuclear code change is UAR=Q (stop codons UAA and UAG to encode glutamine). Zahonova *et al.* found instead that UAR=E and that UGA=W (which is common in bacteria). With all three stop codons specifying an amino acid, what means 'stop'? The answer is that UAR codons do dual-duty, encoding glutamate, but also specifying stop [2]. How this actually works in the cell is a bit mystifying, but it hints at some level at a hidden context surrounding codons or stop codons that we do not yet totally understand. Indeed, independently of Zahonova *et al.*, Swart *et al.* recently reported evidence for just such a context-dependent process linked to a similar genetic code change in a ciliate, where UAR=Q and UGA=W or stop, depending on the downstream sequence [5]. These codes are also exciting because they fulfill a central prediction of one of the models for how the genetic code evolves, the Ambiguous Intermediate model. This model posits that codons shift between two meanings through an intermediate where they have both meanings [6]. In contrast, the other major type of model posits that codons shift through an intermediate where they have neither meaning (e.g., where codons are lost entirely and reappear with a new meaning) [7]. While we must bear in mind that most non-canonical codes evolved independently from the others, and so may have evolved through different steps, it is nevertheless intriguing to have an extant nuclear genome with an ambiguous intermediate. Further study of how translation functions in *Blastocrithidia* will cast a unique light on how the code works, which will certainly be challenging since it is not only an uncultured symbiont, but also a trypanosome, a lineage which is notorious for re-writing the rule books of molecular biology and genomics [8].

REFERENCES

1. Hinegardner, R.T., and Engelberg, J. (1963). Rationale for a universal genetic code. Science *142*, 1083–1055.

2. Záhonová, K., Kostygov, A.Y., Ševčíková, T., Yurchenko, V., and Eliáš, M. (2016). An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. Curr. Biol. *26*, 2364–2369.

3. Gamow, G. (1954). Possible relation between deoxyribonucleic acid and protein structures. Nature *173*, 318.

4. Crick, F.J. (1968). The origin of the genetic code. J. Mol. Biol. *38*, 367–379.

5. Swart, E.C., Serra, V., Petroni, G., and Nowacki, M. (2016). Genetic codes with no dedicated stop codon: context dependent translation termination. Cell *166*, 691–702.

6. Osawa, S., and Jukes, T.H. (1989). Codon reassignment (codon capture) in evolution. J. Mol. Evol. *28*, 271–278.

7. Schultz, D.W., and Yarus, M. (1994). Transfer RNA mutation and the malleability of the genetic code. Mol. Biol. *235*, 1377–1380.

8. Lukeš, J., Leander, B.S., and Keeling, P.J. (2009). Cascades of convergent evolution: the corresponding evolutionary histories of euglenozoans and dinoflagellates. Proc. Natl. Acad. Sci. USA *106*, 9963–9970.

Evolution: Ocean Models Reveal Life in Deep Seas

Christophe Eizaguirre
Queen Mary University of London, School of Biological and Chemical Sciences, Mile End Road, E14NS, London, UK
Correspondence: c.eizaguirre@qmul.ac.uk
http://dx.doi.org/10.1016/j.cub.2016.07.083

Even though the deep sea represents the largest area in the world, evolution of species from those environments remains largely unstudied. A series of recent papers indicate that combining molecular tools with biophysical models can help us resolve some of these deep mysteries.

Beyond the continental shelf and below 200 m in depth, the deep ocean represents about 65% of the Earth's surface [1]. This environment supports previously unsuspected levels of biodiversity, which have only recently started being assessed using direct observations and indirect genetic and computer-based methods. For example, observations using submarines or remote-operated vehicles have shed light on the distribution and biogeography of deep-sea invertebrates [2]. However, without sample collection, the scope of such modern devices are limited, particularly when it comes to investigating population structure and the adaptive potential of cryptic species.

Many disciplines have benefited tremendously from recent advances in computer science enabling