

## NEWS AND COMMENTARY

Microbial genetics

# Split genes uncovered through science fusion

JL McClay and EJCG van den Oord

Heredity (2005) 95, 1–2. doi:10.1038/sj.hdy.6800682

Published online 13 April 2005

An ingenious mix of approaches in a new study has solved the intriguing mystery of tRNA genes apparently missing from the genome of an unusual prokaryote (Randau *et al*, 2005).

A growing number of completely sequenced and annotated genomes should allow us to build the biological parts list for every organism. Progress in this area might, therefore, seem straightforward, but nature has a tendency to introduce unexpected exceptions, which makes life a little more complex. Randau *et al* (2005), in their recent *Nature* paper, have got to the bottom of one such exception. Specifically, the authors have shown that fundamental tRNA genes thought to be missing from the genome of the parasitic species of Archaea, *Nanoarchaeum equitans*, are probably split into separate chunks in completely different parts of the genome. Using an effective synthesis of bioinformatics and wet lab approaches, the authors demonstrate how the separate halves of the missing tRNAs are joined to create functional molecules.

The recent discovery of *N. equitans* in hot submarine vents north of Iceland (Huber *et al*, 2002) caused something of a stir (Boucher and Doolittle, 2002). Analysis of ribosomal RNA sequence suggested that it diverged early in the archaeal lineage, and might represent a previously unknown phylum. Its total DNA, at a mere 0.5 megabases, is among the smallest prokaryotic genome sequenced to date, 95% of which was predicted to encode proteins or stable RNAs. Particularly surprising was that this tiny genome appeared to lack several apparently crucial genes for lipid, amino acid and nucleotide biosynthesis (Waters *et al*, 2003).

Randau and co-workers focused on the absence of tRNA genes for four amino-acid acceptors, namely glutamate (Glu), histidine (His), tryptophan (Trp) and methionine (Met). The absence of the tRNA<sup>Met</sup>, the universal translation initiator, was a particular puzzle. To tackle this problem, the authors developed a novel bioinformatics

approach. They searched for the missing tRNAs using a programme trained to recognise tRNA gene signatures from a large set of example sequences, taken from Eukarya, Archaea and Bacteria (Marck and Grosjean, 2002). In addition to finding the set of tRNAs that standard approaches predicted, the new algorithm identified nine tRNA halves spread throughout the genome. These tRNA halves could be joined *in silico* to form the missing amino-acid acceptors. Why nine halves and not eight? The answer appears to be that, even more surprisingly, two unique tRNA<sup>Glu</sup> were possible, each encoded by separate 5' half genes, which could combine with the single 3' unit.

These results are interesting, but are they convincing? How do we know that the sequences are not just part of the small amount of extraneous DNA in the *N. equitans* genome? Since the organism is parasitic, it is conceivable that these mysterious half genes are merely remnants from a free-living phase earlier in its evolution. What is more, how do the half genes join together and do they actually make functional tRNAs?

The joining mechanism appears to involve short intervening sequences on the immature half gene transcripts. These motifs form perfect reverse complements, which could allow the two halves to pair up. The fact that the total portion of sequence capable of forming reverse complement is between 12 and 14 bp makes this seem likely, since motifs of this length would be highly specific and perhaps even unique, with respect to such a small genome. A high GC content in these sequences adds further credibility to the argument. The strong bonding provided by these motifs is thought to allow the formation of the complete tRNA structure *in vivo*, before removal of the intervening sequences when the two halves are spliced.

The confirmation of the predicted tRNAs was attempted using a range of laboratory procedures. Full-length tRNA<sup>Met</sup> and tRNA<sup>Glu</sup> (both versions) were amplified from a total RNA pre-

paration. For unknown reasons, the tRNA<sup>Trp</sup> and tRNA<sup>His</sup> were not pulled out, but aminoacylation allowed the activity of the latter to be demonstrated. These reactions verify the acceptor activity of the joined tRNAs, which was shown for both tRNA<sup>Glu</sup> and tRNA<sup>His</sup>. It is unfortunate that all the amino-acid acceptors were not verified in this way. While the existence of the full-length tRNA<sup>Glu</sup> was confirmed and their functionality verified, the empirical evidence for the other three tRNAs is patchy, with no evidence for the tRNA<sup>Trp</sup> at all. Overall, the results make a convincing story, at least for a subset of the tRNAs under investigation, but the laboratory work could have been executed more comprehensively.

So why does *N. equitans* have this odd half gene arrangement? The authors invoke an argument based on RNA stability in the prebiotic world, and the evolution of RNA hairpins into primitive tRNAs (Tanaka and Kikuchi, 2001). The apparently ancient nature of *N. equitans* might represent an intermediate state whereby the original RNA hairpins are still encoded by separate genes, and need to be fused with the aid of the GC-rich reverse complementary duplex.

These results, while interesting from an evolutionary point of view, are unlikely to worry the bioinformatics community over the possibility of such missed half genes in other organisms. Randau *et al* were unable to find further examples after searching available bacterial and archaeal genome sequences. The notion of split genes is already well established in eukaryotic genetics, where exons might be located several hundreds of kilobases away from their next neighbour. This study, therefore, represents an unusual case in an unusual organism, but its message is certainly one of an elegant and complex natural world that can continue to surprise us. Although imperfect, the great strength of the study, from a methodological point of view, was the fruitful combination of bioinformatics and traditional wet lab approaches. Without the complete genome of *N. equitans* available for analysis, the fact that certain key genes were 'missing' would probably have gone unnoticed, and the putative half genes would have been much more difficult to locate. Without the laboratory work to confirm the presence of predicted tRNA constructs in the living organism, the results would have remained highly speculative. Such a blend of informatics

and empirical verification will undoubtedly become central to biology in the postgenomic era.

JL McClay and EJCG van den Oord are at Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Medical College of Virginia of Virginia Commonwealth University, PO Box 980126, Richmond, VA 23298-0126, USA

and for express mail/FEDEX: EJCG van den Oord, Virginia Biotechnology Park, 800 E Leigh St Room 1-119, Richmond, VA 23219, USA.

e-mail: [ejvandenoord@vcu.edu](mailto:ejvandenoord@vcu.edu), [jmccclay@vcu.edu](mailto:jmccclay@vcu.edu)

Boucher Y, Doolittle WF (2002). *Nature* **417**: 27–28.

Huber H *et al* (2002). *Nature* **417**: 63–66.

Marck C, Grosjean H (2002). *RNA* **8**: 1189–1232.

Randau L *et al* (2005). *Nature* **433**: 537–541.

Tanaka T, Kikuchi Y (2001). *Viva Origino* **29**: 134–142.

Waters E *et al* (2003). *Proc Natl Acad Sci USA* **100**: 12984–12988.