Journal of Biology

MINIREVIEW

SnoPatrol: how many snoRNA genes are there?

Paul P Gardner*[1], Alex Bateman[1] and Anthony M Poole[2,3]

See research articles http://www.biomedcentral.com/1471-2164/10/515 and http://www.biomedcentral.com/1471-2164/11/61

Abstract

Small nucleolar RNAs (snoRNAs) are among the most evolutionarily ancient classes of small RNA. Two experimental screens published in *BMC Genomics* expand the eukaryotic snoRNA catalog, but many more snoRNAs remain to be found.

What are snoRNAs?

The biosynthesis of eukaryote ribosomes is complex, involving numerous processing events to generate mature ribosomal RNAs (rRNAs) and the subsequent assembly of processed rRNAs with dozens of ribosomal proteins. Small nucleolar RNAs (snoRNAs) are central to ribosome maturation, being required in key cleavage steps to generate individual rRNAs, and in their capacity as guides for site-specific modification of rRNA. In the rRNA of the budding yeast *Saccharomyces cerevisiae*, on the order of 100 snoRNA-guided modifications are made during the biosynthesis of a single ribosome; this number is approximately double in humans. Around half of these modifications are methylations of the 2' position on ribose, and are carried out by C/D-box small nucleolar ribonucleoproteins (snoRNPs), which consist of a guide snoRNA acting in concert with several proteins, including Nop1p, the RNA methylase component of the snoRNP. The remaining modifications produce pseudouridine, an isomer of uridine, and are guided by H/ACA-box snoRNPs, with the Cbf5p subunit performing the pseudouridylation reaction [1]. Figure 1 illustrates the interaction between the two types of snoRNA and their respective RNA targets.

Over the past decade, the snoRNA universe has expanded rapidly. H/ACA- and C/D-family RNAs have been discovered in Archaea (where they are dubbed sRNAs, as Archaea lack nucleoli), and likewise modify rRNA, and in the Cajal body of the eukaryote cell (small Cajal body scaRNPs), where they modify small nuclear RNAs (snRNAs), the RNA constituents of the spliceosome [2]. Recently, HBII-52, a human C/D snoRNA, has been shown to regulate splicing of serotonin receptor 2C mRNA, indicating a wider role in gene regulation [3], and another C/D snoRNA has been shown to be expressed from the Epstein-Barr virus genome [4]. As our knowledge of snoRNAs expands beyond RNA modification and hints at wider regulatory roles, there is a need to identify the full repertoire of snoRNAs in a genome and establish when and on what RNAs they act. Against this backdrop, experimental screens that trawl organism-by-organism for snoRNAs are vital, as bioinformatic screens have so far failed to provide a robust computational alternative to labour-intensive experimental methods of RNA identification. Two recent papers in *BMC Genomics* by Zhang *et al*. [5] and Liu *et al*. [6] report the identification of novel snoRNAs from the rhesus monkey *Macaca mulatta* and the filamentous fungus *Neurospora crassa*, respectively. Both sets of authors experimentally investigated snoRNA pools by sequencing cDNAs derived from RNA extracted from their species of interest. Subsequent bioinformatics analysis was used by each group to classify sequences as either of the two snoRNA classes or otherwise. These approaches netted 48 H/ACA and 32 C/D box snoRNAs in the monkey and 20 H/ACA and 45 C/D box snoRNAs in the fungus. Studies like these are vital to the extension of our knowledge of how complements of snoRNAs vary through evolution. Given the intense effort required for such analyses, it is worth taking stock and asking, where are the current gaps in our knowledge of snoRNAs?

The taxonomic distribution of known snoRNAs

To investigate the taxonomic distribution of the known snoRNAs and highlight where potential new discoveries can be made, we have gathered data from the Pfam (protein families), Rfam (RNA families), Genomes Online (GOLD) and EMBL databases (Figure 2). The Rfam database uses experimentally validated ncRNA sequences that have been deposited in EMBL to search for homologous sequences across all nucleotide sequences (see the red and pink bars in Figure 2). The results show

*Correspondence: pg5@sanger.ac.uk
[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK
Full list of author information is available at the end of the article

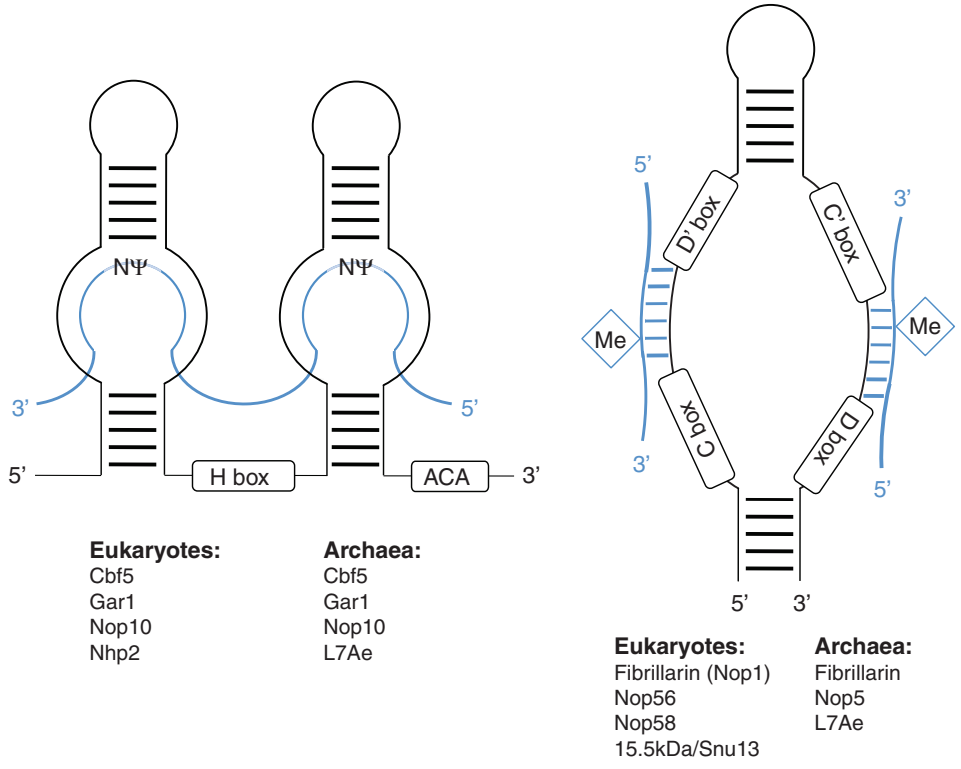

Figure 1. snoRNA structure. The structure of a H/ACA snoRNA (left) and a C/D box snoRNA (right). The targets for RNA modification are shown in blue. The most important snoRNA-associated proteins are listed below.

that for many major taxonomic clades there are few or no known snoRNAs annotated.

In the Archaea, annotated snoRNAs are notably absent from the taxon *Halobacterium*, for which a genome sequence has been available for nearly 10 years and which has been proposed to contain snoRNAs on the basis of the presence of the snoRNP-associated proteins fibrillarin and Nop56/58 [7]. In fact, only 33% of the crenarchaeal and 60% of the euryarchaeal groups carry known or predicted snoRNAs, and numbers of snoRNAs are very low in the Euryarchaeota. Still within the Archaea, snoRNAs have been annotated in some methanococcal genomes, predicted on the basis of homology to experimentally validated snoRNAs from members of the Thermoprotei [8].

Some eukaryotic taxa fare little better. For example, in the unicellular diplomonads (Diplomonadida; Figure 2), such as *Giardia lamblia*, there are no snoRNA families listed in Rfam, although putative snoRNA-like RNAs have been reported from *G. lamblia* [9,10]. Databases such as Rfam inevitably lag behind the current literature; we expect that these missing snoRNAs will be included in future releases.

The case of the microsporidia (unicellular organisms allied to the fungi) is interesting in that one genome sequence was published nearly a decade ago and eight further projects are in progress, yet despite this apparent wealth of information no snoRNAs have been identified. But like diplomonads, microsporidia clearly have components of the snoRNA machinery and almost certainly utilize snoRNAs. The absence, therefore, is due to the fact that snoRNAs have not been experimentally determined, and current bioinformatics methods are not sensitive enough to reliably identify snoRNAs in these taxa from sequence analyses alone, so none have been inferred by homology.

As expected, the Metazoa are comparatively well studied; there is a host of supporting experimental and bioinformatics evidence for snoRNAs across the metazoa, with the exception of the Cnidaria and the Platyhelminthes, which currently only have bioinformatically predicted snoRNAs based upon sequence similarity to other metazoan snoRNAs.

The genome sequence for the parasitic protozoan *Trichomonas vaginalis* (a parabasalid; Figure 2) bears one lonely C/D-box snoRNA annotation for a homolog of the fungal snoRNA snR52/Z13. Furthermore, this is a rather low-scoring hit (26.12 bits, E-value = 1.04e+02) to an otherwise exclusively fungal family and the *Trichomonas* sequence has some differences from the canonical C- and

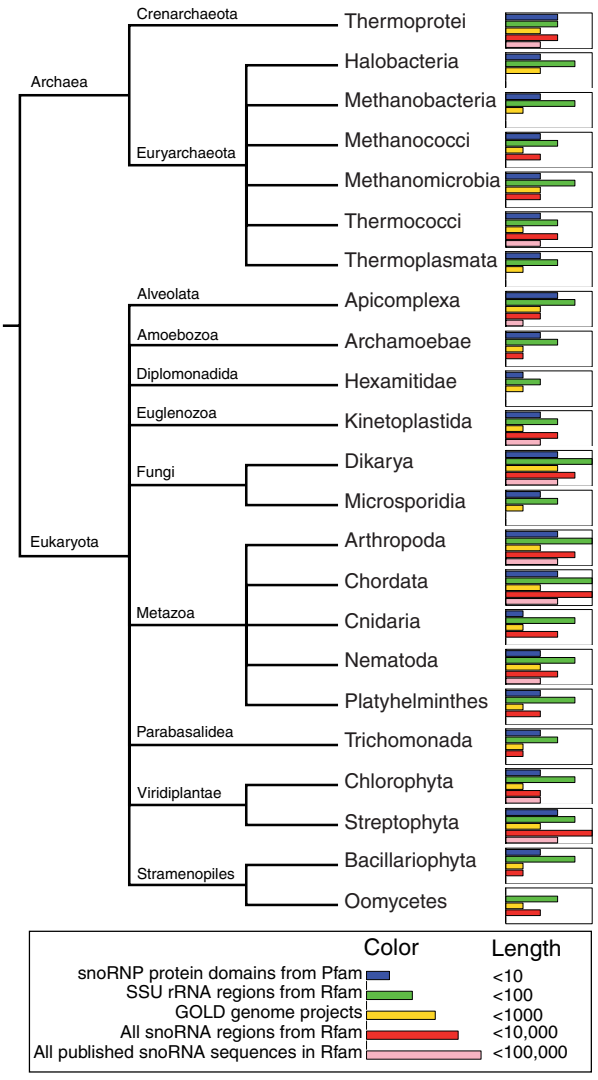

Figure 2. The taxonomic distribution of existing snoRNA annotations. The figure displays a tree derived from the top three levels of the National Center for Biotechnology Information (NCBI) taxonomy. Mapped onto this are counts of: (1) the snoRNP-associated Pfam 24.0 domains Nop, Nop10p, Gar1, SHQ1, fibrillarin and TruB_N (blue); (2) the small subunit (SSU) rRNA regions annotated by Rfam 10.0 (green); (3) genome projects registered as completed, draft or in progress from the GOLD database (version 3.0, October 22, 2009) (gold); (4) all snoRNA regions annotated by Rfam 10.0 (red); (5) EMBL sequences annotated as snoRNAs that are also annotated by Rfam 10.0 (pink). We only show here the lineages where a significant amount of sequencing effort has been directed (see Supplementary Table 1 in Additional data file 1 for the full results). Lengths of the bars correspond to counts in each taxa for each category. The shortest bar length corresponds to counts between 1 and 10 (exclusive), the next shortest is between 10 and 100 (exclusive), and so on.

D-box motifs, suggesting that the prediction may be spurious (Additional file 1). In contrast, the two main groups of green plants (Viridiplantae), the Streptophyta (multicellular green plants and some green algae) and Chlorophyta (green algae) (Figure 2), both have good snoRNA coverage, which is based on both bioinformatics and intensive experimental study of green plant snoRNAs.

Finally, the Stramenopiles (Figure 2) have five completed and one draft genome project according to the GOLD database. Both the two main lineages of stramenopiles, Bacillariophyta and Oomycetes, have reasonable numbers of predicted snoRNAs based on homology to other lineages (9 and 75, respectively), though none has been experimentally validated. Whereas counts of Pfam domains and rRNAs indicate that the snoRNP machinery is present in all known taxa of Archaea and Eukaryota, surprisingly it seems to be absent from Oomycetes. However, this lack is likely to be due to the protein sequences not yet being included in the public sequence databases rather than *bona fide* loss of the snoRNP machinery.

Future directions for snoRNA research

Up to now, bioinformatics approaches for *de novo* prediction of snoRNAs have not been a great success. As shown by Figure 2, a homology search using experimentally verified snoRNAs, as performed by the Rfam database, has some success in identifying snoRNAs in taxonomic lineages where no experiments have yet been performed. But many of these predictions need further validation before they can be entirely trusted. Using additional information such as genomic context and target information could prove quite useful in this regard [11,12]. The growing host of orphan snoRNAs - that is, snoRNAs lacking a target-modification site - are especially interesting in that several lines of evidence hint at a possible regulatory role, as with human HBII-52 [3]. The snoRNA universe is thus likely to expand in function, phylogenetic diversity, and through the discovery of new snoRNAs. Fortunately, discovery has never been easier, thanks to the growing power of new sequencing technologies.

Additional file 1: Supplementary methods and results. It contains details of how the data for Figure 2 were collected, the full dataset summarized in Figure 2 in a tabular format, and an alignment of a *T. vaginalis* candidate snoRNA and the fungal homologs.

Acknowledgements

PPG and AGB are supported by the Wellcome Trust (grant number WT077044/Z/05/Z). AMP is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

Author details

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK.
[2]Department of Molecular Biology and Functional Genomics, Stockholm University, SE-106 91 Stockholm, Sweden.
[3]School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand.

Published: 25 January 2010

References

1. Reichow SL, Hamma T, Ferré-D'Amaré AR, Varani G: **The structure and function of small nucleolar ribonucleoproteins.** *Nucleic Acids Res* 2007, **35:**1452-1464.
2. Darzacq X, Jády BE, Verheggen C, Kiss AM, Bertrand E, Kiss T: **Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs.** *EMBO J* 2002, **21:**2746-2756.
3. Kishore S, Stamm S: **The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C.** *Science* 2006, **311:**230-232.
4. Hutzinger R, Feederle R, Mrazek J, Schiefermeier N, Balwierz PJ, Zavolan M, Polacek N, Delecluse HJ, Huttenhofer A: **Expression and processing of a small nucleolar RNA from the Epstein-Barr virus genome.** *PLoS Pathog* 2009, **5:**e1000547.
5. Zhang Y, Liu J, Jia C, Li T, Wu R, Wang J, Chen Y, Zou X, Chen R, Wang X, Zhu D: **Systematic identification and evolutionary features of rhesus monkey snoRNAs.** *BMC Genomics* 2010, **11**:61
6. Liu N, Xiao ZD, Yu CH, Shao P, Liang YT, Guan DG, Yang JH, Chen CL, Qu LH, Zhou H: **SnoRNAs from the filamentous fungus *Neurospora crassa*: structural, functional and evolutionary insights.** *BMC Genomics* 2009, **10:**515.
7. Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, Shukla HD, Lasky SR, Baliga NS, Thorsson V, Sbrogna J, Swartzell S, Weir D, Hall J, Dahl TA, Welti R, Goo YA, Leithauser B, Keller K, Cruz R, Danson MJ, Hough DW, Maddocks DG, Jablonski PE, Krebs MP, Angevine CM, Dale H, Isenbarger TA, Peck RF, Pohlschroder M, Spudich JL, *et al.*: **Genome sequence of *Halobacterium* species NRC-1.** *Proc Natl Acad Sci USA* 2000, **97:**12176-12181.
8. Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP: **Homologs of small nucleolar RNAs in Archaea.** *Science* 2000, **288:**517-522.
9. Yang CY, Zhou H, Luo J, Qu LH: **Identification of 20 snoRNA-like RNAs from the primitive eukaryote, *Giardia lamblia*.** *Biochem Biophys Res Commun* 2005, **328:**1224-1231.
10. Chen XS, Rozhdestvensky TS, Collins LJ, Schmitz J, Penny D: **Combined experimental and computational approach to identify non-protein-coding RNAs in the deep-branching eukaryote *Giardia intestinalis*.** *Nucleic Acids Res* 2007, **35:**4619-4628.
11. Hoeppner MP, White S, Jeffares DC, Poole AM: **Evolutionarily stable association of intronic snoRNAs and microRNAs with their host genes.** *Genome Biol Evol* 2009, **1:**420-428.
12. Bazeley PS, Shepelev V, Talebizadeh Z, Butler MG, Fedorova L, Filatov V, Fedorov A: **snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions.** *Gene* 2008, **408:**172-179.

doi:10.1186/jbiol211
Cite this article as: Gardner PP, *et al.*: SnoPatrol: how many snoRNA genes are there? *Journal of Biology* 2010, **9:**4.