Identification of circular RNA in Archaea

Alice Heliou, supervisors: Hubert Becker et Mireille Regnier
École Polytechnique, Laboratoire d'Optique et Biosciences & Laboratoire d'Informatique

Objectives

- Identification of circular RNAs in *Pyroccocus Abyssi* genome,
- Understanding the function of ligase PAB1020 in the circularization.

Context

PAB1020 is a ligase which is supposed to have an important role in the circularization of small RNAs.

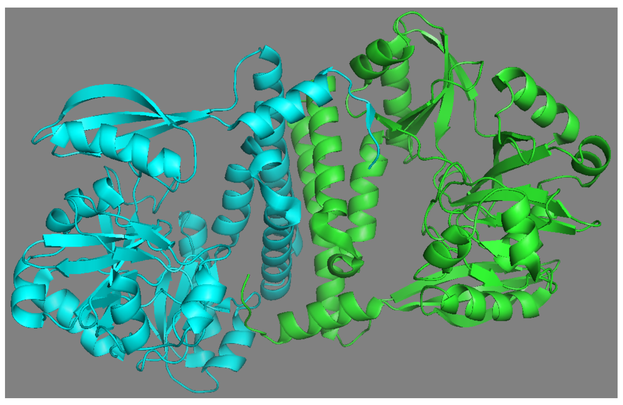


Figure 1: PDB of PAB1020

Materials & Methods

Four sets of NGS data (Ion Torrent) :

- 2 Pulldown ligase
- 1 Pulldown ligase + RNase R
- Total RNA + RNase R

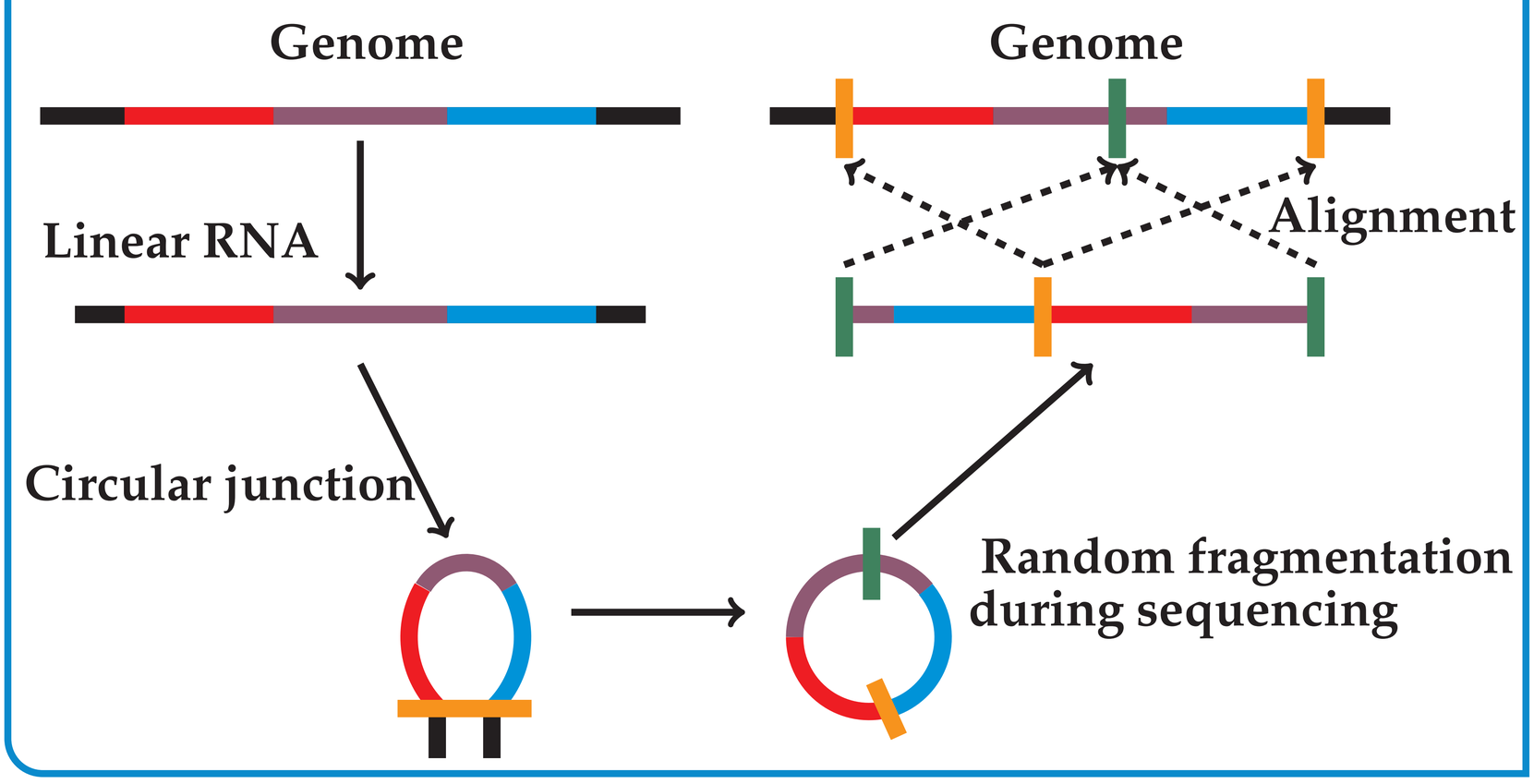
RNase R is an exoribonuclease that degrades linear RNAs, but leaves circular transcripts intact.

We analyzed those sets using **Blastn** to map against the genome.
We identified as putative circular reads those having **two matchs** such that they **cover the read entirely** and align uniquely to the genome within 1000nt in **a chiastic order**.
Then we aligned them using **Clustalw** and used different algorithms to reveal shared motifs.

State-of-the-Art

Recently, Hoffmann et al. designed an algorithm to detect splice junctions and they integrated it into their mapping tool **segemehl** [2].
Their tool is not specific to circular junctions, but detects also cis-splicing and chimeric transcript, so we used it only as a comparison.
In [1], they identified circular RNAs in three different archaea using **Blastn** and a biological validation with RNase R and RT-PCR.
Our approach is very similar but we aim to go further in the analysis of the characterization of circular RNA and the understanding of the function of ligase PAB1020.

Alignment of circular RNA on the genome



Results

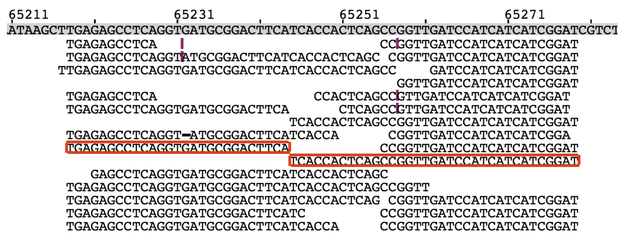


Figure 2: Reads mapped to the reference in a non-linear, chiastic manner at PABsnRNA21 locus. PABsnRNA21 is thus identified as putative circular.

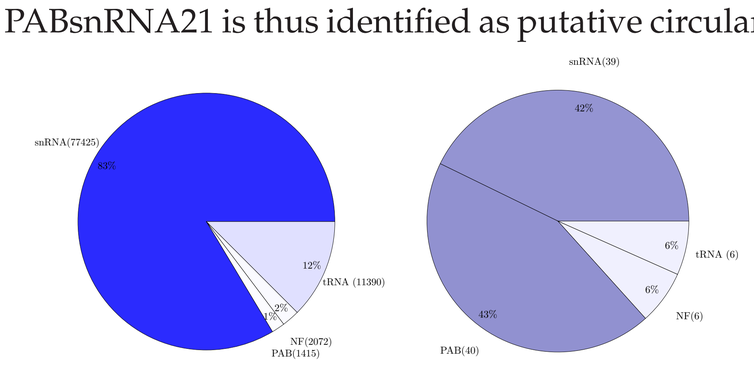


Figure 3: Annotations corresponding to the circular identified transcripts. On the left per read, on the right per transcript.

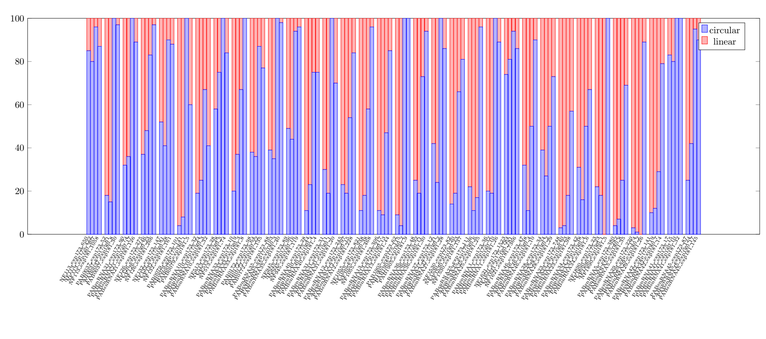


Figure 4: The percentage of circular reads over all the reads that map on a transcript, for every selected snRNA on the four experiments.

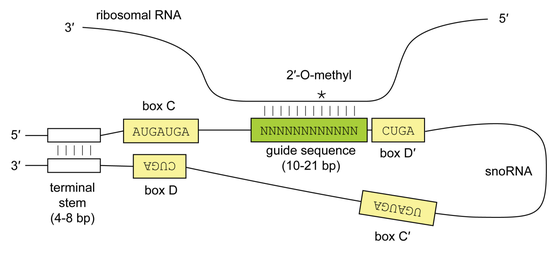


Figure 5: Drawing of a snoRNA, from http://lowelab.ucsc.edu/research.html

Limitations

Some RNAs, like the 5S ribosomal RNA have been identified as interacting with PAB1020 and as being circular using RNase R and RT-PCR.
However, they do not appear as circular when analyzing the NGS data sets with our protocol. We are maybe too restrictive.

Figure 6: Reads mapped to the genome in a non-linear, chiastic manner at PABr04 (5S sub-unit) locus. We cannot identified clearly circular junctions.

Future work

- Devise a less restrictive protocol, and understand why we do not find biologically tested RNAs,
- Characterize circular RNAs identified,
- Clarify the case of ribosomal RNAs, there is a huge amount of reads mapped in a linear and non-linear manner at PABr02 and PABr03 loci. However we are not able to identified any clear circular junction.

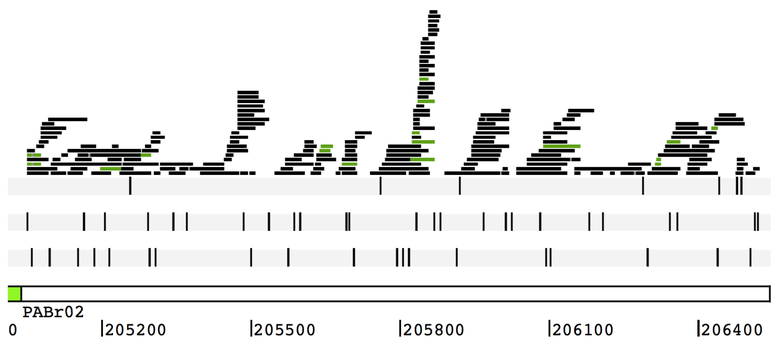


Figure 7: Reads mapped to the genome in a non-linear, chiastic manner at PABr02 locus.

References

[1] Miri Danan et al. Transcriptome-wide discovery of circular rnas in archaea. *Nucleic Acids Research*, 2012.

[2] Steve Hoffmann et al. A multi-split mapping algorithm for circular rna, splicing, trans-splicing and fusion detection. *GenomeBiology*, 2014.

Contact Information

Web www.lix.polytechnique.fr/~alheliou/

Email alice.heliou@polytechnique.edu