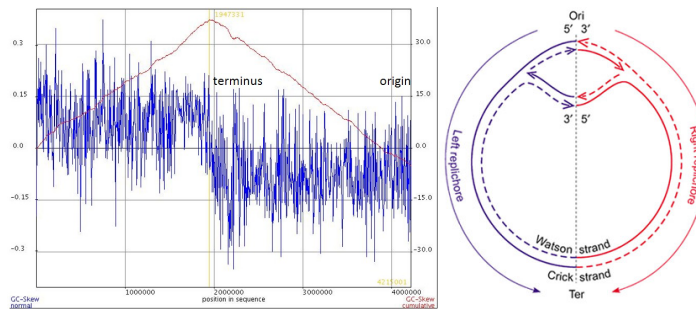


Automatic detection of origin and terminus of replication in bacteria



Bacteria (and some archaea) are replicated **bidirectionally**, from the origin of replication (**Ori**) to the terminus (**Ter**). This property structures the whole genome, impacting – among others – the nucleotide composition and the location and orientation of genes, emphasizing the importance of these two locations.

Since the beginning of the genomic era, researchers have started to use these properties to identify Ori and Ter. The most reliable indicator is the GC-skew, i.e. the difference between the number of Gs and Cs on one strand, but other sources of evidence have been used (some marker sequences, gene orientation, etc.). However, there is currently no software that can use all this to confidently predict the location of both Ori and Ter.

During the exponential growth phase to the respective copy numbers of Ori and Ter changes, with up to 4 Ori copies for 1 Ter copy in *E. coli*. Thanks to the easy access to whole genome sequencing (WGS) data, it has also been possible to use coverage information to help identifying Ori and Ter.

Goal

- Design and write a program that calculates classical metrics (GC-skew, Z-curve) and identifies OriC and dif sequences, if present. Integrates uncertainty assessment.
- From WGS data, design a new metric to locate the origin and terminus of replication.
- Investigate the possibility to use data from NCBI's SRA.
- Integrate all this data in a single predictor for Ori and Ter.

Methods & Material

WGS data from NCBI; existing software that calculates skews and other metrics; general bioinformatics methods. Programming language: Perl, python, R (negotiable).

Requirements

Curiosity and wits indispensable. Skills in **bioinformatics and programming** necessary. Knowledge of **molecular evolution** highly desired.

Useful information

- Start date: open for discussion.
- Contact: Lionel Guy, IMBIM, lionel.guy@imbim.uu.se, 018-471 4246.

References

- Rocha & Touchon, Cold Spring Harb Perspect Biol 2016;8:a018168
<http://cshperspectives.cshlp.org/content/8/1/a018168.short>
- OriLoc: <http://bioinformatics.oxfordjournals.org/content/16/6/560.full.pdf>