



## Sequence Bioinformatics, WS 2019/20

### Project No. 15: Detection of the Replication Origin in Bacterial Genomes

**Challenge of this project:** This task addresses the biological question of determining the origin of replication in a bacterial genome.

**Contact person:** Kay Nieselt

**Summary:** Initiation of replication is not a random but a deterministic process, that always starts at the same position(s) on the chromosome. The origin of replication in a circular bacterial genome (called *oriC*) in most cases starts from a single position, and replication is bidirectional. The *oriC* contains short repeat motifs, one of which serves as the binding motif for the protein DnaA. In this project we ask you to devise a computational approach to locate the replication origin of a genome knowing only the genome's nucleotide sequence. That this is a fascinating topic is for example visible in the latest Science publication, where the authors have computed growth dynamics of bacteria from metagenomics NGS data [1].

#### Prerequisites:

- Java or Python
- Knowledge of pattern matching algorithms, motif search, ...
- R for plotting or other methods for visualisation.

#### Tasks:

1. Compute the skew, defined as the difference between total number of occurrences of G and C in the genome, and draw the skew diagram of the chromosome. Apply it to the genome of *E. coli*, *Vibrio cholerae*, *Thermotoga petrophila*, and *Salmonella enterica*. How does this help to detect the DnaA motif?
2. The main part is to develop and implement an algorithm that allows one to find the DnaA motif, indicating the replication origin. Think about possibilities to visualize your results appropriately.
3. Compute the DnaA motifs for the 4 genomes of above. Find a source to confirm (or possibly not) your findings.
4. Compute the DnaA motifs for further genomes of your choice.
5. Biologists long believed that each bacterial chromosome has only one *oriC*. Find a paper in which this has been shown not to be true.
6. For the bacterium *Wigglesworthia glossinidia* how many minima do you detect in the skew diagram? Do you find DnaA boxes in the vicinity of both? What do you conclude?
7. Write a report in the form of a scientific paper (see general instructions).

#### Literature:

[1] Korem, T., Zeevi, D., Suez, J., *et al.* (2015). Growth dynamics of gut microbiota in health and disease inferred from single metagenomics sample. Science doi: 10.1126/science.aac4812