Leveraging 4f-SAMMY-seq protocol in Prokaryotes as a high-resolution technology to study Nucleoid accessibility

Ugo Maria Iannacchero[1], Elena Chiti[2], Valentina Rosti[1,3], Andrea Vannini[2], Chiara Lanzuolo[1,3], Davide Roncarati[2], Eva Maria Pinatel[1]

[1]Institute of Biomedical Technologies, National Research Council, Segrate, Italy ; [2]Department of Pharmacy and Biotechnology (FaBiT), University of Bologna, Bologna, Italy ;
[3] INGM, National Institute of Molecular Genetics "Romeo ed Enrica Invernizzi", Milan, Italy

• Background

The bacterial chromosome self-organizes into the nucleoid, where Nucleoid-Associated Proteins (NAPs) and RNAs regulate DNA accessibility. While bacterial Hi-C protocols have recently provided insights into Nucleoid organization, their use remains technically demanding and limited to a few bacterial species, due to high-sequencing depth requirements, crosslinking biases and species-specific restriction enzymes. Moreover, the limited achievable resolution (>5 kb) fails to capture gene-level interactions, leaving key aspects of nucleoid organization unresolved.

• A novel epigenomic technology

4f-SAMMY-seq (**4**-**f**ractions **S**equential **A**nalysis of **M**acro**M**olecules accessibilit**Y**) is a high-throughput sequencing technology originally developed in mammalian systems [1]. It relies on the biochemical separation of chromatin into four solubility-based fractions, each reflecting a different level of DNA accessibility. This enables genome-wide profiling of **both open and closed regions**, and has been demonstrated to reconstruct chromatin compartmentalization with Hi-C like resolution and accuracy at lower sequencing depth [2].

• Why 4f-SAMMY-seq in Prokaryotes?

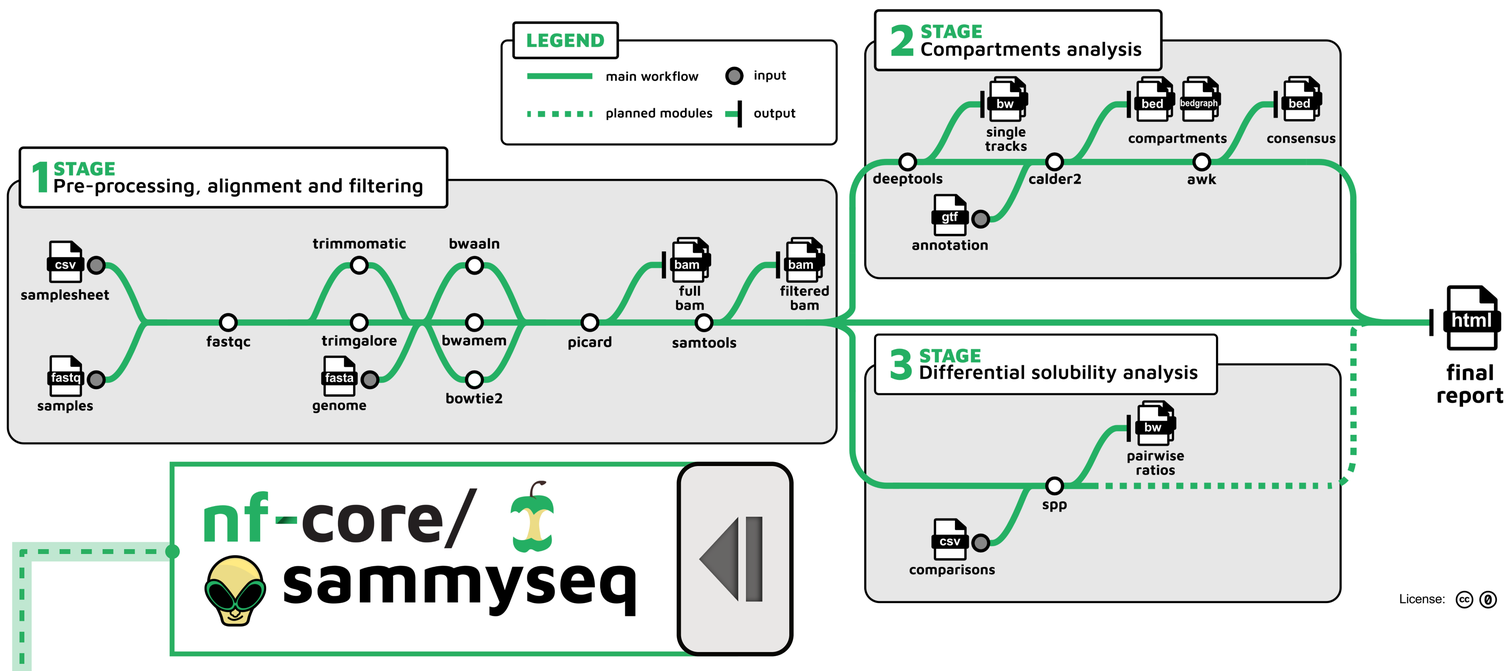
- Broadly **applicable across different bacterial species**.
- **Does not involve cross-linking or restriction enzymes**.
- **Does not rely on the use of antibodies**.
- **Cost-effective** and **fast experimental workflow**.
- Work efficiently at **low sequencing depth** (~0.5–2M reads per fraction).

• Discussion and next steps

We adapted the **nf-core/sammyseq** [3,4] pipeline to **prokaryotes** by integrating read processing **tools better suited for bacterial genomes**, accommodating non-standard chromosome names **ensuring compatibility with Hi-C tools** and **refining genome binning** to support finer-scale resolution through smaller window-sizes. We applied it to **Helicobacter pylori in exponential growth**, where we observed **distinct signal differences between more and less accessible DNA fractions**, and **identified DNA accessibility domains at sub-kilobase resolution** down to the scale of gene-promoters, with as few as **500,000 reads per fraction.**

Ongoing work will focus on finalizing the integration of the Differential Solubility analysis module into the pipeline, with further adaptation for prokaryotic genomes. Additionally, we aim to enhance data interpretation by incorporating KEGG pathway enrichment modules to better understand the functional impact of nucleoid accessibility changes.

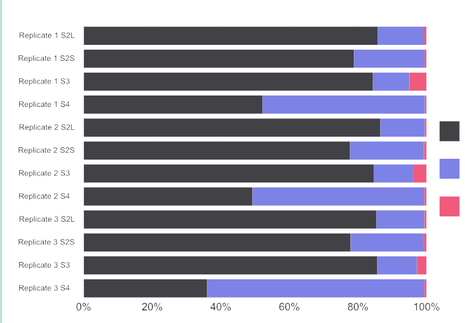
Methods of analysis



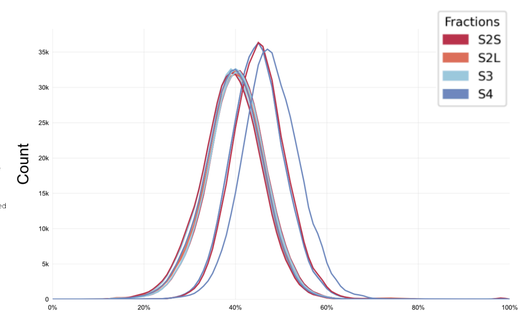
Analytical output

1 STAGE Pre-processing, alignment and filtering

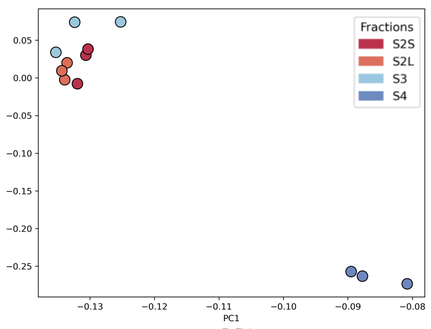
Raw sequencing reads from each solubility fraction are aligned to the reference genome, with quality control metrics including **GC content** and **mapping quality** assessed. To evaluate sample similarity and variability, **Principal Component Analysis (PCA)** is performed on the processed coverage profiles.



Mapping quality



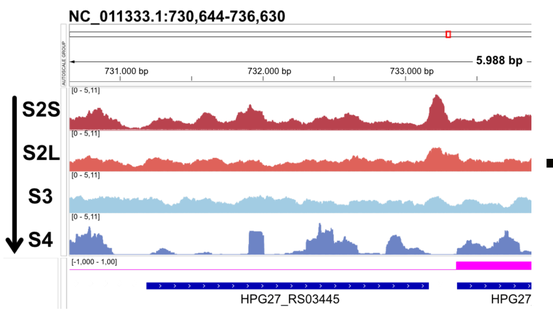
GC content



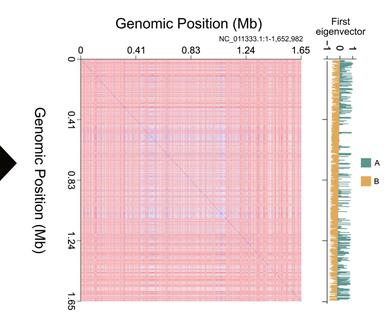
PCA

2 STAGE Compartments analysis (Hi-C like)

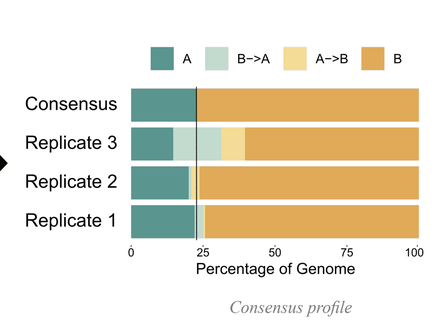
Genome-wide coverage profiles are generated for each fraction, and cross-correlation values among the four fractions are computed using Hi-C analysis tools to generate an **NxN correlation matrix** computed over user-defined genomic bin size. Eigenvector decomposition of this matrix enables identification of genome-wide compartmental reorganizations, such as transitions between active and inactive DNA domains, leading to the **generation of a consensus profile across replicates**.



Genomic tracks



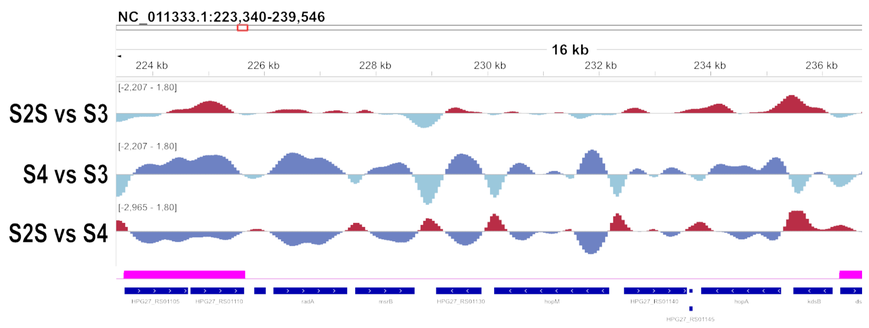
Correlation Matrix



Consensus profile

3 STAGE Differential solubility analysis (ChIP-seq like)

Pairwise relative ratios between the most representative fraction tracks are calculated to detect continuous changes in DNA accessibility, enabling identification of gene-level alterations.



Pairwise comparison tracks (ratios)

References

1. Sebestyén, E., Marullo, F., Lucini, F., Petrini, C., Bianchi, A., Valsoni, S., Olivieri, I., Antonelli, L., Gregoretti, F., Oliva, G., Ferrari, F., & Lanzuolo, C. (2020). SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome. Nature Communications, 11(1), 6274. https://doi.org/10.1038/s41467-020-20048-9
2. Lucini, F., Petrini, C., Salviato, E., Pal, K., Rosti, V., Gorini, F., Santarelli, P., Quadri, R., Lembo, G., Graziano, G., Di Patrizio Soldateschi, E., Tagliaferri, I., Pinatel, E., Sebestyén, E., Rotta, L., Gentile, F., Vaira, V., Lanzuolo, C., & Ferrari, F. (2024). Biochemical properties of chromatin domains define genome compartmentalization. Nucleic Acids Research, 52(12), e54–e54. https://doi.org/10.1093/nar/gkae45
3. Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A.,Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology, 38(3), 276–278. https://doi.org/10.1038/s41587-020-0439-x
4. https://nf-co.re/sammyseq/dev/ (https://github.com/nf-core/sammyseq/tree/dev)

Contacts

ugo.iannacchero@itb.cnr.it
@ugoiann.bsky.social

FEMSMICROMilan 2025:Congress & Exhibition • **POSTER N. 422.15P** • This project was funded by **PRIN 20223BM3NP**