

GMS-Artic: A containerized Nextflow pipeline for detecting Pangolin typing by analysing SARS-CoV-2 whole genome from short and long sequence reads

Jyotirmoy Das^{1*}, Sofia Stamouli², Tanja Normark², Henning Onsbring², Isak Sylvén³

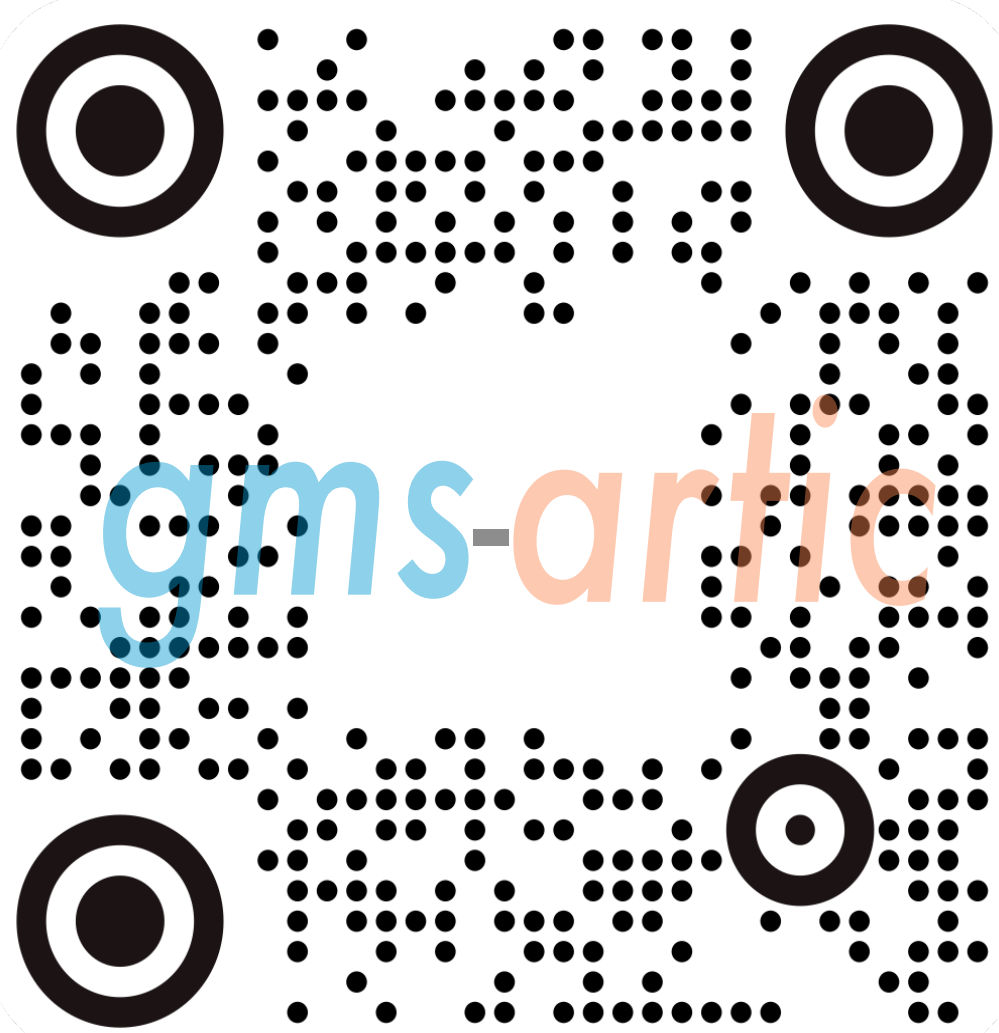
Affiliations:

¹Clinical Genomics Linköping, SciLife Laboratory, Department of Biomedical and Clinical Sciences, Linköping University, Sweden

²Clinical Genomics Stockholm, Karolinska Institutet, Science for Life Laboratory, Department of Microbiology, Tumour and Cell Biology, Stockholm, Sweden

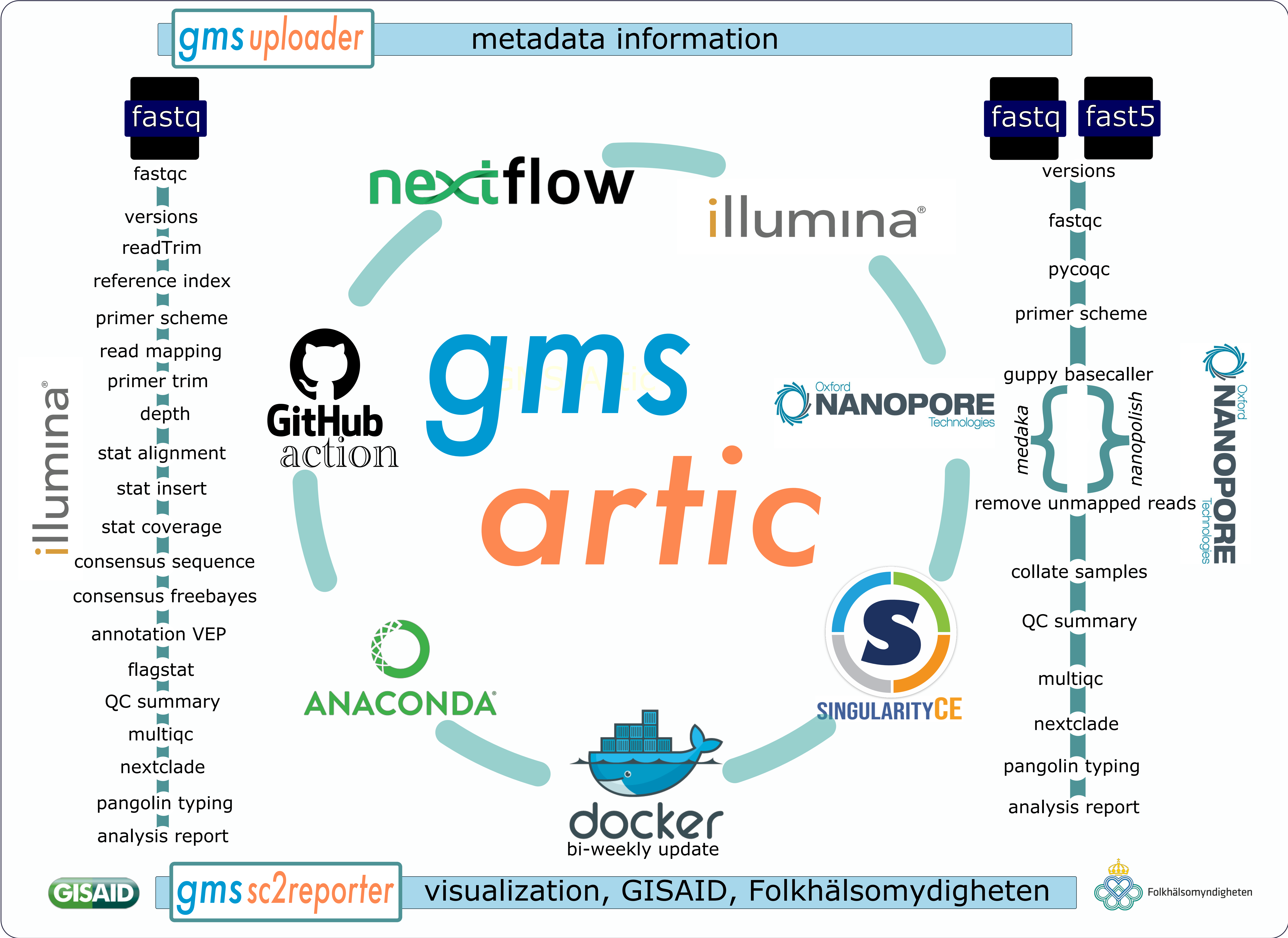
³Clinical Genomics Gothenburg, Gothenburg University, Sweden

*Corresponding: jyotirmoy.das@liu.se



PURPOSE: Development of unified nationwide rapid bioinformatics pipeline for SARS-CoV-2 whole genome sequence analysis using both Illumina® short-read and Nanopore® long-read sequencing data.

Implementation: We improved Conner *et al* (2019) *nCoV2019-artic-nf* semi-automated Nextflow pipeline to execute on the SGE cluster using the docker container with Singularity process caller. To inspect the Illumina® sequencing reads, we applied distinctive quality processes necessary for clinical practices and enhanced them with the freebayes algorithm for sequence depth calculation and variant calling. The pipeline restructured for Nanopore® artic and midnight protocols with customized primer recognition support. Both Illumina® and Nanopore® data analyses reinforced by Nextclade and Pangolin typing. An automated workflow designed to update the Pangolin typing version once every two weeks.



Development(s):

1. Separate container for Illumina®, Nanopore®, pangolin
2. Singularity, MPI support
3. additional statistical analyses
4. Nextclade, Pangolin typing detection
5. biweekly updated Pangolin typing container
6. Integration of metadata and automated uploading FoHM & GISAID.

Conclusion:

The GMS-Artic aims to develop a nationwide standard open-source bioinformatic workflow to investigate SARS-CoV-2 whole genome sequencing for real-time clinical practices on the **National Genomic Platform (NGP)**. A parallel work using the GMS-Artic framework designed to update all metadata information from the clinics and upload the output result to the Swedish Health Agency routinely. Continuous support on the pipeline development and an additional process like variant visualization is currently in progress.

Reference: <https://github.com/connor-lab/ncov2019-artic-nf>

Acknowledgement: The authors would like to acknowledge Clinical Genomics Linköping, Science for Life Laboratory, Sweden, Clinical Genomics Stockholm, Karolinska Institutet, Science for Life Laboratory, Department of Microbiology, Tumour and Cell Biology, and Clinical Genomics Gothenburg, Gothenburg University, Sweden for their support on this application development.