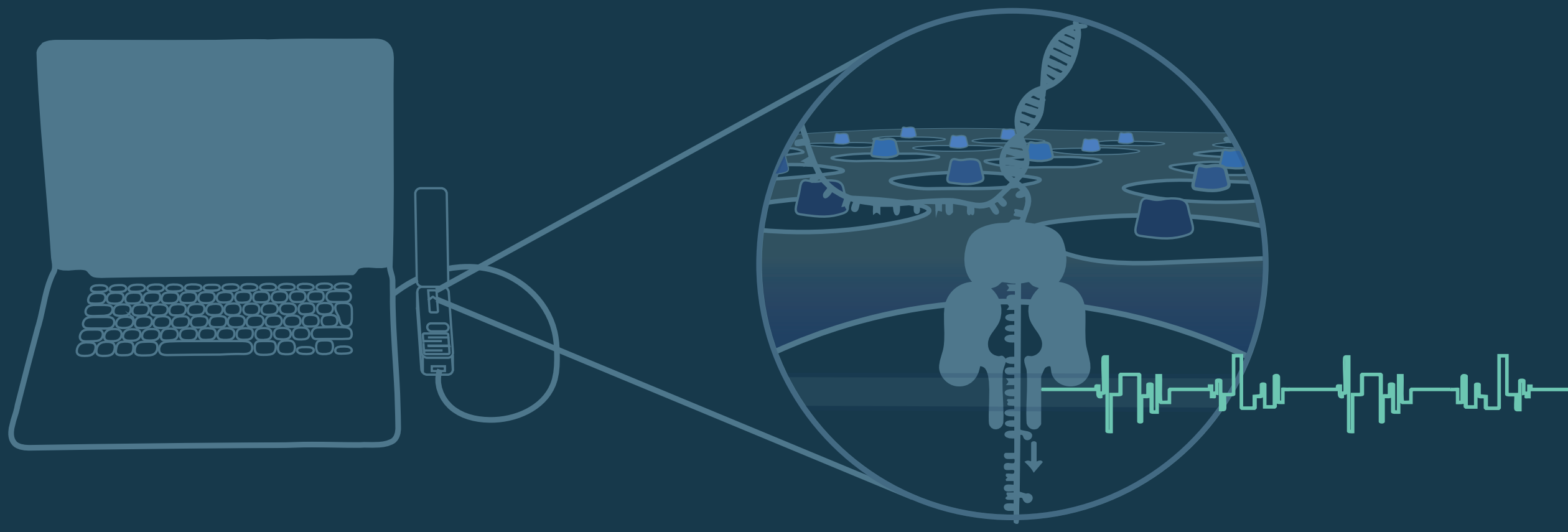


ESTABLISHING A REPRODUCIBLE BASELINE FOR SQUIGGLE-SPACE ALGORITHM DEVELOPMENT

Laura Raes, Wim L. Cuypers, Halil Ceylan, Eline Turcksin, Julia M. Gauglitz, Wout Bittremieux, Kris Laukens

Analysis at the level of raw ionic current signals



Our research focuses on **raw ionic current signals** (“squiggles”) instead of basecalled sequences.

Why?

1. squiggle data provides the **most direct view of the sequencing process**, including epigenetic and kinetic information
2. avoiding basecalling enables **computationally lighter applications**, such as portable pathogen detection

SquiDBase: a centralized FAIR repository of raw microbial and viral signals



www.SquiDBase.org
doi.org/10.1093/nargab/lqaf213

SQUIGGLE DATA

SquiDBase currently contains 2.5 TB *pod5* files, spanning 40 species. The data is **annotated** with taxonomic IDs, library sources, input types, etc; following general ontologies.

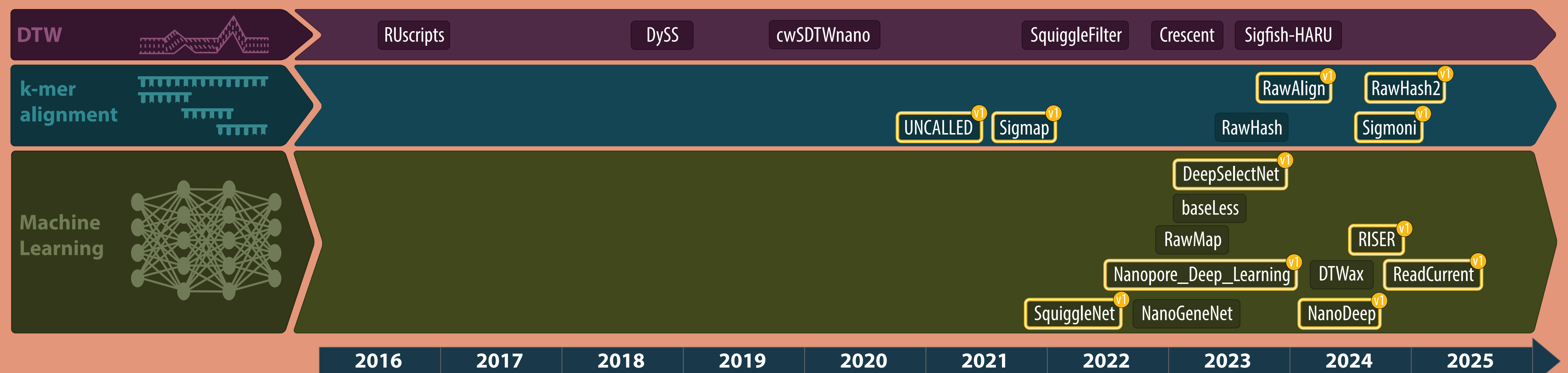
The raw signals are paired with basecalled *bam* files. These are **updated periodically**, using Oxford Nanopore Technologies' latest SUP basecalling models.

SCIENTIFIC APPLICATIONS

- raw signal algorithm development
- microbial genetics and epigenetics research
- supporting reproducible science by providing a repository for the most fundamental data layer in nanopore sequencing

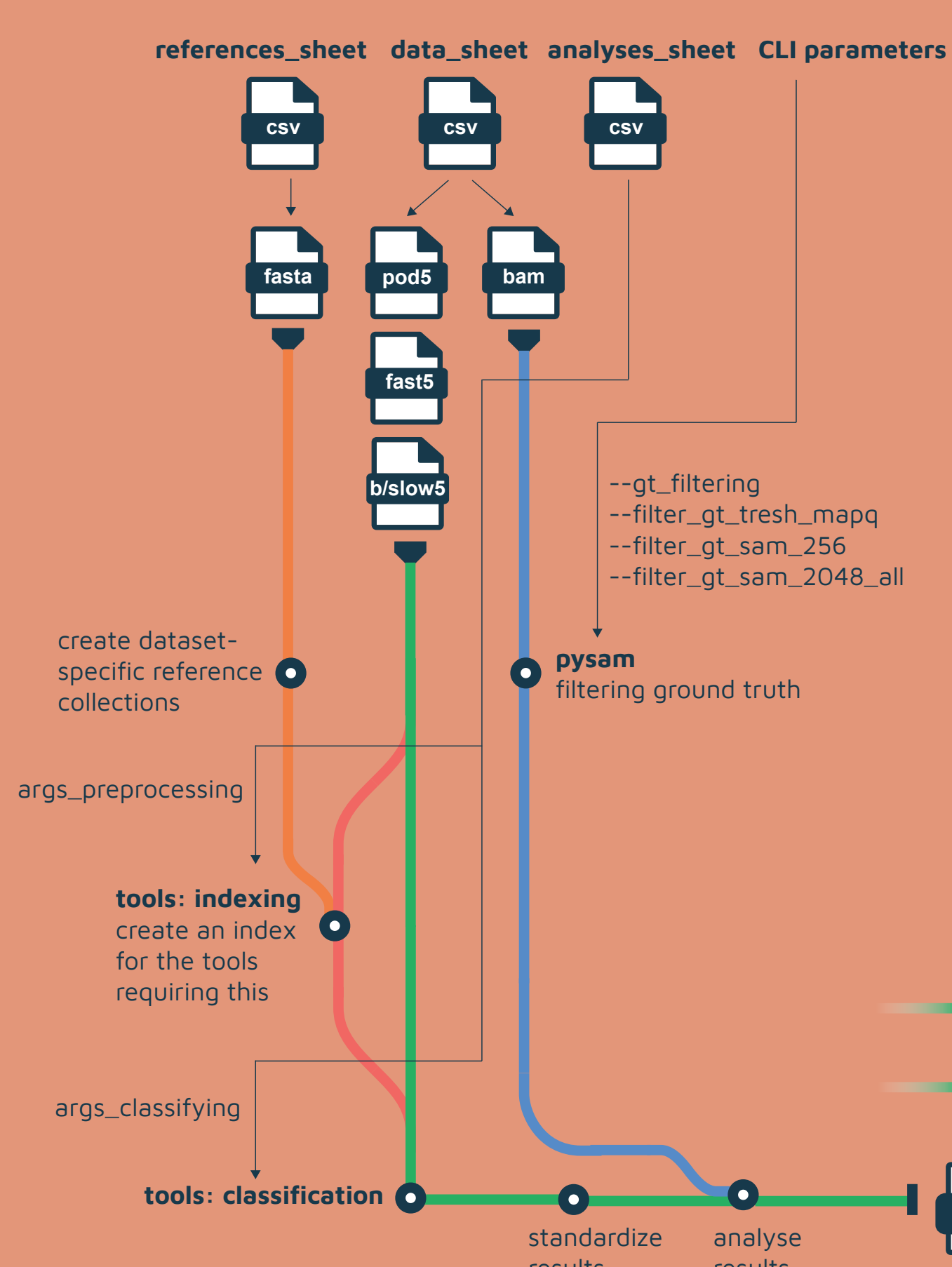


SquiDBench: systematic benchmarking of squiggle-space classification and mapping algorithms



The tools included in the first release of SquiDBench are marked with v1.

PIPELINE ARCHITECTURE

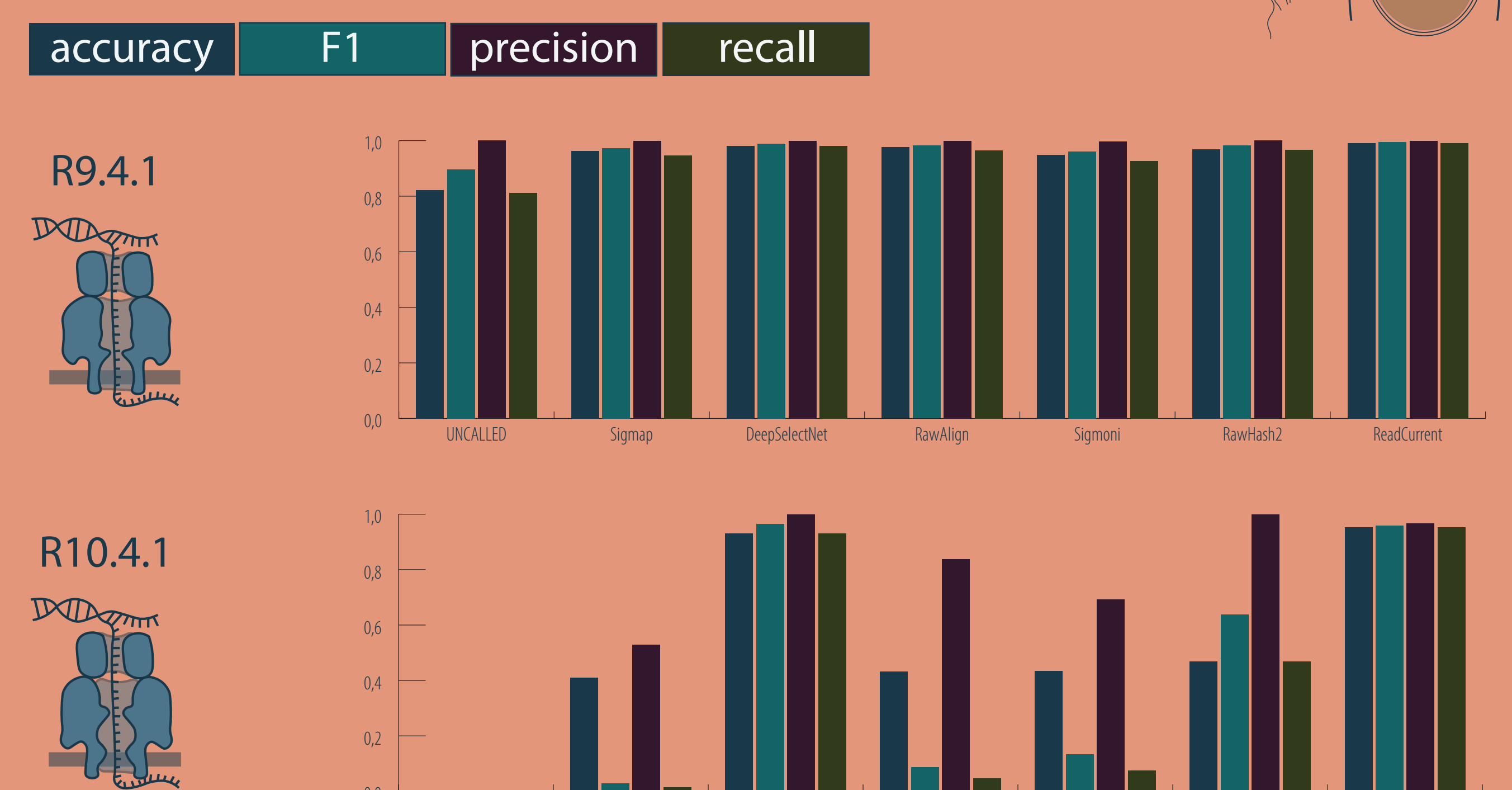


The SquiDBench benchmarking pipeline is implemented in Nextflow, allowing for reproducible execution across different environments. All tools are packaged in Docker containers, with Dockerfiles available in an open-source repository and images accessible via Docker Hub.

New tools can be added with minimal changes and evaluated within the same framework, allowing consistent benchmarking as methods evolve.

PRELIMINARY RESULTS

Binary classification of *E. coli* vs *C. reinhardtii*.



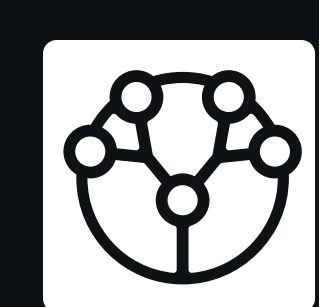
These preliminary results indicate that tools based on machine learning adapt more easily to the R10 sequencing chemistry update than k-mer alignment based tools.

Future perspectives

We are expanding SquiDBase into a **knowledge base** and a **benchmarking hub** for signal-level algorithms. Furthermore, we aim to develop a lightweight squiggle-based classification algorithm for **portable pathogen detection**.



QR code links to a web page containing the poster, references, abstract and additional information.



University of Antwerp
Adrem | Adrem Data Lab