## ONT-HybriFlow: Long-read first hybrid assembly pipeline

## Overview

This pipeline integrates Nanopore and short-read sequencing data processing, offering a streamlined approach from assembly to quality assessment. It utilizes Flye for assembly, Medaka and Polypolish for error correction and polishing, and QUAST for assessing assembly quality.

## Requirements

* Flye,
* Medaka,
* Polypolish,
* QUAST,
* BWA and
* Samtools must be installed and accessible in the directory where you are executing the scripts.

# Installation

1. Copy the git link in from github page

git clone “git\_link”

## Usage

Execute the pipeline with:

./master\_script.sh <nanopore\_fastq\_dir> <short\_reads\_dir> <output\_path> <threads> <scripts\_path>

Parameters include directories for Nanopore FASTQ and short-read data, the output path, the number of processing threads, and the path to scripts.

## Scripts Overview

1. flye.sh: Assembles genomes using Flye.
2. medaka.sh: Applies Medaka for ONT-based error correction.
3. polypolish.sh: Further polishes assemblies using short-read with Polypolish.
4. quast.sh: Assesses assembly quality with QUAST.
5. quast-summary.sh: Summarizes QUAST metrics in a CSV file.
6. Ensure these scripts are executable (`chmod +x` “script\_name”).

## Output Description

Outputs are organized into directories corresponding to each process step:

1. Flye (Path: output\_path/Results/flye): Contains ‘assembly.fasta’ files from the Flye assembler.
2. Medaka (Path: output\_path/Results/medaka): Stores error-corrected assemblies from Medaka.
3. Polypolish sequences (Path: output\_path/Results/polypolish): Holds final polished genome.
4. QUAST (Path: output\_path/Results/quast): Includes detailed assembly quality reports.
5. QUAST Summary: Features a `final\_assembly\_metrics.csv` file summarizing key QUAST metrics.

This structured output facilitates easy access to detailed assembly data and quality metrics.