




AliNe: A Flexible and Efficient Nextflow Pipeline for Read Alignment

Jacques Dainat ¹

¹ Joint Research Unit for Infectious Diseases and Vectors Ecology Genetics Evolution and Control (MIVEGEC), University of Montpellier, French National Center for Scientific Research (CNRS 5290), French National Research Institute for Sustainable Development (IRD 224), 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France.

DOI: [10.21105/joss.07545](https://doi.org/10.21105/joss.07545)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: Julia Romanowska 

Reviewers:

- [@rcannood](#)
- [@gchure](#)

Submitted: 27 September 2024

Published: 05 March 2025

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

Alignment of sequencing reads against a reference genome is a fundamental step in many bioinformatics workflows. Aligner performance varies by speed, memory efficiency, and accuracy, with some tailored to specific sequencing technologies and others more versatile, which makes the selection of an appropriate tool context-dependent. To streamline this process, we present AliNe (Alignment in Nextflow), a flexible and efficient read alignment pipeline built on the Nextflow framework ([Tommaso et al., 2017](#)). AliNe contains a broad range of commonly used aligners and is designed to accommodate any high-throughput sequencing projects.

AliNe supports short reads (both paired-end and single-end) as well as long reads generated by PacBio and Oxford Nanopore Technologies (ONT). It is designed to minimize user inputs and avoid common parameter mistakes (e.g., scoring system, strandedness, orientation). Currently, AliNe supports 16 widely used alignment tools, including BMap ([Bushnell, 2014](#)), Bowtie ([Langmead et al., 2009](#)), Bowtie2 ([Langmead & Salzberg, 2012](#)), BWA ([Li & Durbin, 2009](#)), BWA-MEM ([Li, 2013](#)), BWA-SW ([Li & Durbin, 2010](#)), GraphMap2 ([Marić et al., 2019](#)), HISAT2 ([Kim et al., 2019](#)), Kallisto ([Bray et al., 2016](#)), Minimap2 ([Li, 2021](#)), ngmlr ([Sedlazeck et al., 2018](#)), novoAlign ([novoAlign / Novocraft, n.d.](#)), nucmer ([Marçais et al., 2018](#)), STAR (single or two-pass mode) ([Dobin et al., 2013](#)), subread ([Liao et al., 2013](#)), and sublong ([Liao et al., 2013](#)). These aligners are integrated into a single, easy-to-use workflow, providing a unified entry point for any project requiring alignment and giving users the flexibility to choose the tool that best suits their specific data and objectives. To facilitate aligner comparison, AliNe provides a comprehensive set of metrics, including computational resource usage (CPU, memory, runtime, and disk I/O) reported by Nextflow, as well as alignment statistics generated by FastQC ([FastQC, 2015](#)) and Samtools stats ([Petr Danecek, 2021](#)).

Statement of Need

The complexity of sequencing data analysis, especially with the advent of next-generation sequencing (NGS), has created a demand for robust, scalable, and reproducible workflows that can handle large datasets while allowing users to flexibly choose their preferred analysis tools. Numerous bioinformatics analyses rely on the quality of read alignment, having a major role in the reliability of the downstream results. Each alignment tool has its own strengths and weaknesses. Users may need to try out several alignment tools in order to achieve optimal results specific to the project's requirements. AliNe empowers researchers with a single-entry point for alignment, eliminating the need to juggle between disparate tools. By integrating multiple aligners into a unified pipeline, AliNe simplifies the process of alignment algorithm selection and establishes a standardized workflow that ensures both reproducibility and transparency across analyses.

Majority of pipelines integrating read alignment steps are tailored to specific sequencing technologies and types of analyses (Riesgo-Ferreiro et al., 2022) (Patel et al., 2024) (Garcia et al., 2024) (Chafin et al., 2024). There remains a lack of unified, versatile workflows capable of integrating aligners into a single, user-friendly platform. AliNe addresses this challenge by providing a comprehensive, streamlined solution for alignment tasks across a wide range of sequencing technologies, whatever the field of application. Furthermore, leveraging the concept of daisy-chaining Nextflow workflows (Mahesh Binzer-Panchal, 2024), AliNe can be integrated as a sub-workflow within other Nextflow pipelines (Jacques Dainat, 2025), providing a standardized alignment step while offering a diverse selection of alignment tools.

Workflow Overview

AliNe takes sequencing reads in FASTQ format and aligns them to a user-supplied reference genome in FASTA format, using user-selected aligners. The workflow is fully configurable to accommodate various data types and gives users the ability to fine-tune tool alignment parameters independently. Nevertheless, AliNe is designed to minimize user inputs and avoid common mistakes like the definition of scoring system used by the sequencing machine (Phred+33, Phred+64, and Solexa) or, in the case of RNAseq data, the strandedness (stranded or unstranded) or the orientation in case of paired-end reads (forward reverse i.e., fr; reverse forward i.e., rf; forward forward i.e., ff). If not provided by the user, AliNe will deduce this information and adapt the option of the different downstream aligners accordingly. AliNe suggests also a proper aligner, depending on the input data provided.

To run AliNe you need to select one or several aligners and specify the type of input data: short single-end reads, short paired-end reads, long reads generated by PacBio or long reads generated by Oxford Nanopore Technologies (ONT). AliNe integrates an optional read trimming step using Fastp (Chen, 2023). It also includes optional quality control at every stage with FastQC (FastQC, 2015). For RNA-seq data, the pipeline automatically detects the library type (strandedness and orientation) via Salmon (Patro et al., 2017) and configures the appropriate parameters for downstream alignment tools accordingly. Additionally, AliNe uses SeqKit (Shen et al., 2016) to detect the sequence quality scoring system and standardizes it to Phred+33 if necessary. Once preprocessing is complete, alignment is performed using the aligners selected by the user. Finally, with the help of Samtools (Petr Danecek, 2021), AliNe outputs the aligned reads in sorted BAM format, ready for downstream analyses such as variant calling, gene expression quantification, or structural variation detection. The final stage resumes all FastQC analyses and logs output from compatible aligners using MultiQC (Ewels et al., 2016). All these steps are summarized in Figure 1.

The pipeline supports parallel execution, ensuring efficient processing of large datasets by distributing tasks across available CPU cores. With Nextflow's inherent support for reproducibility, users can confidently rerun analyses with consistent results.

The pipeline is modular, enabling future expansion and the integration of additional tools or customization of downstream analysis steps. Users can adapt AliNe to their specific research needs, whether for RNA-seq, ChIP-seq, or whole-genome resequencing projects.

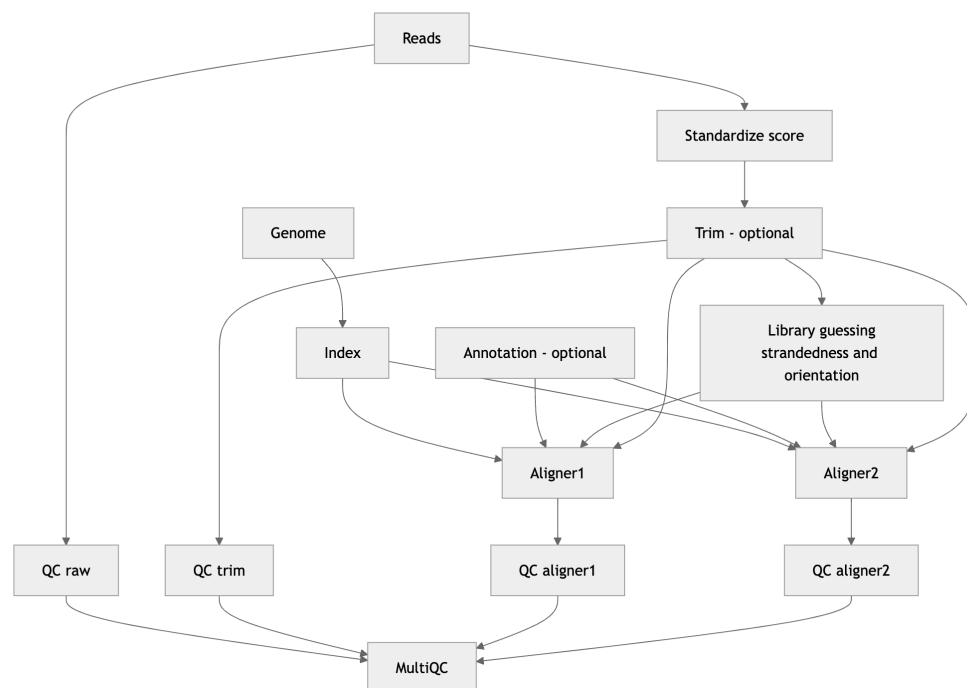


Figure 1: Overview of the different steps implemented in AliNe.

Availability

AliNe is freely available under an open-source license. The source code, along with detailed documentation and examples, can be found at <https://github.com/Juke34/AliNe>. Contributions and community feedback are welcome.

References

- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34:5, 34, 525–527. <https://doi.org/10.1038/nbt.3519>
- Bushnell, B. (2014). *BBMap: A fast, accurate, splice-aware aligner*. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). <https://www.osti.gov/biblio/1241166>
- Chafin, T., Babirye, S. R., Duong, C., Muffato, M., Qi, G., & Surana, P. (2024). *Sanger-tol/readmapping: Antipodean opaleye (patch 1)* (Version 1.3.1). Zenodo. <https://doi.org/10.5281/zenodo.13837834>
- Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, 2. <https://doi.org/10.1002/imt2.107>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32, 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- FastQC. (2015). <https://qubeshub.org/resources/fastqc>

- Garcia, M. U., Hanssen, F., Pedersen, A. S., Gabernet, G., WackerO, SusiJo, Talbot, A., bot, nf-core, James, C., Ergüner, B., Peltzer, A., nicksmith, Navarro, J. F., Miller, E., Pedrol, J. M., Lescai, F., Neilson, G., Pearce, S., Martínez, F., ... Lärkeryd, A. (2024). *Nf-core/sarek: 3.5.0 - áhkájiegná* (Version 3.5.0). Zenodo. <https://doi.org/10.5281/zenodo.14412016>
- Jacques Dainat. (2025). *Daisy chaining AliNe*. <https://github.com/Juke34/daisy-chaining-AliNe>
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 2019 37:8, 37, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9, 357–359. <https://doi.org/10.1038/NMETH.1923>
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10, 1–10. <https://doi.org/10.1186/GB-2009-10-3-R25>
- Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. <https://arxiv.org/abs/1303.3997>
- Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, 37, 4572–4574. <https://doi.org/10.1093/BIOINFORMATICS/BTAB705>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25, 1754–1760. <https://doi.org/10.1093/BIOINFORMATICS/BTP324>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics (Oxford, England)*, 26, 589–595. <https://doi.org/10.1093/BIOINFORMATICS/BTP698>
- Liao, Y., Smyth, G. K., & Shi, W. (2013). The subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, 41. <https://doi.org/10.1093/NAR/GKT214>
- Mahesh Binzer-Panchal. (2024). *A proof of concept daisy-chaining nextflow workflows*. <https://github.com/mahesh-panchal/nf-cascade/>
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, 14. <https://doi.org/10.1371/JOURNAL.PCBI.1005944>
- Marić, J., Sović, I., Križanović, K., Nagarajan, N., & Šikić, M. (2019). Graphmap2 - splice-aware RNA-seq mapper for long reads. *bioRxiv*, 720458. <https://doi.org/10.1101/720458>
- novoAlign* / *novocraft*. (n.d.). <https://www.novocraft.com/products/novoalign/>
- Patel, H., Ewels, P., Manning, J., Garcia, M. U., Peltzer, A., Hammarén, R., Botvinnik, O., Talbot, A., Sturm, G., bot, nf-core, Zepper, M., Moreno, D., Vemuri, P., Binzer-Panchal, M., Greenberg, E., silviamorins, Pantano, L., Syme, R., Kelly, G., ... Gabernet, G. (2024). *Nf-core/rnaseq: Nf-core/rnaseq v3.18.0 - lithium lynx* (Version 3.18.0). Zenodo. <https://doi.org/10.5281/zenodo.14537300>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14, 417–419. <https://doi.org/10.1038/nmeth.4197>
- Petr Danecek, J. L., James K Bonfield. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10. <https://doi.org/10.1093/gigascience/giab008>
- Riesgo-Ferreiro, P., tronbfxdev, & Mohr, J. (2022). *TRON-bioinformatics/tronflow-alignment:*

v1.8.0 (Version v1.8.0). Zenodo. <https://doi.org/10.5281/zenodo.7219617>

Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Haeseler, A. V., & Schatz, M. C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods*, 15, 461–468. <https://doi.org/10.1038/s41592-018-0001-7>

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE*, 11. <https://doi.org/10.1371/journal.pone.0163962>

Tommaso, P. D., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35, 316–319. <https://doi.org/10.1038/nbt.3820>