



# NanoForms: an integrated server for processing, analysis and assembly of raw sequencing data of microbial genomes, from Oxford Nanopore technology

Anna Czmil<sup>1,\*</sup>, Michal Wronski<sup>1,\*</sup>, Sylwester Czmil<sup>1</sup>, Marta Sochacka-Pietal<sup>2</sup>, Michal Cmil<sup>1</sup>, Jan Gawor<sup>3</sup>, Tomasz Wołkiewicz<sup>4</sup>, Dariusz Plewczynski<sup>5,6</sup>, Dominik Strzalka<sup>1</sup> and Michal Pietal<sup>1</sup>

<sup>1</sup> Department of Complex Systems, Rzeszow University of Technology, Rzeszow, Subcarpathian, Poland

<sup>2</sup> Department of Biotechnology and Bioinformatics, Rzeszow University of Technology, Rzeszow, Subcarpathian, Poland

<sup>3</sup> DNA Sequencing and Oligonucleotide Synthesis Laboratory, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Masovian, Poland

<sup>4</sup> Department of Bacteriology and Biocontamination Control, National Institute of Public Health-National Institute of Hygiene, Warsaw, Masovian, Poland

<sup>5</sup> Laboratory of Functional and Structural Genomics, Centre of New Technologies, University of Warsaw, Warsaw, Masovian, Poland

<sup>6</sup> Laboratory of Bioinformatics and Computational Genomics, Warsaw University of Technology, Warsaw, Masovian, Poland

\* These authors contributed equally to this work.

## ABSTRACT

**Background.** Next Generation Sequencing (NGS) techniques dominate today's landscape of genetics and genomics research. Though Illumina still dominates worldwide sequencing, Oxford Nanopore is one of the leading technologies currently being used by biologists, medics and geneticists across various applications. Oxford Nanopore is automated and relatively simple for conducting experiments, but generates gigabytes of raw data, to be processed by often ambiguous set of alternative bioinformatics command-line tools, and genomics frameworks which require a knowledge of bioinformatics to run.

**Results.** We established an inter-collegiate collaboration across experimentalists and bioinformaticians in order to provide a novel bioinformatics tool, free for academics. This tool allows people without extensive bioinformatics knowledge to simply process their raw genome sequencing data. Currently, due to ICT resources' maintenance reasons, our server is only capable of handling small genomes (up to 15 Mb). In this paper, we introduce our tool, NanoForms: an intuitive and integrated web server for the processing and analysis of raw prokaryotic genome data, coming from Oxford Nanopore. NanoForms is freely available for academics at the following locations: <http://nanofoms.tech> (webserver) and <https://github.com/czmilanna/nanofoms> (GitHub source repository).

**Subjects** Bioinformatics, Genomics, Microbiology, Computational Science

**Keywords** NGS, Bioinformatics, Oxford Nanopore, Genomics, Webserver, DNA sequencing, DNA assembly, Microbial genomes

Submitted 23 August 2021  
Accepted 13 February 2022  
Published 29 March 2022

Corresponding author  
Michal Pietal, [m.pietal@prz.edu.pl](mailto:m.pietal@prz.edu.pl)

Academic editor  
Adam Witney

Additional Information and  
Declarations can be found on  
page 9

DOI 10.7717/peerj.13056

© Copyright  
2022 Czmil et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

## INTRODUCTION

Next Generation Sequencing technologies, such as Illumina ([Gloor et al., 2010](#)), Pacific BioSciences ([Rhoads & Au, 2015](#)) and Oxford Nanopore ([Jain et al., 2016](#)), dominate today's landscape of genetics and genomics research. Each technology has its advantages, disadvantages and market share in specific applications and niches. The applications of Oxford Nanopore technology include eDNA extraction and sequencing (*i.e.*, [Garlapati et al., 2019](#)), rapid viral sequencing (including the current challenge of SARS-CoV2 ([Wang et al., 2020](#))), human genome comparative sequencing ([De Coster et al., 2019](#)) and many others. Short-read sequencing technologies such as Illumina have made bacterial genome sequencing relatively cheap and accessible. However, the procedure of closing microbial genomes is often costly and laborious. Assembly of short reads from genomes that are repetitive and/or have extreme %GC content remains challenging. These difficulties can be mostly overcome by using single-molecule, long-read sequencing technologies such as the Oxford Nanopore. Nanopore helps with closing bacterial genomes ([Risse et al., 2015](#); [Kawalek et al., 2020](#)), can deliver two strategies for bacterial genome assembly ([Goldstein et al., 2019](#)), even helps to obtain complete bacterial chromosomes from microbiomes ([Moss, Maghini & Bhatt, 2020](#)) or is used in routine microbial genome sequencing ([Wick et al., 2017a](#)). Two main strategies are used to assemble bacterial genomes using long read sequencing. In the first, nanopore reads are used for long read only genome assembly followed by polishing with Illumina reads. Alternatively, long reads are used to enhance genome assemblies that are generated from short-read Illumina data. In such case nanopore reads can scaffold contigs generated by short read sequencing. It is also now possible to extract 3D structures of the genome using Oxford Nanopore ([Ulahannan et al., 2019](#)).

Oxford Nanopore is relatively user-friendly, easily operated, inexpensive, and can be simply adjusted to allow for rapid sample processing (including outdoor usage). The downside is, after the experiment is done, researchers are left with a huge amount of raw data. There is a wide variety of software tools available to perform taxonomic classification of the raw data. There are also many comparative studies that evaluate the top performing bioinformatics tools, provide recommendations for use cases, and show how to run these tools ([McIntyre et al., 2017](#); [Escobar-Zepeda et al., 2018](#); [Simon et al., 2019](#)). An inexperienced user, however, could easily become overwhelmed by the complexity of the data, the fast pace of tool development, and the version updates, command changes, installation problems, etc.

## MATERIALS & METHODS

We used the following technologies to create the NanoForms server: Python language, Linux/UNIX/BSD operating system, Django application server, Workflow Description Language and Cromwell, Crontab, Docker and BioContainers ([Da Veiga Leprevost et al., 2017](#)), and a custom set of bioinformatics tools. The NanoForms server is freely available for academic use and a commercial release of the server (for non-academics, businesses, etc.) is planned. We also provide its source code (under GPLv3 license) for non-commercial uses. The server is fully virtualized, with about 30 processor cores and 120 GB RAM available

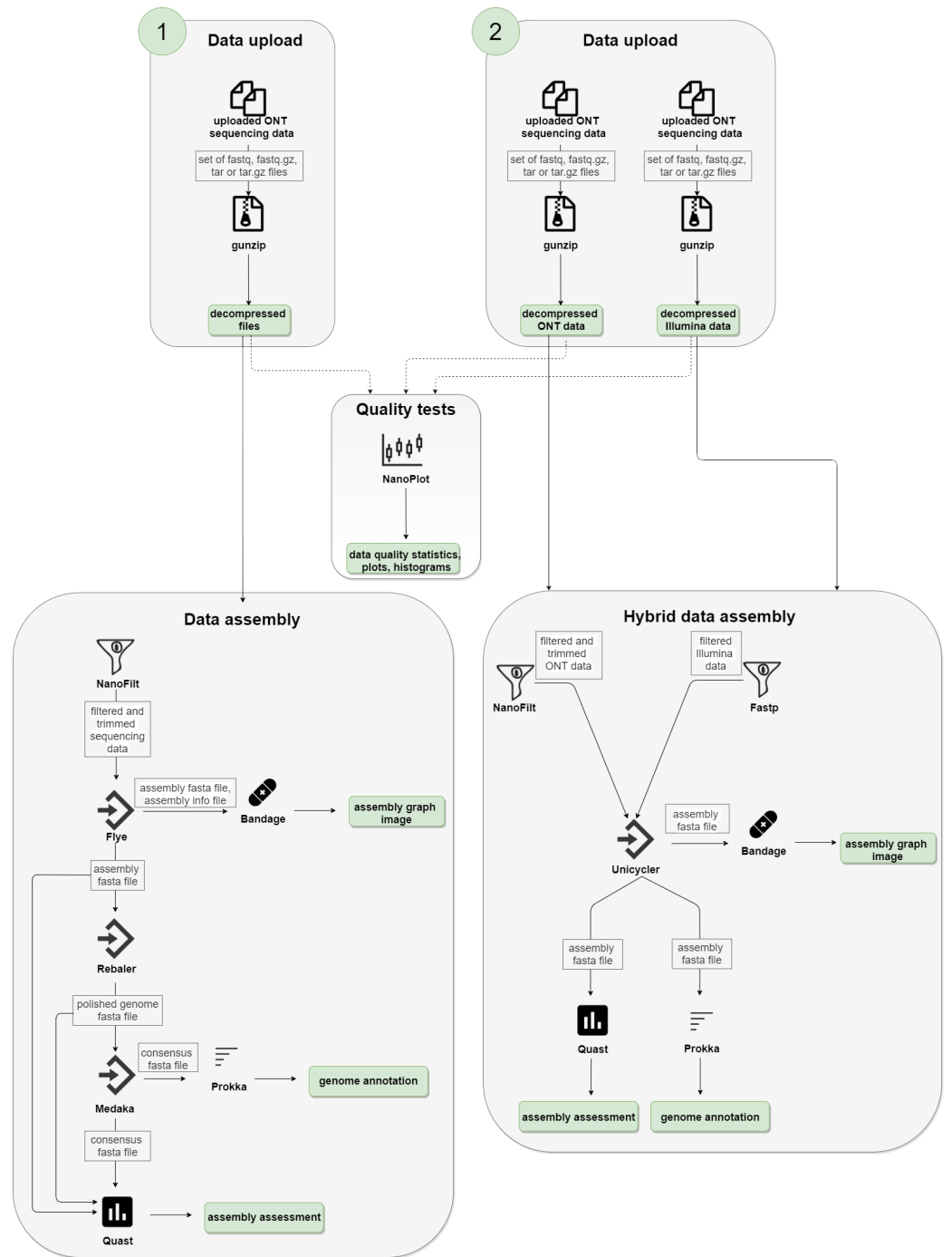
on average. The infrastructure is also hosted in a virtual environment. It can handle about 5-10 parallel jobs (taking into account dataset size limits of 15 GB). In general, single run processing time depends on configuration, sample size and the assembly type (short reads, long reads or both). It takes nearly two hours to process a 220 MB sample of ONT data, yielding the results. Computation time grows in a near linear manner, with a 1 GB dataset taking about five hours. Hybrid assemblies, which include a Unicycler polishing stage, can take up to 10 h for a 1 GB dataset combined (300 MB of ONT and 600 MB of Illumina data). However, actual run times will depend on server usage, so for complete control over timings one can install a local version of nanoforms. The detailed diagram of the server workflow is shown in Fig. 1.

The current version of the server includes the use of the following bioinformatics applications: Nanoplot 1.32.0 NanoFilt 2.7.1 (*De Coster et al., 2018*), FastQC 0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>), Flye 2.8.1 (*Lin et al., 2016*), Bandage 0.8.1 (*Wick et al., 2015*), Rebaler 0.2.0 (<https://github.com/trwick/Rebaler>), Medaka 1.0.1 (<https://github.com/nanoporetech/medaka>), Quast 3.2 (*Mikheenko et al., 2018*), Fitlong 0.1.0 (<https://github.com/trwick/Fitlong>), Prokka 1.14.6 (*Seemann, 2014*), Kraken Tools 0.1 (*Davis et al., 2013*), Kraken 2.1.0 (*Wood, Lu & Langmead, 2019*) and Krona 2.7.1 (*Ondov, Bergman & Phillippy, 2011*). For hybrid genome assembly, we use Unicycler 0.3.0b (*Wick et al., 2017b*) and Fastp 0.18.0 (*Chen et al., 2018*) for filtering data from Illumina. A hybrid assembly strategy has been developed to overcome the limitations of both Illumina and Oxford Nanopore sequencing and to unlock their full potential for genome assembly. Oxford Nanopore long reads can scaffold contigs generated by Illumina short reads to disambiguate regions of the assembly graph that cannot be resolved by Illumina short reads alone, as implemented in the Unicycler assembler (*Chen, Erickson & Meng, 2020*).

The human genome comprises approximately 3 Gb of nucleotides while a typical raw data set from Oxford Nanopore sequencing (before base-calling) exceeds 1 TB. This makes it difficult to upload the data with even high-speed bandwidth (a 100 Mbps transfer would take over 24 h to transfer the data alone). In addition, such large amounts of data would require substantial funding for computing resources as the required calculations would be considered Big Data. Because of these technological issues, we narrowed the analyses to genomes of prokaryotic sizes (up to 15 Mb in length, up to 15 GB in file size) but our server also can handle small eukaryotic genomes (such as *S. cerevisiae*). Unicycler may not be the best tool for yeast genomes but Flye and nanopore assembly pipeline from NanoForms, might be easily used for small eukaryotic genomes (see: *Martín-Hernández et al., 2021*). After deploying this server and gathering remarks from the users, we are considering designing another milestone which might address this problem. We also plan to launch the commercial version of the server for institutional clients.

## RESULTS

We introduce NanoForms: an intuitive and integrated web server for the processing and analysis of raw data from small genomes, yielding from Oxford Nanopore technology. The



**Figure 1** NanoForms server workflow. Subsequent computation steps are run but in between, the user is given the partial results can take action, *i.e.*, give more specific parameters for the next programmes, based on the data available (*i.e.*, quality, quantity *etc.*). In the end, the report is generated and sent in a form of a PDF file.

Full-size DOI: [10.7717/peerj.13056/fig-1](https://doi.org/10.7717/peerj.13056/fig-1)

user uploads an archived, single sequence file (FASTQ) or a list of archived, sequence files, then the data is preprocessed. The user then chooses several options on the go and, after subsequent steps, the user obtains the DNA/RNA sequence in a form of FASTA file as well as the HTML summary with reports, images or statistics of the calculations performed.

The initial output from the NanoPlot program (read length *vs.* the read quality) gives the user a quick outlook (as shown in [Fig. 2](#)) that helps him or her decide whether to continue the analysis or to go back to the lab to fix the sample. The final output of the NanoForms service is an assembled genome in fasta format, prokka annotation files and Bandage diagram, allowing for easy graphical assessment of assembly completeness. An example, derived for the *Bacillus subtilis*, is depicted in [Fig. 3](#). Since the MinION is marketed as needing only minor laboratory skills to operate, NanoForms needs practically no bioinformatics skills to produce the sequence (however more skilled users can benefit from extra, but optional, commands and options that they can provide during the course of the analysis). Therefore, we claim that our NanoForms server, in combination with Oxford Nanopore technology, has ultimately made NGS available for all, including both biologists and bioinformaticians, as specialized skills are no longer needed to perform certain NGS tasks and analyses. On the server website, after logging in, there are several toy datasets already provided to the user. Users can use these datasets for quality tests or data assembly using the respective forms provided in NanoForms, such as *Bacillus subtilis* SRX6978160.

## DISCUSSION

We performed a detailed analysis and comparison of similar services available for researchers. The first service to be tested was the CGE ([Larsen et al., 2017](#)) server. We checked and tested this service at the very beginning of this project and at that time, it was an up-to-date and convenient service, and easy to use for biologists without technical knowledge. Shortly before drafting this manuscript, the part of the server dedicated to genome assembly went offline, so the only input is the contigs in FASTA ([Pearson, 1990](#)) format. The rest of the tools tested are free to use, but as standalone programs, they are not interconnected with the comprehensive pipeline. In addition, no figures are generated, which makes the qualitative analyses more difficult.

The Enterobase server ([Zhou et al., 2020](#)) is aimed at wgMLST analyses. This server is mainly designed for genotyping isolates. Users can screen the database against specific STRs etc., and the service can also generate phylogenetic trees. Enterobase is dedicated to analyses of gut bacteria and supports Illumina or PacBio reads. The user only needs to provide the FASTQ ([Cock et al., 2010](#)) files, which need to be compressed by the gzip tool and also, manually curated, before running the service. The figures can be generated or plotted, however this requires additional manual user intervention. The Enterobase server does not accept nanopore data.

Another interesting tool, though with significantly diminished accessibility, is the Galaxy Tools service ([Cock et al., 2013](#)). Our experience testing this service suggests that the stand-alone version of the server needs to be installed and run locally for optimal use, but for smaller analysis it can be also run on the public server. The server provides

## Read lengths vs Average read quality plot

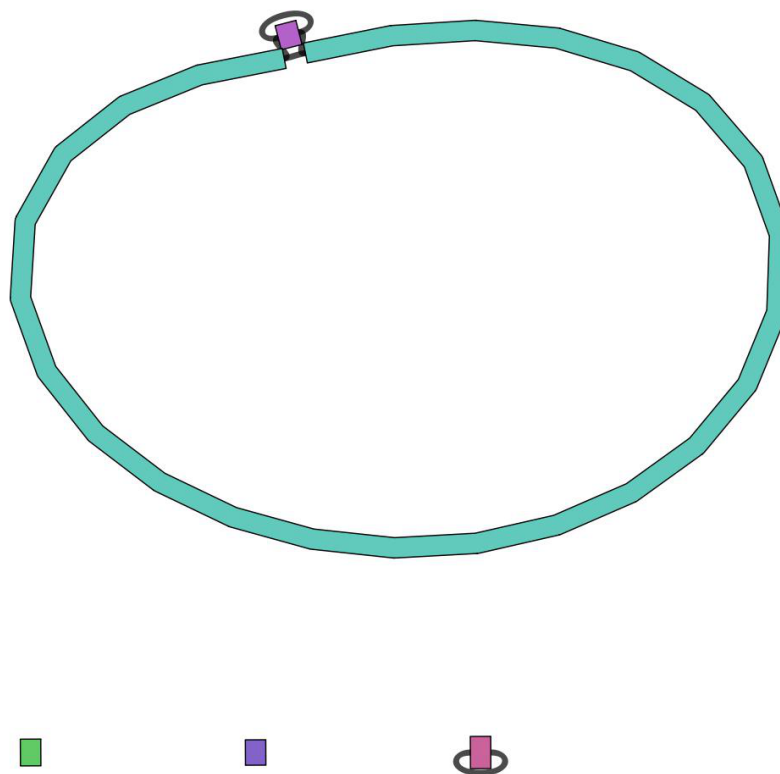


**Figure 2** An example histogram for read length vs. the read quality, as the output from the NanoPlot tool. Based on this information on the server website, the user can decide if the quality of the data is good enough to continue the analysis and thus, save time of the project or decide to which extent to crop the data to exclude low-quality short reads.

Full-size  DOI: [10.7717/peerj.13056/fig-2](https://doi.org/10.7717/peerj.13056/fig-2)

workflows (though only an empty set, for the novel user) which the user can adjust upon request. The tool supports both Illumina and long read (nanopore, PacBio) data input. It is worth noting that both the CGE and Galaxy Tools offer additional applications for practical genome analysis such as: resistance, serotype or virulence, and maintains the updated databases of these genetic targets.

While preparing and programming our tool, a software report was published about the new toolkit, NanoGalaxy, dedicated to the nanopore data processing (*De Koning et al., 2020*). NanoGalaxy is an extension of the aforementioned Galaxy Tools, in the area of support for nanopore data. In some ways NanoForms and NanoGalaxy seems to have similar features and functionalities. Similar to standard Galaxy Tools, NanoGalaxy seems to be more powerful but on the other hand it is aimed at more advanced user with more bioinformatics skills. In our subjective opinion NanoForms is easier to use for non-bioinformatics users but on the other hand can be treated as a quite bordered “black box”. NanoGalaxy is more complex and has more functionalities, but getting familiar with the numerous available algorithms requires some bioinformatics experience. NanoGalaxy and NanoSPC (*Xu et al., 2020*) deliver similar results and has similar capabilities as our server.



**Figure 3** An example of a circular sequence, being assembled at the end, as visualized by the Bandage tool. Sometimes the long-read nanopore data gives a few separate genome fragments as sequencing output (contigs), so the hybrid assembly option, provided by NanoForms, can often resolve these ambiguities. This image is normally the last stage of NanoForms sequencing protocol, however all figures and statistics that arose on the course of sequencing, are delivered to the user as a report.

Full-size  DOI: [10.7717/peerj.13056/fig-3](https://doi.org/10.7717/peerj.13056/fig-3)

NanoSPC is focused only on Nanopore data and as a result it is not possible to perform e.g., hybrid assembly there. It is also focused mostly on metagenomics, identification of pathogens and variant calling, but it seems to be easy to use also for users with limited dry-lab knowledge.

Another tool we tested was Patric ([Gillespie et al., 2011](#)). This platform provides bioinformatics analyses of all bacteria, with special focus on pathogens. It supports hybrid genome assembly in the formula of short + long reads (PacBio and nanopore are supported). [Table 1](#) provides the comprehensive summary of the features about the quoted services, compared based on NanoForms functionality, in terms of nanopore data processing. We decided not to include “demultiplexing reads support” in our table as the most current version of MinKnow software (the native software to Oxford Nanopore) supports this feature already.

As we were pursuing this project in mid-2020, Oxford Nanopore announced the availability of their EPI2ME (<http://epi2me.nanoporetech.com/requireslogin>) cloud-based workflow to process raw nanopore data. The platform’s intended use is only nanopore data analysis, without options for sequence assembly or trimming. After performing the



**Table 1** NanoForms server vs. other services: the comparison. Our service successfully fills in the gap for NGS genome assembly, in regard to the fully automated (but interactive) pipelined nanopore data processing.

Feature	CGE	Enterobase	Galaxy tools	EPI2ME	NanoPipe	Patric	Nano galaxy	NanoSPC	NanoForms
Raw data processing	+	+	+	+	+	+	+	+	+
Interactive interface	+	+	+	+	+	+	+	+	+
QA, reports	+	+	+	+	+	+	+	+	+
Free for academics	+	+	+	+	+	+	+	+	+
Sequence assembly	-	+	+	-	+ <sup>a</sup>	+	+	+	+
Qualitative analyses: images	-	+	+	+	+	+	+	+	+
Nanopore data processing	+	-	+	+	+	+	+	+	+
Tools connected into the pipeline	-	+	+/- <sup>b</sup>	-	-	+/- <sup>b</sup>	+	+	+
Hybrid assembly: nanopore + Illumina	-	-	+	-	-	+	+	-	+
Special nanopore visualization tools	-	-	+	+	-	-	+	+	+
Summary report generated	-	-	-	-	+	-	-	+	+
Ease of use <sup>c</sup>	+	+/-	+/-	+	+	+	+	+	+
Dry-lab knowledge unnecessary <sup>c</sup>	+	+	-	-	+	-	-	+	+

#### Notes.

<sup>a</sup>But not the novo assembly.

<sup>b</sup>There are no automatic pipelines, but users can develop them themselves.

<sup>c</sup>Subjective assessments of the authors.

base calling, reads are uploaded to the server *via* the EPI2ME Agent, which is a stand-alone program. The user can pick one of the several workflows (ie. microbiological classification or human genome analysis) which are triggered and executed in real time. Also in the table, the reader might find the characteristics of the last tool we reviewed: NanoPipe ([Shabardina et al., 2019](#)).

## CONCLUSIONS

In summary, our NanoForms server, freely available for all academics, bridges the high-speed of prokaryotic genome assembly with an intuitive, interactive interface. According to the Oxford Nanopore MinION specification product page, it is sufficient for the user to have a mid-range laptop and the device to obtain the sequence of the sample. No further resources are needed and the user can continue the genomic analyses after a short break in sequencing with the use of NanoForms service.

## ACKNOWLEDGEMENTS

MP wants to thank Przemyslaw Wroblewski and Giovanni Mazzocco who (with Dariusz Plewczynski) pioneered the NanOline web server at CeNT University of Warsaw, implementing the initial idea for nanopore genome assemblies of human genomes preceding the release of the NanoForms server. MP would also like to thank to Kacper KroczeK for additional testing and for checking documentation consistency on the server



homepage. In addition, MP would like to thank Michal Madera and the SoftSystem Sp. z o.o. (Ltd.) company for providing some parts of the code necessary to the server and for delivering auxiliary server infrastructure.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The work of Dominik Strzalka, Michal Pietal, Michal Cmil, Marta Sochacka-Pietal, Michal Wronski and Anna Czmil as well as the paper itself, was financed by Subcarpathian Center for Innovation (Podkarpackie Centrum Innowacyjności - PCI), Teofila Lenartowicza 4 Street, 35-051 Rzeszów, Poland, under the grant no. F3\_116 contract no. 05/PRZ/1/DG/PCI/2019. “Oxford Nanopore technology: optimization of enzymes and analysis of genomic data for commercial applications”. Michal Pietal and Dariusz Plewczynski were supported by the National Science Center grant (2018/02/X/NZ2/00622) “Identification of structural variants in the human genome using long fragments from next generation sequencing, based on Oxford Nanopore technology”. Dariusz Plewczynski was supported by Polish National Science Centre (2019/35/O/ST6/02484), and the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund (TEAM to Dariusz Plewczynski) “Three-dimensional Human Genome structure at the population scale: computational algorithm and experimental validation for lymphoblastoid cell lines of selected families from 1000 Genomes Project”. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Subcarpathian Center of Innovation (PCI): 05/PRZ/1/DG/PCI/2019.

Oxford Nanopore technology: optimization of enzymes and analysis of genomic data for commercial applications.

National Science Center: 2018/02/X/NZ2/00622.

Identification of structural variants in the human genome using long fragments from next generation sequencing, based on Oxford Nanopore technology.

Polish National Science Centre: 2019/35/O/ST6/02484.

Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund (TEAM to Dariusz Plewczynski).

Three-dimensional Human Genome structure at the population scale: computational algorithm and experimental validation for lymphoblastoid cell lines of selected families from 1000 Genomes Project.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Anna Czmil, Sylwester Czmil and Michal Cmil analyzed the data, prepared figures and/or tables, application programming, and approved the final draft.

- Michal Wronski analyzed the data, authored or reviewed drafts of the paper, application programming, and approved the final draft.
- Marta Sochacka-Pietal conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Tomasz Wołkiewicz analyzed the data, prepared figures and/or tables, and approved the final draft.
- Dariusz Plewczynski and Dominik Strzalka conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Michal Pietal conceived and designed the experiments, authored or reviewed drafts of the paper, application programming, and approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

The source code of the NanoForms server is available at Github: <https://github.com/czmlanna/nanoforms>. This is the source for standalone server installation. The server is available at <https://nanoforms.tech/>.

The *Bacillus subtilis* sequences are available at ENA: [SRX6978160](https://ena.ebi.ac.uk/ena/record/SRX6978160).

### REFERENCES

- Chen Z, Erickson DL, Meng J. 2020.** Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. *BMC Genomics* **21**(1):1–21 DOI [10.1186/s12864-019-6419-1](https://doi.org/10.1186/s12864-019-6419-1).
- Chen S, Zhou Y, Chen Y, Gu J. 2018.** fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**(17):i884–i890 DOI [10.1093/bioinformatics/bty560](https://doi.org/10.1093/bioinformatics/bty560).
- Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. 2010.** The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research* **38**(6):1767–1771 DOI [10.1093/nar/gkp1137](https://doi.org/10.1093/nar/gkp1137).
- Cock PJ, Grüning BA, Paszkiewicz K, Pritchard L. 2013.** Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. *PeerJ* **1**:e167 DOI [10.7717/peerj.167](https://doi.org/10.7717/peerj.167).
- Da Veiga Leprevost F, Grüning BA, Alves Aflitos S, Röst HL, Uszkoreit J, Barsnes H, Vaudel M, Moreno P, Gatto L, Weber J, Bai M. 2017.** BioContainers: an open-source and community-driven framework for software standardization. *Bioinformatics* **33**(16):2580–2582 DOI [10.1093/bioinformatics/btx192](https://doi.org/10.1093/bioinformatics/btx192).
- Davis MP, Van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. 2013.** Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* **63**(1):41–49 DOI [10.1016/j.ymeth.2013.06.027](https://doi.org/10.1016/j.ymeth.2013.06.027).
- De Coster W, D’Hert S, Schultz DT, Cruets M, Van Broeckhoven C. 2018.** NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**(15):2666–2669 DOI [10.1093/bioinformatics/bty149](https://doi.org/10.1093/bioinformatics/bty149).

- De Coster W, De Rijk P, De Roeck A, De Pooter T, D’Hert S, Strazisar M, Slegers K, Van Broeckhoven C. 2019. Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Research* 29(7):1178–1187 DOI 10.1101/gr.244939.118.
- De Koning W, Miladi M, Hiltmann S, Heikema A, Hays J, Flemming S, Van den Beek M, Mustafa D, Backofen R, Grüning B, Stubbs A. 2020. NanoGalaxy: nanopore long-read sequencing data analysis in Galaxy. *GigaScience* 10(9):giaa105.
- Escobar-Zepeda A, Godoy-Lozano EE, Raggi L, Segovia L, Merino E, Gutiérrez-Rios RM, Juárez K, Licea-Navarro AF, Pardo-Lopez L, Sanchez-Flores A. 2018. Analysis of sequencing strategies and tools for taxonomic annotation: defining standards for progressive metagenomics. *Scientific Reports* 8(1):1–13.
- Garlapati D, Charankumar B, Ramu K, Madeswaran P, Murthy MR. 2019. A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. *Reviews in Environmental Science and Bio/Technology* 18(3):389–411 DOI 10.1007/s11157-019-09501-4.
- Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, Dalay O, Driscoll T, Hix D, Mane SP, Mao C, Nordberg EK. 2011. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infection and Immunity* 79(11):4286–4298 DOI 10.1128/IAI.00207-11.
- Gloor GB, Hummelen R, Macklaim JM, Dickson RJ, Fernandes AD, MacPhee R, Reid G. 2010. Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. *PLOS ONE* 5(10):e15406 DOI 10.1371/journal.pone.0015406.
- Goldstein S, Beka L, Graf J, Klassen JL. 2019. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. *BMC Genomics* 20(1):1–17 DOI 10.1186/s12864-018-5379-1.
- Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* 17(1):239 DOI 10.1186/s13059-016-1103-0.
- Kawalek A, Kotecka K, Modrzejewska M, Gawor J, Jagura-Burdzy G, Bartosik AA. 2020. Genome sequence of *Pseudomonas aeruginosa* PAO1161, a PAO1 derivative with the ICE Pae 1161 integrative and conjugative element. *BMC Genomics* 21(1):1–12 DOI 10.1186/s12864-019-6419-1.
- Larsen MV, Joensen KG, Zankari E, Ahrenfeldt J, Lukjancenko O, Kaas RS, Roer L, Leekitcharoenphon P, Saputra D, Cosentino S, Thomsen MCF. 2017. The CGE tool box. In: *Applied genomics of foodborne pathogens*. Cham: Springer, 65–90.
- Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner P. 2016. Assembly of long error-prone reads Using de Bruijn graphs. *Proceedings of the National Academy of Sciences of the United States of America* 113(52):E8396–E8405.
- Martín-Hernández GC, Müller B, Chmielarz M, Brandt C, Hölzer M, Viehweger A, Passoth V. 2021. Chromosome-level genome assembly and transcriptome-based annotation of the oleaginous yeast *Rhodotorula toruloides* CBS 14. *bioRxiv*.

- McIntyre AB, Ounit R, Afshinnekoo E, Prill RJ, Hénaff E, Alexander N, Minot SS, Danko D, Foox J, Ahsanuddin S, Tighe S. 2017. Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. *Genome Biology* **18**(1):1–19 DOI [10.1186/s13059-016-1139-1](https://doi.org/10.1186/s13059-016-1139-1).
- Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**(13):i142–i150 DOI [10.1093/bioinformatics/bty266](https://doi.org/10.1093/bioinformatics/bty266).
- Moss EL, Maghini DG, Bhatt AS. 2020. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nature Biotechnology* **38**(6):701–707 DOI [10.1038/s41587-020-0422-6](https://doi.org/10.1038/s41587-020-0422-6).
- Ondov BD, Bergman NH, Phillippy AM. 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12**(1):1–10 DOI [10.1186/1471-2105-12-1](https://doi.org/10.1186/1471-2105-12-1).
- Pearson WR. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* **183**:63–98 DOI [10.1016/0076-6879\(90\)83007-v](https://doi.org/10.1016/0076-6879(90)83007-v).
- Rhoads A, Au KF. 2015. PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics* **13**(5):278–289 DOI [10.1016/j.gpb.2015.08.002](https://doi.org/10.1016/j.gpb.2015.08.002).
- Risse J, Thomson M, Patrick S, Blakely G, Koutsovoulos G, Blaxter M, Watson M. 2015. A single chromosome assembly of *Bacteroides fragilis* strain BE1 from Illumina and MinION nanopore sequencing data. *GigaScience* **4**(1):s13742–015.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**(14):2068–2069 DOI [10.1093/bioinformatics/btu153](https://doi.org/10.1093/bioinformatics/btu153).
- Shabardina V, Kischka T, Manske F, Grundmann N, Frith MC, Suzuki Y, Makalowski W. 2019. NanoPipe—a web server for nanopore MinION sequencing data analysis. *GigaScience* **8**(2):giy169.
- Simon HY, Siddle KJ, Park DJ, Sabeti PC. 2019. Benchmarking metagenomics tools for taxonomic classification. *Cell* **178**(4):779–794 DOI [10.1016/j.cell.2019.07.010](https://doi.org/10.1016/j.cell.2019.07.010).
- Ulahannan N, Pendleton M, Deshpande A, Schwenk S, Behr JM, Dai X, Tyler C, Rughani P, Kudman S, Adney E, Tian H. 2019. Nanopore sequencing of DNA concatemers reveals higher-order features of chromatin structure. *bioRxiv* 833590.
- Wang M, Fu A, Hu B, Tong Y, Liu R, Liu Z, Gu J, Xiang B, Liu J, Jiang W, Shen G. 2020. Nanopore targeted sequencing for the accurate and comprehensive detection of SARS-CoV-2 and other respiratory viruses. *Small* **16**(32):2002169 DOI [10.1002/sml.202002169](https://doi.org/10.1002/sml.202002169).
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017a. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics* **3**(10):e000132.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017b. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology* **13**(6):e1005595 DOI [10.1371/journal.pcbi.1005595](https://doi.org/10.1371/journal.pcbi.1005595).
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualisation of de novo genome assemblies. *Bioinformatics* **31**(20):3350–3352 DOI [10.1093/bioinformatics/btv383](https://doi.org/10.1093/bioinformatics/btv383).

- Wood DE, Lu J, Langmead B. 2019.** Improved metagenomic analysis with Kraken 2. *Genome Biology* **20(1):257** DOI [10.1186/s13059-019-1891-0](https://doi.org/10.1186/s13059-019-1891-0).
- Xu Y, Yang-Turner F, Volk D, Crook D. 2020.** NanoSPC: a scalable, portable, cloud compatible viral nanopore metagenomic data processing pipeline. *Nucleic Acids Research* **48(W1):W366–W371** DOI [10.1093/nar/gkaa413](https://doi.org/10.1093/nar/gkaa413).
- Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M, Brown D, Chattaway M, Dallman T, Delahay R, Kornschober C, Pietzka A. 2020.** The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. *Genome Research* **30(1):138–152** DOI [10.1101/gr.251678.119](https://doi.org/10.1101/gr.251678.119).