**Abstract:**

**Background:** Long-read sequencing can be applied to generate very long contigs and even completely assembled genomes at relatively low cost and with minimal sample preparation. As a result, long-read sequencing platforms are becoming more popular. In this respect, the Oxford Nanopore Technologies-based long-read sequencing ‘nanopore’ platform is becoming a widely used tool with a broad range of applications and end-users. However, the need to explore and manipulate the specialised data generated by long-read sequencing platforms means that accompanying specialised bioinformatics platforms and tools are required in order to correctly process the long-read data generated. Importantly, such tools should additionally help democratise bioinformatics analysis by enabling easy access and ease-of-use solutions for researchers.

**Results:** The Galaxy platform provides a user-friendly interface to computational command-line-based tools, including their software dependencies and refined workflows. The interface enables researchers, who do not necessarily possess programming experience or extended computer skills, to perform powerful bioinformatics analysis, including the assembly and analysis of short- or long-read sequence data. The newly developed ‘NanoGalaxy’ is a Galaxy-based toolkit for analysing long-read sequencing data, which is suitable for diverse applications, including de novo genome assembly from genomic, metagenomic and plasmid sequence reads.

**Conclusions:** A range of best practice tools and workflows for long-read sequence genome assembly have been integrated into a NanoGalaxy platform in order to facilitate easy access and use of bioinformatics tools for researchers. Nanogalaxy is freely available at the European Galaxy server nanopore.usegalaxy.eu with supporting self-learning training material available at training.galaxyproject.org.
| Question                                                                 | Response |
|-------------------------------------------------------------------------|----------|
| Are you submitting this manuscript to a special series or article collection? | No       |
| **Experimental design and statistics**                                  |          |
| Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends. | Yes      |
| Have you included all the information requested in your manuscript?      |          |
| **Resources**                                                           |          |
| A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible. | Yes      |
| Have you included the information requested as detailed in our Minimum Standards Reporting Checklist? |          |
| **Availability of data and materials**                                  |          |
| All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript. | Yes      |
| Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist? |          |
Table 1. The plasmids found by the workflow are BLAST against the plasmids recovered by R. Li et al..

| Barcode | Plasmid | Size  | Structural | No. of resistance genes | No. of resistance genes - paper | Coverage | Identity | Comment                  |
|---------|---------|-------|------------|-------------------------|---------------------------------|----------|----------|---------------------------|
| RB01    | RB01-LZ135-CTX-128976 | 128976 | Circular   | 8                       | 8                               | 100      | 98.615   |                           |
|         | RB01-LZ135-NDM-90845   | 90845  | Circular   | 5                       | 9                               | 99       | 98.788   |                           |
| RB02    | RB02-JN05-IncF-TET-116277-N | 116277 | Circular   | 4                       | 6                               | 99       | 97.872   |                           |
|         | RB02-JN05-IncN-CTX-139496-N | 142307 | Circular   | 10                      | 9                               | 99       | 97.796   |                           |
|         | RB02-JN05-IncN-NDM6-55342 | 55342  | Circular   | 11                      | 3                               | 99       | 86.689   | Not the first hit         |
|         | RB02-JN05-IncX-NDM5-45823 | 45823  | Circular   | 1                        | 1                               | 99       | 98.449   |                           |
|         | RB02-JN05-IncY-CTX-98443 | 98443  | Circular   | 0                       | 0                               | 99       | 98.771   |                           |
| RB03    | RB03-WH96T-IncF-OXA-153088 | 153088 | Circular   | 5                       | 3                               | 100      | 98.355   |                           |
|         | RB03-WH96T-IncN-NDM1-56215 | 56215  | Circular   | 6                       | 2                               | 100      | 98.278   |                           |
|         | RB04-SZ984-rt-IncF-TET-114056 | 114065 | Circular   | 6                       | 7                               | 99       | 98.387   |                           |
|         | RB04-SZ984-rt-IncX3-NDM1-56K-NC | 55919  | Circular   | 6                       | 2                               | 26       | 97.452   | Not the first hit         |
|         | RB04-SZ984-rt-IncY-130821 | 130821 | Circular   | 0                       | 0                               | 99       | 98.322   |                           |
| RB05    | RB05-C267-IncA/C-CTX-166407 | 166407 | Circular   | 8                       | 10                              | 99       | 98.92    |                           |
| RB06    | RB06-C499-IncA/C-CTX-192739 | 192739 | Circular   | 11                      | 11                              | 100      | 98.675   |                           |
| RB07    | RB07-vb0506-IncA/C-CTX-137742 | 137742 | Circular   | 5                       | 6                               | 100      | 98.48    |                           |
| RB09    | RB09-IncN-KPC-68571 | 68571  | Circular   | 34                      | 7                               | 100      | 98.497   |                           |
| RB10    | RB0-29KPC-IncF-TET-136532 | 136532 | Circular   | 10                      | 12                              | 99       | 98.658   |                           |
|         | RB0-29KPC-IncY-KPC-98K-N | 95908  | Circular   | 2                       | 1                               | 99       | 97.769   |                           |
| RB11    | RB1-IncF-IncHI-KPC-238153 | 238153 | Circular   | 2                       | 2                               | 99       | 98.48    |                           |
| RB12    | RB2-74T-KPC-IncF-115K-N | 115689 | Circular   | 0                       | 0                               | 99       | 97.948   |                           |
|         | RB2-74T-KPC-IncN-IncX1-KPC-108K-N | 107969 | Circular   | 5                       | 5                               | 100      | 97.927   |                           |

| Total / Average | 146 | 100 | 95.86 | 97.76 |
TECHNICAL NOTE

NanoGalaxy: Nanopore long-read sequencing data analysis in Galaxy

Willem de Koning1,2,†, Milad Miladi3,†, Saskia Hiltemann1, Astrid Heikema4, John P. Hays4, Marius van den Beek5, Dana A. Mustafa2, Rolf Backofen3, Björn Grüning3,⁎ and Andrew Stubbs1

1Department of Pathology, Clinical Bioinformatics Unit, Erasmus University Medical Centre, Wytemaweg 80, 3015 CN, Rotterdam, The Netherlands and 2Department of Pathology, Tumor Immuno–Pathology Laboratory, Erasmus University Medical Centre, 's Gravendijkwal 230, 3015 CE, Rotterdam, The Netherlands and 3Department of Computer Science, Bioinformatics Group, University of Freiburg, 79110 Freiburg im Breisgau, Germany and 4Department of Medical Microbiology and Infectious Diseases, Erasmus University Medical Centre, 's Gravendijkwal 230, 3015 CE, Rotterdam, The Netherlands and 5Department of Stem Cells and Tissue Homeostasis, Institut Curie, PSL Research University, 75005 Paris, France

†Contributed equally.
⁎To whom correspondence should be addressed. w.dekoning.1@erasmusmc.nl; gruening@informatik.uni-freiburg.de

Abstract

Background: Long-read sequencing can be applied to generate very long contigs and even completely assembled genomes at relatively low cost and with minimal sample preparation. As a result, long-read sequencing platforms are becoming more popular. In this respect, the Oxford Nanopore Technologies-based long-read sequencing 'nanopore' platform is becoming a widely used tool with a broad range of applications and end-users. However, the need to explore and manipulate the specialised data generated by long-read sequencing platforms means that accompanying specialised bioinformatics platforms and tools are required in order to correctly process the long-read data generated. Importantly, such tools should additionally help democratise bioinformatics analysis by enabling easy access and ease-of-use solutions for researchers.

Results: The Galaxy platform provides a user-friendly interface to computational command-line-based tools, including their software dependencies and refined workflows. The interface enables researchers, who do not necessarily possess programming experience or extended computer skills, to perform powerful bioinformatics analysis, including the assembly and analysis of short- or long-read sequence data. The newly developed ‘NanoGalaxy’ is a Galaxy-based toolkit for analysing long-read sequencing data, which is suitable for diverse applications, including de novo genome assembly from genomic, metagenomic and plasmid sequence reads.

Conclusions: A range of best practice tools and workflows for long-read sequence genome assembly have been integrated into a NanoGalaxy platform in order to facilitate easy access and use of bioinformatics tools for researchers. Nanogalaxy is freely available at the European Galaxy server https://nanopore.usegalaxy.eu with supporting self-learning training material available at https://training.galaxyproject.org.

Key words: Long-read sequencing; Nanopore; Galaxy; Reproducibility; Workflows

Compiled on: April 17, 2020.
Draft manuscript prepared by the author.
Background

Short-read sequencing has become a routine technique within clinical diagnostics [1]. However, the short length of the reads obtained (150–300 bp) complicates the assembly of genomes, especially with respect to highly repetitive regions and the detection of structural variance [2, 3, 4]. Furthermore, even comprehensive algorithms cannot overcome the issues associated with genome mapping or assembly using short read sequences. Importantly, advances in sequencing technology now allows long-read sequencing to be performed. Such platforms generate sequence reads much longer than the classic short-read technologies, including long-reads from single DNA molecules and without the need of PCR amplification (>10 kilobase on average). Moreover, utilising these technologies, library preparation and sequencing may be performed outside of traditional research laboratories, with sequencing outputs generated in real-time [5]. Protocols that require no PCR amplification also permit the direct detection of base modifications [6]. Analyzing the large amount of data generated by these short- and long-read sequencing technologies is a complex, multi-step process that is computationally intensive and often require bioinformatics expertise [7]. Specifically, for each step in the analysis, a set of different tools or software may be needed. For example, de novo assembly is performed via combination of multiple alignment, assembly and polishing tools, each utilizing its own input parameters. Such tools are typically executed from a UNIX command-line and require extensive computational resources, adding to the complexity of the analysis process.

To reduce this complexity, the Galaxy platform has implemented a standardized and user-friendly interface that accommodates command-line tools and refined workflows complete with their dependencies. The platform hosts a wide range of more than 7200 tools/software, has over 7000 citations, and is widely used for bioinformatics analysis within the biological science community [8, 9]. Here we introduce the NanoGalaxy toolkit which comprises a series of integrated Galaxy-based tools that enable researchers to generate powerful short- or long-sequence read assemblies for genomic and plasmid bioinformatics analyses. The NanoGalaxy toolkit is a user-friendly environment that can be utilized inside or outside of traditional research laboratories.

Findings

Tools

We have integrated a large collection of long-read sequence tools into the Galaxy platform, the NanoGalaxy toolkit, including diverse applications for the analyses of long-read sequences (Table 1). This toolkit is freely available from the Galaxy ToolShed, and has additionally been made available as a specialized GalaxyEU subdomain (https://nanopore.usegalaxy.eu).

Workflows

In order to increase the utility of this toolkit, we have developed a set of Galaxy workflows performing common analysis tasks using the tools in the NanoGalaxy toolkit.

Metagenomics taxonomic classification

The base quality of nanopore sequencing reads is constantly improving, making the actual assembly of reads more reliable. Further, the long-reads generated by nanopore sequencing can be used to provide valuable information from metagenomics data, including taxonomic classifications.

Kraken2 is a k-mer based classification technique that can efficiently assign the taxa of long reads that are resilient to the noisy nature of long-read data. The input reads for Kraken2 are compared to a database containing different classes and domains of life that are pre-indexed for algorithm efficiency. Within the NanoGalaxy toolkit we provide a workflow for taxonomic classification using Kraken2, including the post-processing of data and visualization of the results as interactive pie charts using the Krona tool [29].

Nanopolish tutorials

Nanopolish includes an extensive set of software tools for analysing nanopore long-read information at the raw signal level. Further, accompanying Nanopolish documentation provides intuitive tutorials on common scenarios, such as variation analysis and base methylation calling from the raw and mapped signals – Loman et al. [19]. We have integrated Nanopolish and its tutorials into Nanogalaxy in the form of workflows that can be used by researchers to analyse and interpret common quality values for their data.

De novo assembly of genome with high repeats

Compared to short reads, long-read data has the advantage of facilitating the assembly of large genomes that contain high numbers of repetitive elements. Schmid et al. utilised Flye and several other tools to generate a comprehensive assembly of the Pseudomonas koreensis genome, identifying that the genome has near identical repeat pairs up to 70 kilobase pairs in length [30]. These workflows have also been integrated in the NanoGalaxy toolkit.

Worked example: Antimicrobial resistance

As a further illustration of the utility of the NanoGalaxy toolkit and workflows, we describe below a full end-to-end workflow within Galaxy. This analysis pipeline performs a micro-

Table 1. NanoGalaxy toolkit.

| Category                      | Tool name          |
|-------------------------------|--------------------|
| De novo genome assembly       | Flye [10]          |
|                               | Canu [11]          |
|                               | Unicycler [12]     |
|                               | Wtdbg2 [13]        |
|                               | Miniasm [14]       |
|                               | Racon [15]         |
|                               | Spades [16]        |
| Long-read mapping             | Minimap2 [17]      |
|                               | GraphMap (2 tools) [18] |
| Polishing, QC and preprocessing| Nanopolish (3 tools) [19] |
|                               | Porechop [20]      |
|                               | Filllong [21]      |
|                               | Poretools (13 tools) [22] |
|                               | Pilon [23]         |
| Visualization                 | Nanoplot [24]      |
|                               | Bandage (2 tools) [25] |
| Taxonomy and metagenomics     | Kraken2 [26]       |
|                               | FlasFlow [27]      |
|                               | Stararrr [28]      |
| Methylation                   | Nanopolish (1 tool) [19] |
bial resistance detection in clinical samples. We describe this workflow in more detail in our training manual on the Galaxy Training materials repository (https://training.galaxyproject.org).

Background
According to the World Health Organization (WHO) and the Organisation for Economic Co-operation and Development (OECD), antimicrobial resistance (AMR) has become one of the biggest threats to global health, food security and economic development [31, 32]. Approximately 50,000 lives per year are lost due to AMR infections within the USA and Europe [33] and AMR infections are expected to increase, reaching 10 million deaths per year by 2050 [34].

Further, the misuse of antibiotics in medical, veterinary and agricultural sectors continues to contribute to the alarming global rise in antibiotic resistant infections – an increase that may ultimately lead to an era where common infections could once again be lethal. However, the (rapid) detection of antimicrobial resistant pathogens and their resistances in diseases, food and the environment are pillars by which increasing AMR could be detected, monitored and prevented.

Conventional methods for the identification of antimicrobial resistances involves microbial isolation (via culture) and phenotypic typing, which together can take a few days or weeks to complete [35]. Moreover, not all microbial species are amenable to laboratory-based culturing [36]. DNA-sequencing technologies may be utilised to sequence the genomes of cultured microorganisms for the presence of antimicrobial resistance genes, which reduces the time-to-result time. Currently, Illumina sequencing is most widely used, but using this sequencing technology generates difficulties in correctly identifying repetitive insertion sequences, sequences that may flank horizontally acquired genes associated with AMR [37]. Newer technologies such as ‘nanopore sequencing’ allow the generation of so-called long-read sequences, although the error rate obtained in the sequence data may be relatively high compared to short-read technologies. One solution is to provide long-read tools and to combine the strengths of both short- and long-read sequencing methods to overcome the problems of repetitive sequences and sequence read error rates.

Use case 1: Long-read sequencing analysis
The Nanogalaxy toolkit incorporates a rapid long-read assembly workflow employing minimap2 [17], miniasm [14] and Racon [38]. Tools for further analysis in the toolkit include Staramr [28] for resistance gene detection, PlasFlow [27] and Bandage [25] for microbial species/plasmid determination and NanoPlot [24] for quality assessment.

In this worked example, the outcome of the Nanogalaxy pipeline was compared to the plasmid sequences recovered by Li et al. [39] (Table S1). The pipeline recovered 19 out of 21 plasmids, with an average identity of 97.76%. The number of detected resistance genes was higher than that found by Li et al. [39], which was expected as Staramr includes the PointFinder (chromosomal point mutations) database [40] and current long-read sequencing may generate relatively high sequence error rates.

Use case 2: Combining short- and long-read sequencing
The previously described long-read assembly workflow rapidly assembles genomes but cannot reliably detect single nucleotide polymorphisms (SNPs). This is because long-read sequencing has a higher error rate than short-read sequencing. Therefore, the NanoGalaxy toolkit includes a workflow that processes both long- and short-read sequences. In this respect, Unicycler was integrated into the NanoGalaxy toolkit in order to combine the best features of long- and short-sequencing technologies.

The workflow recommended by the Unicycler developers [12] includes: Trim Galore [41], Porechop [20] and Filtlong [21] for quality trimming; Unicycler [12] for de novo assembly and Bandage [25] for plasmid visualization. These tools are available as stand-alone tools and combined in a NanoGalaxy workflow.

The assembly graphs shown in Fig. 1, compare the NanoGalaxy toolkit with the results from Wick et al. [12]. The Illumina-only (short-read sequencing) graphs show no clear structure(s) present, whereas Nanopore-only (long-read sequencing) is able to generate the circularized structure expected of plasmids. The combination of both sequence techniques gives the clearest view of the circular assemblage expected of plasmids, analogues to the results obtained by Wick et al. [12] (Figure 1). Note that different combinations of short- and long-read tools can be used individually, or combined, to generate personalized workflows.

Methods
Implementation
The tools and workflows included in the NanoGalaxy toolkit enable non-bioinformatics-trained researchers to perform extensive genomics analysis using long-read sequence data, without the need for any coding skills. All tools and their dependencies are installed on the Galaxy platform and are managed by the Conda framework for dependency management. NanoGalaxy tools and their dependencies are available from the Bioconda Conda channel [42]. The Galaxy wrappers are developed openly on GitHub, utilizing the Travis continuous integration framework [43] for testing, and have been made available on the Galaxy ToolShed [9].

Training Materials
An online training manual for the AMR use case described in this publication, as well as a description of NanoGalaxy tools and end-to-end workflows can be found on the Galaxy training materials website [44].

Future Work
The availability of long-read sequencing platforms and data analysis tools is relatively new, with improvements in technology and software continually being developed. As more tools become available these will need to be assembled into existing or new toolkits. Additionally, the future availability of toolkits such as NanoGalaxy will help popularise long-read sequencing, while making it accessible to non-bioinformatics-trained researchers of the future.
Availability of source code and requirements

- Project name: NanoGalaxy
- Project home page: https://nanopore.usegalaxy.eu
- Training Manual: https://training.galaxyproject.org/training-material/topics/metagenomics/tutorials/plasmid-metagenomics-nanopore/tutorial.html
- License: GNU GPL

All developed Galaxy wrappers are available for installation from the Galaxy ToolShed (https://toolshed.g2.bx.psu.edu/). The corresponding code repositories for the tool wrappers are listed in Table 2. The workflows described in this work are publicly available from the European Galaxy server, as well as published Galaxy histories with an example run of each of these workflows (3).

Table 2. Tool availability.

| Tool          | Github repository                                      |
|---------------|--------------------------------------------------------|
| Bandage       | https://github.com/galaxyproject/tools-iuc/tree/master/tools/bandage |
| Canu          | https://github.com/bgruening/galaxytools/tree/master/tools/canu       |
| Filtlong      | https://github.com/galaxyproject/tools-iuc/tree/master/tools/filtlong  |
| Flye          | https://github.com/galaxyproject/tools-iuc/tree/master/tools/flye        |
| GraphMap      | https://github.com/galaxyproject/tools-iuc/tree/master/tools/graphmap   |
| Kraken2       | https://github.com/galaxyproject/tools-iuc/tree/master/tools/kraken2     |
| Miniasm       | https://github.com/galaxyproject/tools-iuc/tree/master/tools/miniasm    |
| Minimap2      | https://github.com/galaxyproject/tools-iuc/tree/master/tools/minimap2   |
| Nanopolish    | https://github.com/galaxyproject/tools-iuc/tree/master/tools/nanopolish |
| Pilon         | https://github.com/galaxyproject/tools-iuc/tree/master/tools/pilon       |
| PlasFlow      | https://github.com/galaxyproject/tools-iuc/tree/master/tools/plasflow    |
| Porechop      | https://github.com/galaxyproject/tools-iuc/tree/master/tools/porechop    |
| Poretools     | https://github.com/galaxyproject/tools-iuc/tree/master/tools/poretools   |
| Unicycler     | https://github.com/galaxyproject/tools-iuc/tree/master/tools/unicycler   |
| Racon         | https://github.com/bgruening/galaxytools/tree/master/tools/racon        |
| Spades        | https://github.com/galaxyproject/tools-iuc/tree/master/tools/spades      |
| Staramr       | https://github.com/phac-nml/galaxy_tools/tree/master/tools/staramr      |
| Wtdbg2        | https://github.com/bgruening/galaxytools/tree/master/tools/wtdbg2        |

Table 3. Workflow availability.

| Workflow                                                                 | Link                                      | History                              |
|--------------------------------------------------------------------------|-------------------------------------------|--------------------------------------|
| Basic workflows inspired by the Nanopolish tutorials                     | https://nanopore.usegalaxy.eu/u/milad/w  | eu/u/milad/h/nanopolish-tutorial     |
| Genome assembly: Flye-based WF for highly repetitive genomes             | https://nanopore.usegalaxy.eu/u/milad/w  | eu/u/milad/h/ont-assembly-flye-ahrens|
| Genome assembly: Unicycler-based WF for Klebsiella pneumoniae           | https://usegalaxy.eu/u/milad/h/vick-stal-nanopore | eu/u/milad/h/vick-stal-nanopore |
| Metagenomics classification                                              | https://nanopore.usegalaxy.eu/u/milad/w  | eu/u/milad/h/nanopolish-kram2        |

Availability of supporting data and materials

The data presented here to illustrate the functionality of the tools was obtained from previous publications [45, 39] and was collected and made available from Zenodo https://doi.org/10.5281/zenodo.3741446 [46].

Declarations

List of abbreviations

- AMR: Antimicrobial Resistance
- OECD: Organisation for Economic Co-operation and Development
- ONT: Oxford Nanopore Technologies
- SNPs: Single Nucleotide Polymorphisms
- WHO: World Health Organization

Competing Interests

The authors declare that they have no competing interests.

Funding

This project was made possible with the support of Support Casper and the Albert Ludwig University of Freiburg. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement 835775.

Author’s Contributions

WdK, MM and SH contributed to toolkit development and writing of the manuscript. AH tested and evaluated the tools and suggested modifications, feature requests and user improvements. MvdB contributed to the tool development. BG contributed to the tool development and supervised the project. DM, RB, and AS supervised the project. JH contributed to AMR tool
and nanopore sequencing discussions and the writing of the manuscript.

All authors approved the final version of the manuscript.

Acknowledgements

The authors would like to thank the Galaxy community for their help in reviewing, testing, and validating the tools presented here.

References

1. Gillissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. Genome biology 2011;12(9):228.
2. de Koning AJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS genetics 2011;7(12):e1002384.
3. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 2016;17(6):333.
4. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nature Reviews Genetics 2006;7(2):85.
5. Tsai YC, Greenberg D, Powell J, Hoijer I, Ameur A, Strahl M, et al. Amplification-free, CRISPR–Cas9 targeted enrichment and SMRT sequencing of repeat-expansion disease causative genomic regions. bioRxiv 2017; p. 203919.
6. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature methods 2010;7(6):461.
7. Hemlata, Tiwari A. Applications of bioinformatics tools to combat the antibiotic resistance. In: International Conference on Soft Computing Techniques and Implementations, ICSCTI 2015 IEEE; 2016. p. 96–98. http://ieeexplore.ieee.org/document/7489645/.
8. Zotero list of Citations of the Galaxy project; https://www.zotero.org/groups/1732893/galaxy.
9. Galaxy Tool Shed; https://toolshed.g2.bx.psu.edu/.
10. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nature biotechnology 2019;37(5):540.
11. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome research 2017;27(5):722–736.
12. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Reconstructing bacterial genome assemblies from short and long sequencing reads. PLoS Computational Biology 2017 Jun;13(6):e1005595.
13. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. BioRxiv 2019;p. 530972.
14. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics (Oxford, England) 2016;32(44):2103–10.
15. Vasar R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome research 2017;27(5):737–746.
16. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, et al. Assembling genomes and mini-metagenomes from highly chimeric reads. In: Annual International Conference on Research in Computational Molecular Biology Springer; 2013. p. 158–170.
17. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018 sep;34(18):3094–3100.
18. Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nature communications 2016;7:11307.
19. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nature methods 2015;12(8):733.
20. Wick R, Porechop. Github https://github.com/rrwick/Porechop; 2017.
21. Wick R, Filtlong. Github https://github.com/rrwick/Filtlong; 2017.
22. Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics (Oxford, England) 2015;31(20):3350–3352.
23. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one 2014;9(11):e112961.
24. De Coster W, D’Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics (Oxford, England) 2018 aug;34(15):2666–2669.
25. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: Interactive visualization of de novo genome assemblies. Bioinformatics 2015 oct;31(20):3350–3352.
26. Wood DE, Li J, Langmead B. Improved metagenomic analysis with Kraken 2. BioRxiv 2019;p. 762302.
27. Krawczyk PS, Lipinski S, Dziewiowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. Nucleic acids research 2018 apr;46(6):e35.
28. Staramr. Github https://github.com/phac-nml/staramr; 2018.
29. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC bioinformatics 2011;12(1):385.
30. Schmid M, Frei D, Patrignani A, Schlapbach R, Frey JE, Remus-Emsermann MN, et al. Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats. Nucleic acids research 2018;46(17):8953–8965.
31. Organisation for Economic Co-operation and Development, Antimicrobial Resistance; 2017.
32. World Health Organization, Antibiotic resistance; 2018.
33. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. Review on Antimicrobial Resistance 2014;.
34. O’Neill J, Tackling a crisis for the health and wealth of nations; 2014.
35. Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. Genome Biology 2015 may;16(1):114.
36. Mitsuhashi S, Kryukov K, Nakagawa S, Takeuchi J, Shiraiishi Y, Asano K, et al. A portable system for metagenomic analyses using nanopore-based sequencer and laptop computer; can realize rapid on-site determination of bacterial compositions. bioRxiv 2017;p. 101865.
37. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nature Biotechnology 2014 dec;33:296.
38. Vasar R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Research 2017;27(5):737–746.
39. Li R, Xie M, Dong N, Lin D, Yang X, Wong MHY, et al. Efficient generation of complete sequences of MDR–encoding plasmids by rapid assembly of MinION barcoding sequencing data. Gigascience 2018;7(3):1–9.
40. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aare-
strup FM. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. Journal of Antimicrobial Chemotherapy 2017;72(10):2764–2768.

41. Kreuger F, Trim Galore! Github https://github.com/FelixKrueger/TrimGalore; 2016.

42. Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, et al. Bioconda: sustainable and comprehensive software distribution for the life sciences. Nature methods 2018;15(7):475.

43. Travis CI: Test and Deploy with Confidence; https://travis-ci.org/.

44. Batut B, Hiltemann S, Bagnacani A, Baker D, Bhardwaj V, Blank C, et al. Community-Driven Data Analysis Training for Biology. Cell Systems 2018 jun;6(6):752–758.e1. https://doi.org/10.1016/j.cels.2018.05.012.

45. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microbial Genomics 2017;3(10):e000132.

46. NanoGalaxy Zenodo; https://doi.org/10.5281/zenodo.3529597.