Identifying Cyanobacteria through Next-Generation Sequencing Technology for Modern Agriculture

Joko Pebrianto Trinugroho[1*], *Faisal* Asadi[1], and *Bens* Pardamean[2]

[1]Bioinformatics and Data Science Research Center, 11480 Bina Nusantara University, Indonesia
[2]Computer Science Department BINUS Graduate Program, Master of Computer Science, 11480 Bina Nusantara University, Indonesia

Abstract. As the global demand for food continue to increase, it is important to find a way to meet the demand without creating any problems to the environment. Cyanobacteria have a prospective to be utilised for the modern agriculture, as they contribute to the improvement of the soil fertility, the crop yield, and they also do not harm the environment. Therefore, it is crucial to understand the species of cyanobacteria or the characteristics that could be used for modern agriculture. The development of Next-Generation Sequencing (NGS) technologies enables us to study the genome of cyanobacteria. Thus, we can study their characteristics by analysing the NGS data. This paper aims to elaborate a pipeline for genomic analysis on cyanobacteria from NGS data. We used a free Linux-based software tool, namely Breseq to process the NGS sequencing raw data. This tool predicts mutations that occur in the genome of the sample, including single-nucleotide variation, insertions, and deletions which could be beneficial for the identification of a new species or a mutant of cyanobacteria which has the right characteristics for modern agriculture utilisation.

1 Introduction

Cyanobacteria are organisms with a microscopic size and the ability to perform photosynthesis, which can be found in different habitats, including marine, lakes, ponds, rocks, and soils [1-3]. It is believed that cyanobacteria have been living in the Earth for more than 2 billion years, which is much earlier than the other organisms such as plants and animals [4]. Cyanobacteria are also known to play a critical role in the agriculture as they can improve the soil fertility and the crop yield [5]. This makes cyanobacteria have a huge potential to be explored for the modern sustainable agriculture. Although there are a lot of species of cyanobacteria, not all of them could be utilised for agriculture. Therefore, effort is needed in identifying and discovering the right species and characteristics of cyanobacteria for modern agricultural utilisation.

One of the promising approaches to identify cyanobacteria is by studying their genome, which can be achieved using next-generation sequencing (NGS) technology. Unlike the standard sequencing method, the NGS method can easily capture the information of the entire genome of an organism. The development of NGS technologies has been extremely rapid for the last 20 years, which enables us to obtain the genomic or transcriptomic sequencing data quickly in a much more affordable price [6-15]. Currently, there are several NGS technologies that are widely known, including Illumina, Roche 454, IonTorrent, Pacific Biosciences (PacBio), and Oxford Nanopore [16, 17].

Analysing NGS data is one of the most important steps in studying the genome of cyanobacteria. Most often, we compare the genome of a newly identified cyanobacterium or a mutant to the reference organism, a cyanobacterium which its genome has been studied before [18]. In order to compare the difference between the genome of a new cyanobacterium with the reference sequence, it is crucial to have a parameter. There are several parameters that is commonly used for genomic analysis: single-nucleotide variation (SNV) or single-nucleotide polymorphism (SNP), short insertion or deletion (indel), and structural variation (SV) which comprises of long deletions or insertions, and rearrangements [19-22]. To perform genomic analysis, many software tools have been created and developed. However, most of the software tools for genomic analysis are created to analyse a large genome, such as human genome, so some reads pointing to the repetitive sequences are overlooked to speed up the process [23, 24]. Therefore, finding the right software tool to perform genomic analysis on cyanobacteria is important so that we can study the genome of cyanobacteria well.

Here, we describe a pipeline that is designated for performing identifying cyanobacteria through their genome. The pipeline consists of wet lab and dry lab, which are discussed in this paper. The key steps of the pipeline, which leads to the results, are also demonstrated.

* Corresponding author: joko.trinugroho@binus.edu, faisal.asadi@binus.edu, bpardamean@binus.edu

2 Literature review

2.1 Overview of NGS genome analysis of cyanobacteria

Overall, the cyanobacterial genome analysis pipeline consists of two parts, the wet lab and dry lab parts. The wet lab focuses on performing Whole Genome Sequencing (WGS) using Next Generation Sequencing technique. The workflow of the wet lab part is shown in Figure 1. The most important step of the Whole Genome Sequencing is to obtain the correct cyanobacterial sample. It is crucial to make sure that we have a sufficient number of cyanobacterial cells for genomic DNA extraction. After the genomic DNA of the sample has been successfully extracted, it is subsequently subjected to DNA purity measurement and DNA concentration measurement. This process is performed to ensure that the quality and quantity of genomic DNA meet the standard [25, 26]. If both the concentration and purity of the genomic DNA meet the standard, the next step is library preparation. This step is performed to make the genomic DNA sample compatible with the NGS sequencing technology/system that is used [27]. After the library preparation was successfully performed, the sample was then subjected to NGS sequencing with Ilumina Nextseq550. Ilumina Nextseq550 is a powerful system with a high accuracy and robust performance, which has been vastly used to study cyanobacterial genome [28-30]. The system produces sequencing raw data that needs to be further analysed.

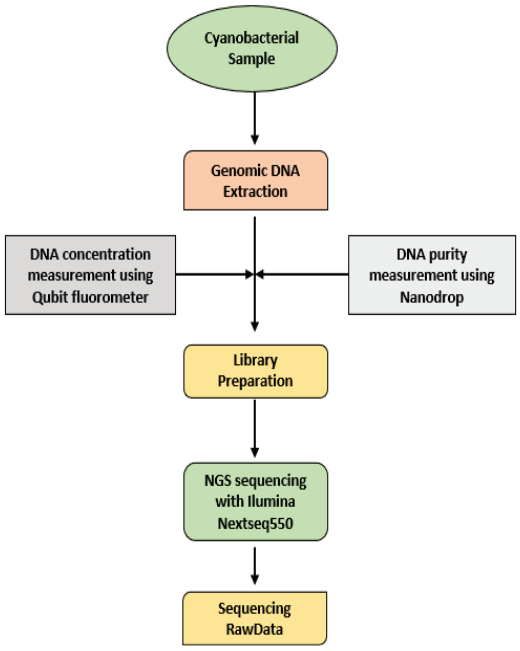

Fig. 1. The workflow of whole genome sequencing (wet lab).

3 Methodology

3.1 Data acquisition

All genome sequences were obtained from an online database. The reference sequence of cyanobacterium Synechocystis sp. PCC 6803 (BA000022.2) was retrieved from https://www.ncbi.nlm.nih.gov, while the sample of genomic data of Synechocystis sp. PCC 6803 GT-1 strain was retrieved from https://www.ebi.ac.uk. Both reference sequence and sample were then subjected to data analysis.

3.2 NGS sequencing data analysis

Analysis of the NGS sequencing data from cyanobacterial sample was performed using breseq (v.0.35.5) software tool [31, 32]. This tool is free-access, based on Linux, and has been extensively used to analyse genome's sequencing data from prokaryotes, including cyanobacteria ([18], [33-36]). To prepare the input for breseq, we first extract fastq.gz files containing raw sample genomic data into fastq files. These files were then used for the analysis. In addition, we used the genebank (.gbk) extension file as the reference sequence, since it has already been annotated. We run breseq in a default mode without any modifications. The detailed steps are explained in the results section.

4 Results and discussion

The dry lab part of the cyanobacterial genome analysis pipeline focuses on analysing NGS sequencing raw data using bioinformatic approach. The workflow of the dry lab part is shown in Figure 2. As mentioned in the methodology, we used breseq software tool for the genomic analysis. Both sequencing raw data (FASTQ) and reference sequences are needed for the input. The first step is mapping reads using bowtie2 [37], which is already integrated in breseq software. This step generates read alignments (SAM/BAM extension file). From the best-read alignments, the software then creates several evidences, such as New Junction Evidence, Missing Coverage Evidence, and Read Alignment Evidence, that are used to predict mutations [31]. The predicted mutations include large deletions or insertions, short deletions or insertions, and substitutions. The software than created a list of mutations and evidence in three different formats, genome diff, Variant Call Format (VCF) [38], and Genome Variant Format (GVF) [39], which can be visualised interactively using Integrative Genomics Viewer [40]. Finally, the output of the data analysis is created in Hypertext Markup Language (HTML) extension file, which contains annotated mutation and evidence presented in a table [31].

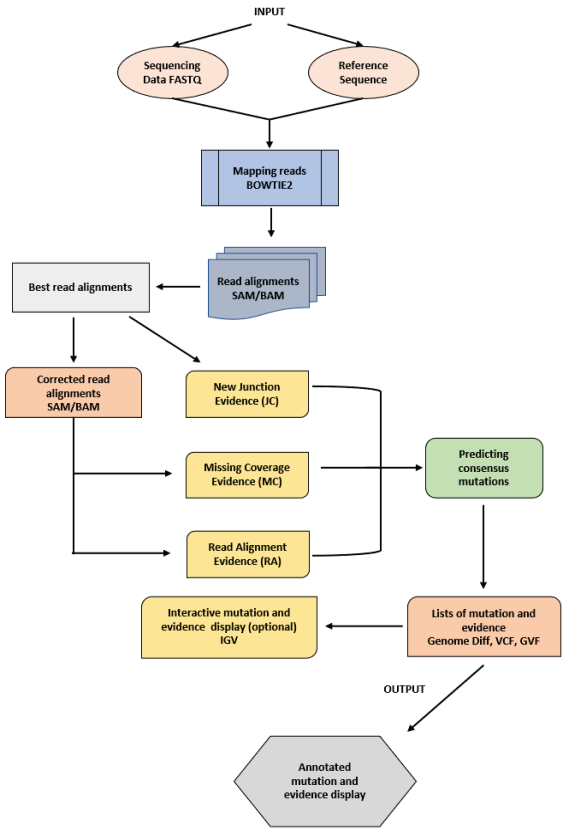

Fig. 2. The workflow of genomic analysis on NGS data from cyanobacterial sample (dry lab) [31].

The result of genomic analysis from cyanobacterial sample (*Synechocystis* sp. PCC 6803 GT-1 strain) is displayed in Table I. The table informed important information from the analysis, including evidence, position, type of mutation, annotation, gene, and description. The evidence tells the type of evidence that are used to predict the mutations as mentioned previously. The position reflects the location of the mutation within the genome. The mutation gives information about the type of mutation, while the annotation tells the change in coding sequence/amino acid of the genomic DNA. If the reference sequence has been well curated and annotated, the detailed information about the name of the gene and its description are also shown. This helps us to further analyse the data.

The genomic analysis predicted 27 mutations from the sample (Table 1). This includes substitutions, short insertions or deletions, and large deletions. Single-nucleotide variations (SNV) and short insertions or deletions (indels) were the most predicted mutations from the analysis. SNV and Indels commonly occurred in the genome, so they are used as genetic markers [41-43] . Among all predicted mutations, the mutation in the *psaA* and *sps* genes may be beneficial for the growth of cyanobacteria, as both genes are involved in the metabolism [44]. *psaA* gene encodes photosystem I protein subunit, which is crucial for photosynthesis, while *sps* gene functions in sucrose synthesis [44, 45]. Variation in these genes could lead to the improvement of cells' biomass, which could be further utilised for agriculture [45]. Overall, our genomic analysis has successfully predicted mutations from NGS sequencing data of the cyanobacterial genomic DNA.

Table 1. List of predicted mutations.

	Predicted mutations					
	Evidence	Position	Mutation	Annotation	Gene	Description
1	RA	387,006	C→T	P109L (CCT→CTT)	slr1085 →	ORF_ID:slr1085; unknown protein
2	RA	842,060	C→T	R185Q (CGG→CAG)	rpl3 ←	50S ribosomal protein L3
3	RA	909,360	C→T	E93K (GAG→AAG)	pmgA ←	PmgA
4	RA	943,495	G→A	V604I (GTC→ATC)	psaA →	P700 apoprotein subunit Ia
5	RA	1,012,958	G→T	intergenic (-70/+87)	repA ← / ← ftsZ	rare lipoprotein A/cell division FtsZ protein
6	MC JC	1,200,306	Δ1,183 bp		[slr1862]–sll1780	[slr1862], sll1780
7	RA	1,364,187	A→G	L116L (TTG→CTG)	pyrF ←	orotidine 5' monophosphate decarboxylase
8	RA	1,392,586	T→C	L204S (TTA→TCA)	pstB →	phosphate transport ATP-binding protein; PstB
9	RA	1,470,212	G→A	R46C (CGC→TGC)	fabZ ←	(3R)-hydroxymyristol acyl carrier protein dehydrase
10	RA	1,764,198	T→G	F158C (TTC→TGC)	slr1962 →	ORF_ID:slr1962; unknown protein
11	MC JC	2,048,412	Δ1,183 bp		slr1635–[slr1636]	slr1635, [slr1636]
12	RA	2,092,571	A→T	L313* (TTA→TAA)	sll0422 ←	asparaginase
13	RA	2,198,893	T→C	L689L (TTA→TTG)	sll0142 ←	cation or drug efflux system protein
14	RA	2,204,584	(G)9→8	coding (428/498 nt) coding (109/897 nt)	gspF → pilC →	general secretion pathway protein F pilin biogenesis protein
15	RA	2,301,721	A→G	K403E (AAG→GAG)	slr0168 →	ORF_ID:slr0168; unknown protein
16	RA	2,350,285	+A	intergenic (-29/-87)	psbI ← / → slr0363	photosystem II PsbI protein/ORF_ID:slr0363; unknown protein
17	RA	2,360,246	+C	coding (8924/9090 nt)	slr0364 →	ORF_ID:slr0364; unknown protein
18	RA	2,409,244	Δ1 bp	coding (301/351 nt)	sll0762 ←	ORF_ID:sll0762; unknown protein
19	RA	2,419,399	Δ1 bp	coding (496/510 nt)	ycf22 ←	ORF_ID:sll0751; hypothetical protein
20	RA	2,544,044	+C	coding (280/300 nt)	ssl0787 ←	ORF_ID:ssl0787; unknown protein
21	RA	2,602,717	C→A	H82Q (CAC→CAA)	slr0468 →	ORF_ID:slr0468; unknown protein
22	RA	2,602,734	T→A	I88N (ATT→AAT)	slr0468 →	ORF_ID:slr0468; unknown protein
23	RA	2,748,897	C→T	intergenic (+40/-49)	slr0210 → / → ssr0332	sensory transduction histidine kinase/ORF_ID:ssr0332; unknown protein
24	RA	2,817,683	G→T	intergenic (-230/-179)	ssl1045 ← / → cyp	ORF_ID:ssl1045; unknown protein/cytochrome P450
25	RA	3,142,651	A→G	L75L (CTT→CTC)	sps ←	sucrose phosphate synthase
26	RA	3,260,096	(C)7→6	intergenic (-209/+41)	sll0529 ← / ← sll0528	ORF_ID:sll0529; unknown protein/ORF_ID:sll0528; hypothetical protein
27	MC JC	3,400,332	Δ1,183 bp		[sll1475]–[sll1473]	[sll1475], sll1474, [sll1473]

5 Conclusion

In this paper, we presented a genomic analysis pipeline from NGS data of cyanobacteria, which was obtained

from an online database. The genomic analysis in this study used a free-access software tool with simple commands. Our results have predicted different mutations, including SNV and indels. Hence, this pipeline will be useful to identify and discover a new species or a mutant of cyanobacteria which has the right characteristics for modern agriculture utilisation.

References

1. F. Garcia-Pichel, J. Belnap, S. Neuer, F. Schanz, *Estimates of global cyanobacterial biomass and its distribution*, Algological Studies **109**, pp. 213–227 (2003)
2. A. D. Jungblut, C. Lovejoy, W. F. Vincent, *Global distribution of cyanobacterial ecotypes in the cold biosphere*, The ISME J. **4**, pp. 191–202 (2010)
3. P. Flombaum, J. L. Gallegos, R. A. Gordillo, J. Rincón, L. L. Zabala, N. Jiao, D. M. Karl, W. K. W. Li, M. W. Lomas, D. Veneziano, C. S. Vera, J. A. Vrugt, A. C. Martiny, *Present and future global distributions of the marine cyanobacteria prochlorococcus and synechococcus*, Proceedings of the National Academy of Sciences of the United States of America **110**, 24, pp. 9824–9829 (2013)
4. B. E. Schirrmeister, A. Antonelli, H. C. Bagheri, *The origin of multicellularity in cyanobacteria*, BMC Evolutionary Biology **11**, 1, pp. 45 (2011)
5. J. S. Singh, A. Kumar, A. N. Rai, D. P. Singh, *Cyanobacteria : a precious bio-resource in agriculture, ecosystem, and environmental sustainability*, Frontiers in Microbiology **7**, pp. 1–19 (2016)
6. E. R. Mardis, *Next-generation DNA sequencing methods*, Annual Review of Genomics and Human Genetics **9**, pp. 387–402 (2008)
7. J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. DeWinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, *Real-time DNA sequencing from single polymerase molecules*, Science **323**, 5910, pp. 133–138 (2009)
8. W. J. Ansorge, *Next-generation DNA sequencing techniques*, New Biotechnology **25**, 4, pp. 195–203 (2009)
9. J. W. Baurley, C. S. McMahan, C. M. Ervin, B. Pardamean, A. W. Bergen, *Biosignature discovery for substance use disorders using statistical learning*, Trends in Molecular Medicine **24**, 2, pp. 221–235 (2019)
10. C. Joyner, C. McMahan, J. Baurley, B. Pardamean, *A two-phase bayesian methodology for the analysis of binary phenotypes in genome-wide association studies*, Biometrical J. **62**, 1, pp. 191–201 (2020)
11. D. Sudigyo, G. Rahmawati, D. W. Setiasari, R. H. Poluan, T. W. Cenggoro, A. Budiarto, A. A. Hidayat, S. R. Indrasari, Afiahayati, S. M. Haryana, B. Pardamean, *Bioinformatics pathway analysis pipeline for NGS transcriptome profile data on nasopharyngeal carcinoma*, IOP Conf. Series: Earth and Environmental Science **794**, 1, pp. 1–10 (2021)
12. I. Yusuf, B. Pardamean, J. W. Baurley, A. Budiarto, U. A. Miskad, R. E. Lusikooy, A. Arsyad, A. Irwan, G. Mathew, I. Suriapranata, R. Kusuma, M. F. Kacamarga, T. W. Cenggoro, C. McMahan, C. Joyner, C. I. Pardamean, *Genetic risk factors for colorectal cancer in multiethnic Indonesians*, Scientific Reports **11**, 9988, pp. 1–9 (2021)
13. A. Budiarto, B. Mahesworo, A. A. Hidayat, I. Nurlaila, B. Pardamean, *Gaussian mixture model implementation for population stratification estimation from genomics data*, Procedia Computer Science **179**, pp. 202–210 (2021)
14. D. E. Parung, K. Azizatikarna, D. Amirulloh, E. Listiyaningsih, B. Mahesworo, A. Budiarto, Simon, B. Pardamean, *DNAku consumers profile: one of the first direct to customer genetics testing in Indonesia*, IOP Conf. Series: Earth and Environmental Science **794**, pp. 1–9 (2021)
15. A. Budiarto, B. Pardamean, *Explainable supervised method for genetics ancestry estimation*, in 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI) (2021)
16. B. E. Slatko, A. F. Gardner, F. M. Ausubel, *Overview of next generation sequencing technologies*, Molecular Biology **122**, 1, pp. 1–15 (2018)
17. V. Tripathi, P. Kumar, P. Tripathi, A. Kishore, M. Kamle, *Next-generation sequencing (NGS) platforms: an exciting era of genome sequence analysis*, in Microbial Genomics in Sustainable Agroecosystems, pp. 89–110 (Springer, 2019)
18. J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, J. F. Kim, *Genome evolution and adaptation in a long-term experiment with escherichia coli*, Nature **461**, pp. 1243–1247 (2009)
19. Z. D. Blount, J. E. Barrick, C. J. Davidson, R. E. Lenski, *Genomic analysis of a key innovation in an experimental escherichia coli population*, Nature **489**, pp. 513–518 (2012)
20. A. Budiarto, B. Mahesworo, J. Baurley, T. Suparyanto, B. Pardamean, *Fast and effective clustering method for ancestry estimation*, Procedia Computer Science **157**, pp. 306–312 (2019)
21. B. Mahesworo, A. Budiarto, B. Pardamean, *Systematic evaluation of cross population polygenic risk score on colorectal cancer*, Procedia Computer Science, pp. 1–8 (2020)
22. S. Amadeus, T. W. Cenggoro, A. Budiarto, B. Pardamean, *A design of polygenic risk model with deep learning for colorectal cancer in multiethnic Indonesians*, Procedia Computer Science **179**, 2020, pp. 632-639 (2021)
23. K. Chen, J. W. Wallis, M. D. McLellan, D. E. Larson, J. M. Kalicki, C. S. Pohl, S. D. McGrath,

M. C. Wendl, Q. Zhang, D. P. Locke, X. Shi, R. S. Fulton, T. J. Ley, R. K. Wilson, L. Ding, E. R. Mardis, *BreakDancer: an algorithm for high-resolution mapping of genomic structural variation*, Nature Methods **6**, pp. 677–681 (2009)

24. B. Zeitouni, V. Boeva, I. Janoueix-Lerosey, S. Loeillet, P. Legoix-né, A. Nicolas, O. Delattre, E. Barillot, *SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data*, Bioinformatics **26**, 15, pp. 1895–1896 (2010)
25. S. Linnarsson, *Recent advances in DNA sequencing methods - general principles of sample preparation*, Experimental Cell Research **316**, 8, pp. 1339–1343 (2010)
26. A. Healey, A. Furtado, T. Cooper, R. J. Henry, *Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species*, Plant Methods **10**, 1, pp. 1–8 (2014)
27. S. R. Head, H. Kiyomi Komori, S. A. LaMere, T. Whisenant, F. Van Nieuwerburgh, D. Salomon, P. Ordoukhanian, *Library construction for next-generation sequencing: overviews and challenges*, BioTechniques **56**, 2, pp. 61–77 (2014)
28. R. M. Martin, M. Kausch, K. Yap, J. D. Wehr, G. L. Boyer, S. W. Wilhelm, *Metagenome-assembled genome sequences of raphidiopsis raciborskii and planktothrix agardhii from a cyanobacterial bloom in kissena lake, New York, USA*, Microbiology Resource Announcements **10**, 2, pp. 10–11 (2021)
29. J. S. Boden, M. Grego, H. Bolhuis, P. Sánchez-baracaldo, *Draft genome sequences of three filamentous cyanobacteria isolated from brackish habitats*, J. Genomics **9**, pp. 20–25 (2021)
30. A. V. Bryanskaya, A. A. Shipova, A. S. Rozanov, O. A. Volkova, E. V. Lazareva, Y. E. Uvarova, T. N. Goryachkovskaya, S. E. Peltek, *Metagenomics dataset used to characterize microbiome in water and sediments of the lake solenoe (novosibirsk region, Russia)*, Data in Brief **34**, 106709 (2021)
31. J. E. Barrick, G. Colburn, D. E. Deatherage, C. C. Traverse, M. D. Strand, J. J. Borges, D. B. Knoester, A. Reba, A. G. Meyer, *Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq*, BMC Genomics **15**, 1039, pp. 1–17 (2014)
32. D. E. Deatherage, J. E. Barrick, *Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq*, Methods in Molecular Biology **1151**, pp. 165–188 (2015)
33. S. Diamond, B. E. Rubin, R. K. Shultzaberger, Y. Chen, C. D. Barber, S. S. Golden, *Redox crisis underlies conditional light-dark lethality in cyanobacterial mutants that lack the circadian regulator, RpaA*, Proceedings of the National Academy of Sciences of the United States of America **114**, 4, E580–9 (2017)
34. K. S. Walter, C. Colijn, T. Cohen, B. Mathema, Q. Liu, J. Bowers, D. M. Engelthaler, A. Narechania, D. Lemmer, J. Croda, J. R. Andrews, *Genomic variant-identification methods may alter mycobacterium tuberculosis transmission inferences*, Microbial Genomics **6**, 8, pp. 1–16 (2020)
35. H. Derakhshani, S. P. Bernier, V. A. Marko, M. G. Surette, *Completion of draft bacterial genomes by long-read sequencing of synthetic genomic pools*, BMC Genomics **21**, 519, pp. 1–11 (2020)
36. S. R. Miller, H. E. Abresch, N. J. Ulrich, E. B. Sano, A. H. Demaree, A. R. Oman, A. I. Garber, *Bacterial adaptation by a transposition burst of an invading IS element*, Genome Biology and Evolution **13**, 11, pp. 1–12 (2021)
37. B. Langmead, S. L. Salzberg, *Fast gapped-read alignment with bowtie 2*, Nature Methods **9**, pp. 357–360 (2012)
38. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, *1000 Genomes project analysis, 2011, the variant call format and VCFtools*, Bioinformatics **27**, 15, pp. 2156–2158 (2011)
39. M. G. Reese, B. Moore, C. Batchelor, F. Salas, F. Cunningham, G. T. Marth, L. Stein, P. Flicek, M. Yandell, K. Eilbeck, *A standard variation file format for human genome sequences*, Genome biology **11**, R88, pp. 1–9 (2010)
40. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, *Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration*, Briefings in Bioinformatics **14**, 2, pp. 178–192 (2013)
41. U. Väli, M. Brandström, M. Johansson, H. Ellegren, *Insertion-deletion polymorphisms (indels) as genetic markers in natural populations*, BMC genetics **9**, pp. 1–8 (2008)
42. R. Ohbayashi, S. Hirooka, R. Onuma, Y. Kanesaki, Y. Hirose, Y. Kobayashi, T. Fujiwara, C. Furusawa, S. Miyagishima, *Evolutionary changes in dnaA-dependent chromosomal replication in cyanobacteria*, Frontiers in Microbiology **11**, 786, pp. 1–14 (2020)
43. M. Dann, E. M. Ortiz, M. Thomas, A. Guljamow, M. Lehmann, H. Schaefer, D. Leister, *Enhancing photosynthesis at high light levels by adaptive laboratory evolution*, Nature Plants **7**, pp. 681–695 (2021)
44. W. Xu, H. Tang, Y. Wang, P. R. Chitnis, *Proteins of the cyanobacterial photosystem I*, Biochimica et Biophysica Acta **1507**, 1–3, pp. 32–40 (2001)
45. R. M. Anur, N. Mufithah, W. D. Sawitri, H. Sakakibara, B. Sugiharto, *Overexpression of sucrose phosphate synthase enhanced sucrose content and biomass production in transgenic sugarcane*, Plants **9**, 200, pp. 1–11 (2020)