



Editorial: Curriculum Applications in Microbiology: Bioinformatics in the Classroom

Melanie Crystal Melendrez^{1*}, Sophie Shaw², C. Titus Brown³, Brad W. Goodner⁴ and Christopher Kvaal⁵

¹ Anoka-Ramsey Community College, Cambridge, MN, United States, ² Centre for Genome Enabled Biology and Medicine, University of Aberdeen, Aberdeen, United Kingdom, ³ Department of Population Health and Reproduction, University of California, Davis, Davis, CA, United States, ⁴ Hiram College, Hiram, OH, United States, ⁵ Department of Biology, St. Cloud State University, St. Cloud, MN, United States

Keywords: bioinformatics and computational biology, undergraduate research, microbiology, curriculum—undergrad and postgrad, software, genomics, science technology engineering mathematics

Editorial on the Research Topic

Curriculum Applications in Microbiology: Bioinformatics in the Classroom

OPEN ACCESS

Edited by:

Iain Sutcliffe,
Northumbria University,
United Kingdom

Reviewed by:

Joao Carlos Setubal,
University of São Paulo, Brazil
David W. Ussery,
University of Arkansas for Medical
Sciences, United States

*Correspondence:

Melanie Crystal Melendrez
melanie.melendrez-vallard@
anokaramsey.edu

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 04 May 2021

Accepted: 07 June 2021

Published: 01 July 2021

Citation:

Melendrez MC, Shaw S, Brown CT,
Goodner BW and Kvaal C (2021)
Editorial: Curriculum Applications in
Microbiology: Bioinformatics in the
Classroom.
Front. Microbiol. 12:705233.
doi: 10.3389/fmicb.2021.705233

John Naisbitt stated in his 1982 book Megatrends, “We are drowning in information but starved for knowledge.” The statement, made nearly 40 years ago, seems acutely applicable in today’s scientific and academic world. Reviews by Barba et al. and van Dijk et al., provide a nice historical perspective on the growth of sequencing technology. Over three decades, sequencing technology has improved greatly from 1987 when the first ABI automated sequencing machine went to market up through the mid 2010s when next generation sequencing platforms from 454 Life Sciences, Illumina and other companies were outputting up to 1,800 Gb per run (Barba et al., 2014; van Dijk et al., 2014). Technology has since progressed even further with the development of long-read and single molecule (Pacific Biosciences, Illumina, Oxford Nanopore, and 10X Genomics) sequencing systems that can output terabytes of data, per run, in a matter of days (van Dijk et al., 2018). Specifically, in the areas of genomics, proteomics, and transcriptomics, we are now producing upwards of 1 zetta-bases/year (Stephens et al., 2015). The explosion of data has increased the demand for hardware and software development to manage and analyze the data as well as qualified personnel in bioinformatics to sift through the outputs to draw meaningful conclusions. A report from Reports and Data states the global bioinformatics market is projected to reach 18.96 billion USD by 2026 (Reports and Data, 2019) and this means re-thinking not only how we store data but how we train the next generation of scientists. The greatest needs identified in various surveys compiled by the NSF, ELIXIR-UK, and EMBL-ABR include: (i) data quality and control (ii) data analysis skills in visualization and interpretation, (iii) data mining, manipulation and management, (iv) analysis reproducibility, and (v) statistics (Kanwal et al., 2017; Kim et al., 2018; Attwood et al., 2019).

The large quantity of data available for analysis in many scientific fields is both a strength and a weakness in bioinformatic analysis. There are several databases and repositories available to acquire sequence data (INSDC: ENA+DNA Data bank of Japan and Genbank SRA, GISAID) and it is essential to know that not all data may be handled the same way. This variability in the quality of data that is released to scientists and the public at large can result in low quality data being analyzed and potentially spurious conclusions. A prime example can be seen in the current pandemic where SARS-CoV-2 sequences can be obtained from several different databases and analyzed in real-time. In the haste to make genomes available, the quality of what is released has been variable with challenges in consistent nomenclature (Gozashti and Corbett-Detig, 2021),

and genomes containing errors created by sequencing artifacts, sample preparation, consensus calling approaches, or contamination (De Maio et al., 2020a,b; Van Noorden, 2021). Analysis of these genomes, even among experts, can lead to data misinterpretation, over-interpretation, and confusion on important topics such as SARS-CoV-2 origins (Andersen et al., 2020; Zhang et al., 2020; Wacharapluesadee et al., 2021). However, it is important to recognize these challenges to big data quality control, management, analysis, and reproducibility are not unique to SARS-CoV-2 but are systemic in many subfields of bioinformatics such as microbiome analysis (Katsnelson, 2019), metatranscriptomics (Shakya et al., 2019), and RNA-seq (Simoneau et al., 2021) analysis.

Recognition of these short-comings of big data acquisition, quality control, reproducibility, management, and analysis across bioinformatics disciplines have led to improvements in next generation sequencing workflows and quality control (Charre et al., 2020; De Maio et al., 2020b; Van Damme et al., 2021), efforts to use provenance, github, and docker containers to facilitate reproducibility (Kanwal et al., 2017; Kulkarni et al., 2018; Menegidio et al., 2018; Bolyen et al., 2019; Wercelens et al., 2019), nomenclature clarification (Rambaut et al., 2020, 2021), an increased emphasis on workflow automation (Reiter et al., 2021), and database curation and consolidation (Heard et al., 2021). As the tools and refinements to how scientists manage and analyze data continue to move forward, the demand for qualified big data analysts, statisticians, and bioinformaticians is increasing rapidly (Gómez-López et al., 2019; Terry, 2019; Tammi et al., 2020).

To address the need for big data management and analytical skill sets, many university programs have emerged offering certificates, Master's degrees and even Ph.D. degrees in the field of bioinformatics. The most recent guidance on bioinformatics core competencies has highlighted the importance of developing informatics skill sets early in the undergraduate curriculum (Welch et al., 2014; Vincent and Charette, 2015; Mulder et al., 2018; Wilson Sayres et al., 2018; Tractenberg et al., 2019). However, few curricula at the undergraduate level introduce big data analytics and bioinformatics systematically and many students graduate without a full understanding of what bioinformatics is or how it can be used to solve biological problems.

Several bioinformatic disciplines: i.e., metagenomics, genome construction/annotation, pathogen discovery, phylogenetics, metabolomics, and transcriptomics, have well-known workflows that teach valuable skills in data management, analytics, interpretation, and troubleshooting, but have yet to be translated to the classroom. Additionally, while many microbiology instructors recognize the importance of integrating more research, real-world datasets, and informatics into the classroom, they feel their training is inadequate, their curriculum is already over-full, or students do not appear particularly interested or prepared for such topics in the course (Williams et al., 2019). For many instructors, it can be daunting to put together bioinformatics curriculum modules if you are not familiar with the software or general topics within bioinformatics that students can explore.

This Research Topic focuses on bringing both research and educational communities together; encouraging researchers to translate their studies and pipelines into teaching tools and curriculum, and encouraging educators to dive into messy real-world datasets when teaching microbiology. Much of the challenge in implementing research or bioinformatics focused modules in the undergraduate classroom revolves around implementation. Bennet discusses strategies for blending your classroom to incorporate undergraduate research and bioinformatics modules into your curriculum design (CURE). Bennet takes a “workshop” or “project-based” approach to introduce the often complicated and challenging topic of RNA-Seq analysis (Conesa et al., 2016; Bennett) and discusses the long term outcomes for students experiencing this particular CURE as well as educational applications.

Another challenge in implementation of bioinformatics workflows in the classroom is the requirement for background experience in a variety of topics, both biological and computational. While many biology instructors are comfortable introducing and expanding on biological topics related to research and design, they are less comfortable discussing the computing aspects of bioinformatic analysis such as coding languages, data quality control, and data management. Several papers in the special topic discuss data workflows that utilize Microsoft Excel (Mitchell et al.; Hankey et al.; Kruchten). While many individuals working in advanced bioinformatic analysis may cringe at the idea of excel data analysis and tables, this program is well-used in classrooms globally and many instructors are comfortable with implementing data analysis and mathematical functions in the Excel environment. Programs such as Excel can provide a bridge between the user-friendly, GUI-based interfaces and the world of command-line applications (CLI). Kruchten, in particular, offers a nice comparison of the use of Excel vs. the R statistical language when analyzing metagenomic datasets and this may serve as motivation for instructors to explore other programming and CLI-based workflows (Kruchten). Topic papers in the methods category show instructors how to introduce, discuss, and/or implement coding languages and CLI-based bioinformatics in their classroom such as python/R for microbiome analysis (Rosen and Hammrich), basic command line proficiency in analyzing genome scale data for microbial isolates (Petrie and Xie), and how to conduct metagenomic analysis using the R statistical package (Kruchten) or QIIME, which contains its own language and syntax for implementation (Bolyen et al., 2019; Rosen and Hammrich).

Additional topic papers contain curricular designs for introducing and teaching a variety of bioinformatic analysis skills in the classroom without the need for teaching additional modules on coding skills. Topic papers discuss gene discovery and genome annotation using a variety of free web-accessible programs (Amatore et al.; Koury et al.; Martins et al.), microbiome analysis using PUMAA (Mitchell et al.), 16S amplicon identification using DNALC and NCBI-BLAST databases and the DNA Subway software program (Tawde and Williams; Williams et al., 2014), metagenomics analysis using MG-RAST and the MicrobiomeAnalyst program (Meyer et al.,

2019; Chong et al., 2020; Baker et al.), phage hunting using PHASTER and iTOL programs (Arndt et al., 2019; Letunic and Bork, 2019; Martinez-Vaz and Mickelson), and Cancer data analysis using The Cancer Genome Atlas (TCGA; Hankey et al.).

To account for variable quality of datasets analyzed in the classroom, special topic studies used already published, curated, data from the Cancer Genome Atlas or GENI-ACT toolkit (Hankey et al.; Koury et al.) or pre-curated genomes for genome prediction exercises rather than raw data from databases (Martins et al.). Studies that made use of raw data or minimally curated data utilized embedded quality assessment tools and discussion modules on data cleanup within their curriculum methods and workflow (Amatore et al.; Kruchten; Tawde and Williams; Petrie and Xie; Baker et al.; Mitchell et al.). However, discussion and training on quality and data management needs to be ongoing; especially given data is being reused for educational purposes. Wilkinson and colleagues proposed the FAIR guiding principles to support the accessibility, findability, interoperability, and reusability of data in science (FAIR principles for data stewardship, 2016; Wilkinson et al., 2016) and there are workshops available on how to get started with “FAIR data” (<https://mdibl.org/course/applied-bioinformatics-2021/>). These principles should be considered widely in addition to the use of provenance and contained workflows or containers such as those mentioned earlier. In the overwhelming world of big data analysis it will be important for instructors to translate complex analysis techniques to their novice students; a key challenge is balancing quality and rigor with simplicity.

Finally the topic papers extend into existing scientific communities, where skills needed for data analysis are lacking by a large number of current researchers and professionals tasked to conduct bioinformatics analysis and interpretation. Therefore, workshops to educate existing researchers and laboratory personnel, from the level of graduate student to principal investigator, have become more frequent. These professional development and “train the trainer” workshops are attractive in that they are intensive short term experiences that teach very specific skill sets related to computational jobs in the field (McGrath et al., 2019). The Physalia courses (<https://www.physalia-courses.org/>), Cold Spring Harbor Laboratory Short Courses (<https://meetings.cshl.edu/courses.html>), and various workshops offered by the Evolution and Genomics training team (<http://evomics.org/workshops/>) and the MDI Biological Laboratory (<https://mdibl.org/course/bioinformatics-t3-2021/>) are a few examples of training experiences that undergraduates, graduates and professional personnel can use to augment their skill sets in the field of genetic analysis and computational

biology. Internationally, these short term intensive educational opportunities, putting bioinformatics in the classroom, have proven useful in bringing staff and personnel up to date on the latest technologies and analysis capabilities to increase job performance and institute mission output. The BioCANET network in Central America (Orozco et al., 2013), Walter Reed Army Institute of Research (WRAIR) in South America (Pollett et al., 2016), H3Africa consortium in Africa (Aron et al., 2017; Ahmed et al., 2018; Shaffer et al., 2019), and APBioNet in Asia (Khan et al., 2013; Ahmad et al., 2019) are all aimed at increasing capacity for educational and research institutions in the areas of data management, systems administration, biostatistics, genome wide association studies, next generation sequencing analysis, metagenomics, and virology; and all have had success using this educational format. Our topic supports this educational “workshop” format of continued training for professional personnel through a paper by Maljkovic Berry et al., on implementation of a bioinformatics workshop for laboratory and research personnel at a US Department of Defense laboratory located in Kisumu, Kenya.

Special topic papers detail curriculum set up and implementation of bioinformatics modules or coding contain supplemental material to facilitate readers in their own implementation of the module or curriculum design in their classroom. We hope to convey through this topic the versatility of instructional designs that can be used to teach students at all levels of expertise, from high school to established professionals, how to leverage the strength of coding, software, and computational analysis to accomplish their research goals and further scientific teaching and discovery.

AUTHOR CONTRIBUTIONS

MM contributed to the initial concept of the special topic and wrote the initial draft of the editorial. All authors contributed to the design and proposal of the special topic, participated as active guest editors to manage manuscripts and oversee the special topic, and contributed to editorial revision and final approval for submission.

ACKNOWLEDGMENTS

We would like to thank all of the authors who submitted to this special topic, committed to the furthering of academic creativity, excellence, and rigor in the challenging and virtual instructional world of SARS-CoV-2 (COVID-19). To you and all of our educators globally, you are indispensable.

REFERENCES

- Ahmad, S., Gromiha, M. M., Raghava, G. P. S., Schönbach, C., and Ranganathan, S. (2019). APBioNet’s annual International Conference on Bioinformatics (InCoB) returns to India in 2018. *BMC Genomics* 19:266. doi: 10.1186/s12864-019-5582-8
- Ahmed, A. E., Mpangase, P. T., Panji, S., Baichoo, S., Souilmi, Y., Fadlelmola, F. M., et al. (2018). Organizing and running bioinformatics hackathons within Africa: the H3ABioNet cloud computing experience. [version 2; peer review: 2 approved, 1 approved with reservations]. *AAS Open Res.* 1:9. doi: 10.12688/aasopenres.12847.1

- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.* 26, 450–452. doi: 10.1038/s41591-020-0820-9
- Arndt, D., Marcu, A., Liang, Y., and Wishart, D. S. (2019). PHAST, PHASTER and PHASTEST: tools for finding prophage in bacterial genomes. *Brief. Bioinformatics* 20, 1560–1567. doi: 10.1093/bib/bbx121
- Aron, S., Gurwitz, K., Panji, S., and Mulder, N. (2017). H3abionet: developing sustainable bioinformatics capacity in Africa. *EMBnet J.* 23:886. doi: 10.14806/ej.23.0.886
- Attwood, T. K., Blackford, S., Brazas, M. D., Davies, A., and Schneider, M. V. (2019). A global perspective on evolving bioinformatics and data science training needs. *Brief. Bioinformatics* 20, 398–404. doi: 10.1093/bib/bbx100
- Barba, M., Czosnek, H., and Hadidi, A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 6, 106–136. doi: 10.3390/v6010106
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9
- Charre, C., Ginevra, C., Sabatier, M., Regue, H., Destras, G., Brun, S., et al. (2020). Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol.* 6:veaa075. doi: 10.1093/ve/veaa075
- Chong, J., Liu, P., Zhou, G., and Xia, J. (2020). Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nat. Protoc.* 15, 799–821. doi: 10.1038/s41596-019-0264-1
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17:13. doi: 10.1186/s13059-016-0881-8
- De Maio, N., Gozashti, L., Turakhia, Y., Walker, C., Lanfear, R., Corbett-Detig, R., et al. (2020a). *Updated Analysis With Data From 12th June 2020*. virological.org. Available online at: <https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473/11> (accessed May 28, 2021).
- De Maio, N., Walker, C., Borges, R., Weilguy, L., Slodkowitz, G., and Goldman, N. (2020b). *Issues With SARS-CoV-2 Sequencing Data*. virological.org. Available online at: <https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473> (accessed May 28, 2021).
- FAIR principles for data stewardship. (2016). *Nat. Genet.* 48:343. doi: 10.1038/ng.3544
- Gómez-López, G., Dopazo, J., Cigudosa, J. C., Valencia, A., and Al-Shahrour, F. (2019). Precision medicine needs pioneering clinical bioinformaticians. *Brief. Bioinformatics* 20, 752–766. doi: 10.1093/bib/bbx144
- Gozashti, L., and Corbett-Detig, R. (2021). Shortcomings of SARS-CoV-2 genomic metadata. *BMC Res. Notes* 14:189. doi: 10.1186/s13104-021-05605-9
- Heard, E., Birney, E., Apweiler, R., Bloomberg, N., Cochrane, G., Lauer, K., et al. (2021). *Open Letter: Support Data Sharing for COVID-19*. COVID-19 Data Portal. Available online at: <https://www.covid19dataportal.org/support-data-sharing-covid19> (accessed May 28, 2021).
- Kanwal, S., Khan, F. Z., Lonie, A., and Sinnott, R. O. (2017). Investigating reproducibility and tracking provenance - a genomic workflow case study. *BMC Bioinformatics* 18:337. doi: 10.1186/s12859-017-1747-0
- Katsnelson, A. (2019). Standards seekers put the human microbiome in their sights. *ACS Cent. Sci.* 5, 929–932. doi: 10.1021/acscentsci.9b00557
- Khan, A. M., Tan, T. W., Schönbach, C., and Ranganathan, S. (2013). APBioNet-transforming bioinformatics in the Asia-Pacific region. *PLoS Comput. Biol.* 9:e1003317. doi: 10.1371/journal.pcbi.1003317
- Kim, Y.-M., Poline, J.-B., and Dumas, G. (2018). Experimenting with reproducibility: a case study of robustness in bioinformatics. *Gigascience* 7:giy077. doi: 10.1093/gigascience/giy077
- Kulkarni, N., Alessandri, L., Panero, R., Arigoni, M., Olivero, M., Ferrero, G., et al. (2018). Reproducible bioinformatics project: a community for reproducible bioinformatics analysis pipelines. *BMC Bioinformatics* 19:349. doi: 10.1186/s12859-018-2296-x
- Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- McGrath, A., Champ, K., Shang, C. A., van Dam, E., Brooksbank, C., and Morgan, S. L. (2019). From trainees to trainers to instructors: sustainably building a national capacity in bioinformatics training. *PLoS Comput. Biol.* 15:e1006923. doi: 10.1371/journal.pcbi.1006923
- Menegidio, F. B., Jabes, D. L., Costa de Oliveira, R., and Nunes, L. R. (2018). Dugong: a Docker image, based on Ubuntu Linux, focused on reproducibility and replicability for bioinformatics analyses. *Bioinformatics* 34, 514–515. doi: 10.1093/bioinformatics/btx554
- Meyer, F., Bagchi, S., Chaterji, S., Gerlach, W., Grama, A., Harrison, T., et al. (2019). MG-RAST version 4—lessons learned from a decade of low-budget ultra-high-throughput metagenome analysis. *Brief. Bioinformatics* 20, 1151–1159. doi: 10.1093/bib/bbx105
- Mulder, N., Schwartz, R., Brazas, M. D., Brooksbank, C., Gaeta, B., Morgan, S. L., et al. (2018). The development and application of bioinformatics core competencies to improve bioinformatics training and education. *PLoS Comput. Biol.* 14:e1005772. doi: 10.1371/journal.pcbi.1005772
- Orozco, A., Morera, J., Jiménez, S., and Boza, R. (2013). A review of bioinformatics training applied to research in molecular medicine, agriculture and biodiversity in Costa Rica and Central America. *Brief. Bioinformatics* 14, 661–670. doi: 10.1093/bib/bbt033
- Pollett, S., Leguia, M., Nelson, M. I., Maljkovic Berry, I., Rutherford, G., Bausch, D. G., et al. (2016). Feasibility and effectiveness of a brief, intensive phylogenetics workshop in a middle-income country. *Int. J. Infect. Dis.* 42, 24–27. doi: 10.1016/j.ijid.2015.11.001
- Rambaut, A., Holmes, E. C., O’Toole, Á., Hill, V., McCrone, J. T., Ruis, C., et al. (2021). Addendum: a dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 6:415. doi: 10.1038/s41564-021-00872-5
- Rambaut, A., Holmes, E. C., O’Toole, Á., Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407. doi: 10.1038/s41564-020-0770-5
- Reiter, T., Brooks, P. T., Irber, L., Joslin, S. E. K., Reid, C. M., Scott, C., et al. (2021). Streamlining data-intensive biology with workflow systems. *Gigascience* 10:giaa140. doi: 10.1093/gigascience/giaa140
- Reports and Data (2019). *Global Bioinformatics Market To Reach USD 18.96 Billion By 2026 | Reports and Data*. Available online at: <https://www.globenewswire.com/news-release/2019/05/14/1823970/0/en/Global-Bioinformatics-Market-To-Reach-USD-18-96-Billion-By-2026-Reports-And-Data.html> (accessed July 1, 2020).
- Shaffer, J. G., Mather, F. J., Wele, M., Li, J., Tangara, C. O., Kassogue, Y., et al. (2019). Expanding research capacity in Sub-Saharan Africa through informatics, bioinformatics, and data science training programs in Mali. *Front. Genet.* 10:331. doi: 10.3389/fgene.2019.00331
- Shakya, M., Lo, C.-C., and Chain, P. S. G. (2019). Advances and challenges in metatranscriptomic analysis. *Front. Genet.* 10:904. doi: 10.3389/fgene.2019.00904
- Simoneau, J., Dumontier, S., Gosselin, R., and Scott, M. S. (2021). Current RNA-seq methodology reporting limits reproducibility. *Brief. Bioinformatics* 22, 140–145. doi: 10.1093/bib/bbz124
- Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., et al. (2015). Big data: astronomical or genomic? *PLoS Biol.* 13:e1002195. doi: 10.1371/journal.pbio.1002195
- Tammi, M., Fei, K. T., and Low, L. (2020). *Career Outlook: Data and Bioinformatics Scientists to 2026. Dr Martti Knowledge is Power*. Available online at: https://bioinformatics-home.com/blog/jobs_landscape_2018_2020.html (accessed June 1, 2021).
- Terry, M. (2019). *Careers in Bioinformatics: Hot and Getting Hotter*. BioSpace. Available at: <https://www.biospace.com/article/careers-in-bioinformatics-hot-and-getting-hotter/> (accessed June 1, 2021).
- Tractenberg, R. E., Wilkinson, M. R., Bull, A. W., Pellathy, T. P., and Riley, J. B. (2019). A developmental trajectory supporting the evaluation and achievement of competencies: articulating the Mastery Rubric for the nurse practitioner (MR-NP) program curriculum. *PLoS ONE* 14:e0224593. doi: 10.1371/journal.pone.0224593
- Van Damme, R., Hölzer, M., Viehweger, A., Müller, B., Bongcam-Rudloff, E., and Brandt, C. (2021). Metagenomics workflow for hybrid assembly, differential coverage binning, metatranscriptomics and pathway analysis (MUFFIN). *PLoS Comput. Biol.* 17:e1008716. doi: 10.1371/journal.pcbi.1008716

- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426. doi: 10.1016/j.tig.2014.07.001
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. *Trends Genet.* 34, 666–681. doi: 10.1016/j.tig.2018.05.008
- Van Noorden, R. (2021). Scientists call for fully open sharing of coronavirus genome data. *Nature* 590, 195–196. doi: 10.1038/d41586-021-00305-7
- Vincent, A. T., and Charette, S. J. (2015). Who qualifies to be a bioinformatician? *Front. Genet.* 6:164. doi: 10.3389/fgene.2015.00164
- Wacharapluesadee, S., Tan, C. W., Maneeorn, P., Duengkae, P., Zhu, F., Joyjinda, Y., et al. (2021). Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* 12:972. doi: 10.1038/s41467-021-21240-1
- Welch, L., Lewitter, F., Schwartz, R., Brooksbank, C., Radivojac, P., Gaeta, B., et al. (2014). Bioinformatics curriculum guidelines: toward a definition of core competencies. *PLoS Comput. Biol.* 10:e1003496. doi: 10.1371/journal.pcbi.1003496
- Wercelens, P., da Silva, W., Hondo, F., Castro, K., Walter, M. E., Araújo, A., et al. (2019). Bioinformatics workflows with nosql database in cloud computing. *Evol. Bioinform. Online* 15:1176934319889974. doi: 10.1177/1176934319889974
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18
- Williams, J., McKay, S., Khalfan, M., Hilgert, U., Lauter, S., Jeong, E.-S., et al. (2014). *DNA Subway - An Educational Bioinformatics Platform for Gene and Genome Analysis: DNA Barcoding, and RNA-Seq*. Vancouver, BC: American Society of Animal Science (ASAS). Available online at: https://www.asas.org/docs/default-source/wcgalp-proceedings-oral/227_paper_9798_manuscript_948_0.pdf (accessed April 20 2021).
- Williams, J. J., Drew, J. C., Galindo-Gonzalez, S., Robic, S., Dinsdale, E., Morgan, W. R., et al. (2019). Barriers to integration of bioinformatics into undergraduate life sciences education: a national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction. *PLoS ONE* 14:e0224288. doi: 10.1371/journal.pone.0224288
- Wilson Sayres, M. A., Hauser, C., Sierk, M., Robic, S., Rosenwald, A. G., Smith, T. M., et al. (2018). Bioinformatics core competencies for undergraduate life sciences education. *PLoS ONE* 13:e0196878. doi: 10.1101/170993
- Zhang, T., Wu, Q., and Zhang, Z. (2020). Probable Pangolin Origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr. Biol.* 30, 1346–1351.e2. doi: 10.1016/j.cub.2020.03.022

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Melendrez, Shaw, Brown, Goodner and Kvaal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.