INTRODUCTION

High throughput DNA sequencing technologies, also known as next-generation sequencing (NGS) techniques, have evolved rapidly in recent years and have had an impact on all branches of biological science, including microbiology. Within the field of microbiology, the knowledge that can be gained relating to individual strains (through genomics) and populations of microorganisms (through metagenomics) via the use of these technologies has the potential to play a key role in several applications, for example in medical and veterinary diagnostic, forensic genomics, improving the monitoring of food, and food quality and safety (1). NGS methods provide larger amounts of data, shorter sequencing times and reduced costs than the older ‘Sanger’ sequencing technology (2). NGS techniques produce millions of individual DNA sequence reads attributable to several different samples at the same time with the use of specific adaptors, called indexes, which are assigned to each sample in advance in order to allow researchers to subsequently identify and assign the sequences corresponding to these samples. The chemistry underlying NGS facilitates this process and is thus of great relevance to science students. Among the NGS technologies, those provided by Illumina on their MiSeq, NextSeq, and HiSeq platforms are most commonly used. The common working principle underlying these instruments is a clonal amplification process that incorporates DNA bases into a nucleic acid chain while simultaneously allowing their identification through the emission of a unique fluorescent signal, which is used to determine the order of the DNA sequence. These instruments sequence basic units that are called ‘reads’ through a solid-phase hybridization (3). The reads are then analyzed by bioinformaticians through a process that involves several steps, starting from the selection and trimming of the reads that have good quality (4), then combining overlapping sequences in order to assemble longer continuous fragments called contigs (5), and finally obtaining a draft assembly that, in the case of a whole genome sequencing (WGS) of a bacterium, consists of the complete sequence of bases found within chromosomal DNA and any other stably maintained extra-chromosomal genetic elements that might be present (6). In addition to WGS of microorganisms, NGS can, among other things, be used for 16S ribosomal RNA (rRNA) sequencing to identify the bacteria present within a given sample and to compare the proportions in which they are present across samples. This application is widely used for phylogeny and taxonomy studies, including the investigation of complex microbiomes or environments that are difficult or impossible to study using culture-based techniques (7).

In spite of the widespread application of NGS, and references to it in biology classes, professional courses, and articles in general science magazines, content relating to the principle and chemistry of NGS and the ways in which bioinformatics facilitates data analysis is somewhat lacking and would benefit from the development of new hands-on educational approaches. Here, we address this issue. More specifically, with this exercise, students will take part in a practical activity that will allow the identification of unknown pathogens that have been virtually sequenced. The participants will be provided with DNA read sequences that they will build manually with LEGO blocks (to represent DNA bases) associated with their indexes. The reads will then be oriented and aligned in order to produce contigs and, in turn, the final assembly, thereby mimicking the computational approach employed by bioinformaticians. Finally, the participants will compare the sequences obtained with representative short versions of the genomes of pathogens that are provided in a printed format. The students will “identify” the bacteria from which the sequenced (LEGO) DNA originated by comparison. The material provided is suitable for six different groups of participants and also includes an introductory presentation for the combined class that covers the principle of NGS technology, its applications, and an overview of the activity. In addition, a formative assessment quiz is provided, as are the results of a previous
survey that was administered before and after the pilot activity. This survey showed that the exercise helped to fill gaps in the participants’ knowledge with respect to NGS through a hands-on group activity that is safe, economical, and easy to store.

**PROCEDURE**

**Background for the class**

This activity is developed with a view to being a component of an introductory technical molecular biology laboratory course for students majoring in microbiology and/or other life sciences who have already received some instruction related to cellular and molecular biology and a theoretical introduction to aspects of DNA sequencing. Other potential participants are molecular laboratory technicians involved in NGS who want to familiarize themselves with the principles underlying NGS techniques and associated bioinformatic approaches. The activity is also recommended for instructors who want to add practical exercises and new approaches to encourage interest in new technologies and computational analyses. NGS can be introduced with the support of generic (8–10), and specific (11, 12) papers that relate to the application of this technology. Indeed, the students can be introduced to the specific use of NGS to reflect their interests and the curriculum. This may include, for example, the application of NGS to food microbiology (13), food safety (14), study of the human microbiome (7), or the discovery and evolutionary analysis of plant viruses (15).

**Learning time and learning objectives**

The activity is designed to take place during a 55-minute lesson comprised of a 15-minute slide-based lecture (Appendix 1), 25 minutes of activity, 10 minutes for the post-activity quiz (Appendix 4), and 5 minutes for clean-up. The major goal is to analyze sequences and identify the correct bacterial genome from a selection of six provided. The participants are asked to complete the quiz before and after the activity to determine the activity’s success.

**Materials**

A complete list of the materials and the accompanying instructors’ notes are available in Appendix 2. In order to carry out the activity, instructors should prepare the kit for a maximum of six different groups in advance. The instructors’ notes also contain the “reference genomes”, i.e., the solutions for the activity consisting of sequences that match the contigs created during the exercise.

**Student instructions**

A 15-minute PowerPoint lecture (Appendix 1) covering the introduction, syllabus, and an overview of NGS methods should be presented to the class. The instructions for the participants are included in Appendix 2 as an “Exercise form” available in six different versions (2a/2b/2c/2d/2e/2f).

At the beginning of the practical session, students form groups of no more than three to allow each person an opportunity to manipulate the LEGO bricks and record the data. To simplify the activity, the model reads consist of only one segment, each visualized by a LEGO tower consisting of bricks of the same width with a different color for each of the bases (blue for A, red for G, yellow for C, and green for T). The indexes are represented by Lego round plates that consist of cylinders of the same thickness in size with different heights. These bricks are transparent but respect the color code of the bases (blue for A, red for G, yellow for C, and green for T). Each group receives a box kit with a non-transparent LEGO plate that is used to model the flow-cell on which sequencing templates are immobilized, and an envelope containing the 64 pieces of 1×1 Lego bricks (16 for each base) and 32 pieces of 1×1 Lego round plates (8 for each base). The participants will follow the instructions contained in the exercise form (Appendix 2), and each group will be provided with the list of reads to build and the index sequences that can be chosen. A basic overview of index ligation and the pooling of the samples is provided in slides 18 to 20 of the lecture (Appendix 1). In the second part of the exercise, the students overlap the read sequences in order to create the contigs, and they will write the final sequence within a table printed on a transparent sheet provided with the kit (Appendix 4). The sequence obtained will be compared with the six reference genomes provided (Appendix 3) in order to find the perfect match and identify the unknown microorganism.

The sequences presented in this activity are short stretches of DNA sequences that have been extracted from the genomes of the following common foodborne pathogens: *Listeria monocytogenes* (GenBank: CP011345.1), *Staphylococcus aureus* (GenBank: CP018205.1CP010151), *Escherichia coli* (GenBank: CP010151.1), *Salmonella enterica* (GenBank: CP012151.1), *Campylobacter jejuni* (GenBank: CP017673.1), and *Clostridium botulinum* (GenBank: CP013243.1).

**FIGURE 1. Bricks fixed in the flow-cell. In the simulation are represented 8 reads of 7 bases, each with indexes of 3 base lengths. (See Appendix 2 for the list of materials and notes for the instructor.)
The DNA fragment generated by hand manipulation, contig assembly, and comparison with the reference genomes mimic the computational analysis performed by bioinformaticians. It is recommended that the instructor walk around the room during the class and interact with the students. The remaining class time can be dedicated to the quiz (Appendix 5), which takes about 10 minutes, leaving 5 minutes for clean-up.

**Assessment**

This activity was tested with undergraduate students who were being taught the principles of NGS and bioinformatics for the first time and microbiology technicians who had not previously had the opportunity to have hands-on training in sequencing and bioinformatics. A total of 90 participants engaged in this activity. A quiz (Appendix 5) was administered by e-mail to the participants a day before the activity (pre-quiz) and at the end of the activity (post-quiz). Questions 3, 4, and 5 were used to evaluate the learning outcome at the end of the activity.

**CONCLUSION**

The presented exercise offers a straightforward and efficient way to visualize the mechanisms underlying DNA sequencing and contig generation using sequencing by synthesis. In addition, the hands-on activity provides insight into the processes underlying the computational analysis that is the basis for sequence assembly.

Results of the pre- and post-quiz are shown in Appendix 6. After the activity, all 90 students (100%) answered the questions correctly (including questions number 1, 2, and 6, which were not included in the pre-quiz).

**Possible modifications**

Possible extensions of the activity may include the acquisition of sequential pictures through the use of a smartphone from above the plates in order to simulate the solid-phase cluster amplification and acquisition of the images of the flow-cell by sequencers (Fig. 3). The participants will fix the reads on the surface of the plate in a random manner and will then add the special labeled nucleotide (step 1) that can consist of Lego bricks of the same color with a fluorescent brick attached. This imitates the polymerase activity that adds a complementary nucleotide to the one already fixed in the flow-cell; a picture from above will show the first base added (step 2); and the denaturing stage will remove the fluorescent brick (step 3). Repetition of steps 1 to 3 over a fixed number of cycles will complete the exercise.

In our experience, students engaged actively and enthusiastically in the described activity and were fascinated by the technologies and applications available thanks to this simple but effective way of explaining NGS and associated bioinformatic approaches.

**SUPPLEMENTAL MATERIALS**

Appendix 1: Lecture copy
Appendix 2: List of materials and instructor’s notes
Appendix 2: Exercise forms 2a/2b/2c/2d/2e/2f
Appendix 3: Reference genomes table
Appendix 4: Tables for recording the sequences obtained
Appendix 5: Formative assessment quiz
Appendix 6: Results of the pre- and post-quiz (pilot activity)

**ACKNOWLEDGMENTS**

Sincere thanks to researchers and attendees of the Food Control and Production Hygiene Unit Laboratory course series entitled “New technologies applied in food control laboratory” (Istituto Zooprofilattico Sperimentale del Piemonte, Liguria e Valle d’Aosta, Turin, Italy), who were willing to provide informal feedback on this exercise. GM and PDC are funded by Science Foundation Ireland in the form of a center grant (APC Microbiome Institute grant number SFI/12/RC/2273). Research in the Cotter laboratory...
is also funded by Science Foundation Ireland through the PI award “Obesibiotics” (11/PI/1137). The authors declare that there are no conflicts of interest.

REFERENCES

1. Metzker ML. 2010. Sequencing technologies—the next generation. Nat Rev Genet 11:31–46.
2. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. 2013. The next-generation sequencing revolution and its impact on genomics. Cell 155:27–38.
3. Horn S. 2012. Target enrichment via DNA hybridization capture. Methods Mol Biol 840:177–188.
4. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.
6. Antipov D, Hartwick N, Shen M, Raiko M, Pevzner PA. 2016. PlasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics 32:3380–3387.
7. Clooney AG, Fouhy F, Sleator RD, O’Driscoll A, Stanton C, Cotter PD, Claesson MJ. 2016. Comparing apples and oranges?: Next generation sequencing and its impact on microbiome analysis. PLoS One 11:e0148028.
8. Shendure J, Ji H. 2008. Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145.
9. Mardis ER. 2008. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402.
10. Glenn TC. 2011. Field guide to next-generation DNA sequencers. Mol Ecol Resour 11:759–769.
11. Ansorge WJ. 2009. Next generation DNA sequencing techniques. N Biotechnol 25:195–203.
12. Lu H, Giordano F, Ning Z. 2016. Oxford nanopore MiniON sequencing and genome assembly. Genom Proteom Bioinformatics 14:265–279.
13. Mayo B, Rachid CT, Alegria Á, Leite AM, Peixoto RS, Delgado S. 2014. Impact of next generation sequencing techniques in food microbiology. Curr Genomics 15:293–309.
14. Gilchrist CA, Turner SD, Riley MF, Petri WA, Hewlett EL. 2015. Whole-genome sequencing in outbreak analysis. Clin Microbiol Rev 28:541–563.
15. Roossinck MJ. 2016. Deep sequencing for discovery and evolutionary analysis of plant viruses. Virus Res, in press. doi:10.1016/j.virusres.2016.11.019.