



A Small-Group Activity Introducing the Use and Interpretation of BLAST †

Peter D. Newell¹, Ashwana D. Fricker², Constance Armanda Roco², Pete Chandrangsu²,
and Susan M. Merkel^{2*}

Departments of ¹Entomology, and ²Microbiology, Cornell University, Ithaca, NY 14853

As biological sequence data are generated at an ever increasing rate, the role of bioinformatics in biological research also grows. Students must be trained to complete and interpret bioinformatic searches to enable them to effectively utilize the trove of sequence data available. A key bioinformatic tool for sequence comparison and genome database searching is BLAST (Basic Local Alignment Search Tool). BLAST identifies sequences in a database that are similar to the entered query sequence, and ranks them based on the length and quality of the alignment. Our goal was to introduce sophomore and junior level undergraduate students to the basic functions and uses of BLAST with a small group activity lasting a single class period. The activity provides students an opportunity to perform a BLAST search, interpret the data output, and use the data to make inferences about bacterial cell envelope structure. The activity consists of two parts. Part 1 is a handout to be completed prior to class, complete with video tutorial, that reviews cell envelope structure, introduces key terms, and allows students to familiarize themselves with the mechanics of a BLAST search. Part 2 consists of a hands-on, web-based small group activity to be completed during the class period. Evaluation of the activity through student performance assessments suggests that students who complete the activity can better interpret the BLAST output parameters % query coverage and % max identity. While the topic of the activity is bacterial cell wall structure, it could be adapted to address other biological concepts.

INTRODUCTION

Over the last decade, a number of national reports have called for reform in teaching college-level biology classes (1, 7, 8). One strategy for training future biologists is to develop critical-thinking and problem-solving skills by providing authentic educational activities that help them learn how to implement and interpret current research in biology (2). Bioinformatic tools that allow scientists to explore genome sequence data have become a cornerstone of current biological research and as such should be included in any modern biology curriculum (3, 5, 9). The Basic Local Alignment Search Tool (BLAST; <http://blast.ncbi.nlm.nih.gov/>) is one of the most commonly used tools for comparing sequence information and retrieving sequences from databases and is thus an excellent starting point for teaching bioinformatics (4). BLAST takes a query sequence and searches a given database of sequences for significant matches, generating local alignments that vary in length,

and providing descriptive parameters as well as statistical evaluation of any matches. A few microbiology educational activities that employ BLAST have been published, and range from a basic introduction (e.g., ref. 6) to more involved multi-day activities (e.g., ref. 5). The activity described in this study introduces students to the use of BLAST and uses BLAST to show that differences between Gram-positive and Gram-negative cell envelope structure are reflected in the gene content of bacterial genomes.

We have given students in our General Microbiology courses bacterial 16S ribosomal RNA (rRNA) sequences and asked them to use BLAST to determine the bacterial identity. From this experience, we observed that while students could easily go through the steps to do a BLAST search, most could not interpret the output data and therefore could not draw appropriate conclusions from those searches. The students needed to develop a conception of what constitutes a 'good' BLAST hit, or matching sequence. To address this need, we developed the classroom activity presented here, emphasizing the interpretation of BLAST output parameters.

Our activity was designed for microbiology and biology majors in a general microbiology class, but would be equally appropriate in a biotechnology class. Students should have some understanding of the structure of Gram-positive and

*Corresponding author. Mailing address: Department of Microbiology, 111 Wing Hall, Cornell University, Ithaca, NY 14853. Phone: 607-254-2767. Fax: 607-255-3904. E-mail: smm3@cornell.edu.

†Supplemental materials available at <http://jmbe.asm.org>

Gram-negative cell envelopes, as well as DNA structure and how it relates to protein synthesis in bacteria. Students need computer access to a BLAST site. We used the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov/).

The activity has two parts: Part 1 (Appendix 1) is a handout that was completed by students independently prior to class, while Part 2 (Appendix 3) was handed out and completed during a 50-minute class period. In Part 1 students were given diagrams of Gram-positive and Gram-negative cell envelopes and required to label key components. In addition, students were given an amino acid sequence and asked to follow a tutorial on how to use the NCBI BLAST site to find the predicted function of the protein (in this case, a porin). They answered questions about the BLAST output, including E-value, % query cover, and % max identity. We estimate that it took 30 to 45 minutes to complete.

Students then came to class with laptop computers and had 50 minutes to complete Part 2. In Part 2, students were given a different amino acid sequence (Appendix 5) and again used NCBI BLAST to predict the protein's function (a histidine porin). Next, they restricted the BLAST search to certain bacterial genomes with the same query sequence to determine if other bacteria carry the gene. The restricted searches were designed to highlight the presence of porins in Gram-negative bacteria and their absence in Gram-positive bacteria. Students answered a number of questions about the NCBI BLAST search output. They performed the searches and discussed their answers in groups of three to four students, but wrote individual answers to questions, which were handed in for a grade.

While we had over 180 students divided into multiple sections, each with a facilitator, this activity could be used in smaller classes with one teacher as a facilitator for small group work. Students could work together on Part 1 in class, with Part 2 being given as an out-of-class assignment. Alternatively, both sections could be done in class or as homework.

Learning objectives

Students who complete this activity should be better able to:

1. Explain the basic function of BLAST
2. Predict the function of a protein sequence using BLAST
3. Evaluate sequence similarity based on BLAST outputs: E-value, % query cover, and % max identity
4. Determine if a gene product is present in a specific organism using BLAST

Even though the topic of the activity is cell envelope structure, it is not intended to provide this information comprehensively. Rather, we hope it enhances student understanding of the topic by showing that cell structure differences are reflected in the gene content of bacterial genomes.

PROCEDURE

Materials

Students need the handout for Part 1 (1 per student, Appendix 1) and access to the BLAST tutorial, which we provided as both a Panopto webinar (https://www.youtube.com/watch?v=x_dAyY5-VNc) and a PDF file (Appendix 7). For Part 2, students need at least one laptop computer for a group of three to four students, a copy of the handout (1 per student, given out in class, Appendix 3) and access to an electronic version of the amino acid sequence (Appendix 5). Instructors need the answer keys to both parts (Appendices 2 and 4). Students need Internet access to a BLAST site for all parts of the activity.

Student instructions

In our classes, we instructed students to download a PDF of Part 1 (Appendix 1) before class and answer the questions therein. The instructions for Part 1 were as follows: *This week's activity introduces BLAST (Basic Local Alignment Search Tool), ... [and will] highlight important differences between the cell envelopes of Gram-positive and Gram-negative bacteria. Please note: you need to complete Part 1 to obtain information for the in-class activity (Part 2). Completion of Part 1 will be checked at the beginning of class. Please bring laptops to class!*

One week later, students in the general microbiology lecture came to their assigned discussion section. After reviewing the answers to Part 1, students were given Part 2 (Appendix 3) with this introduction: *Pseudomonas aeruginosa is a pathogenic bacterium that can infect a wide variety of animals. P. aeruginosa is particularly devastating to patients suffering from cystic fibrosis (CF), a genetic disease causing the buildup of thick mucus in the lungs. The dysfunctional lungs of CF patients are chronically infected with P. aeruginosa, which is well adapted to survive in this habitat, in part because it can efficiently utilize amino acids for carbon and energy. Working in a clinical microbiology lab, you isolate a new strain of P. aeruginosa that thrives especially well in CF patients. Comparing its protein expression patterns to previous isolates, you find that one protein is highly expressed in your isolate relative to other P. aeruginosa strains.*

Students downloaded an electronic copy of an amino acid sequence from the course BlackBoard site (Appendix 5), and were instructed as follows: *To find the function of this protein, you perform a BLAST search of its amino acid sequence. To do this, go to the National Center for Biotechnology Information site (<http://blast.ncbi.nlm.nih.gov/>) to do a BLAST-P search to determine a probable identity of this protein following the same procedures as in Part 1 and in the BLAST tutorial. They then answered a series of questions asking them to define and interpret the BLAST output results: E value, % query coverage, and % maximum identity.*

Students were then instructed to search other genomes for similar sequences: *You scroll down the table giving*

descriptions of the BLAST hits and notice that similar proteins also occur in *Pseudomonas* species other than *P. aeruginosa*. You are curious about how widespread this protein may be, so you decide to search the genomes of two well studied bacteria for similar sequences: *Bacillus subtilis* (Gram-positive) and *Escherichia coli* (Gram-negative). After doing this, they answered a series of questions asking them to interpret the BLAST output results within the context of their knowledge of Gram-negative and Gram-positive cell structure.

Faculty instructions

Out-of-class Assignment, Part 1. Completion of Part 1 and viewing the BLAST tutorials prepared students for the in-class activity. While we did not grade student answers to Part 1 per se, we chose to incentivize completion of this part of the activity to ensure student preparedness for Part 2. At the beginning of the class period, the instructor checked whether or not each student's Part 1 was completed. A completed Part 1 counted as two points (out of a total of 10) towards the grade for this activity. We hoped from this activity that students understood the main definitions of E-value, % maximum identity, and % query coverage. E-value represents how well the query sequence matches the database sequence, taking into account both the number of matching residues and the total length of the alignment. The lower the E-value, or the closer it is to zero, the better the match is. % Maximum identity is the percentage of residues that match up in the alignment. % Query coverage is the percentage of the query sequence length that is included in the alignment. When running a BLAST search, often the sequences returned will align with only part of the queried sequence; therefore % query coverage has a significant impact on the E value—the greater the query coverage, the lower the E-value and the better the match.

Small group discussion, Part 1. For the discussion section in our General Microbiology Lecture (~180 students), students met in smaller groups of about 15 led by facilitators (in our case, graduate student teaching assistants). The facilitator began the class period by reviewing the answers to Part 1 (Appendix 2) in an informal discussion format, answering any questions that students had. During this review, facilitators explained and emphasized the meaning of BLAST output terms: E-value, % maximum identity, and % query coverage. Part 1, Question 7 was particularly useful in demonstrating the differences between % maximum identity, and % query coverage.

Small group discussion, Part 2. After reviewing Part 1 (~10 min into the class period), students broke up into groups of three to four to complete Part 2 (Appendix 3). Students were encouraged to work together and discuss the answers to each question with their group members, but each student wrote his/her own answer to each question.

To complete Part 2, students were given an amino acid sequence and used the BLAST-P search hosted by NCBI to identify matches. (See Appendix 7 for step-by-step instructions.) The first set of questions (Appendix 4, questions 1–6) asked students to report and interpret information from the top BLAST hit obtained by searching the entire non-redundant protein sequence database. Next, the students used the same query sequence to search two subsets of sequences individually by using the “Choose Search Set-Organism” drop-down menu. The two subsets were: “*Escherichia coli* genomes [taxid:562]” and “*Bacillus subtilis* genome [taxid:1423].” Once the query sequence had been aligned to either the *Escherichia coli* genome or the *Bacillus subtilis* genome, students entered the following information into a table (Table 1): protein name, % query coverage, E-value, and % maximum identity. Facilitators checked the table to see that students' answers were correct, as failure to obtain the correct values for the table would likely have lead students to answer the remaining questions incorrectly.

The next set of questions (Table 1 and Appendix 4, questions 8–9) compared search result values obtained by searching different genomes. The activity used a porin sequence from *P. aeruginosa* as query, and thus guaranteed a better match to sequences in *E. coli* than to any sequence in *B. subtilis* because the former is Gram-negative and contains porins while the latter is Gram-positive and does not. Please note that the NCBI databases are subject to change due to the addition of new sequences, and thus the exact values for the top BLAST hits may change. This is unlikely to impact the general trends and comparisons that students are asked to make in questions 8–9. Interpreting the output, particularly for question 9 (*How can Bacillus subtilis have a higher % max identity than E. coli but a lower % query coverage? Explain this phenomenon*), required an understanding of the differences between % query coverage and % maximum identity. It is important to look at % query coverage first. For a good match, it precludes having a high % maximum identity. Sequences from unrelated proteins could match across a small segment of amino acids, which would result in a high % maximum identity but a low % query coverage (which was the case for the porin sequence and *B. subtilis*). Therefore, to be confident of a good match, a high % query coverage is needed.

Sample data

Student answers and misconceptions are given in Appendix 2 and Appendix 4.

Suggestions for determining student learning

At the start of the semester, we typically give students pretest questions to determine their understanding of a number of key concepts in microbiology. In this case, students were given pretest questions (Appendix 6) to assess their prior knowledge of what BLAST does and what the

output means. Second, we collected the answers to the student activity, Parts 1 and 2. Third, we gave students posttest questions. Fourth, students filled out an evaluation (Table 2) to assess whether they thought they had achieved all four learning objectives of this activity. Answer keys for the pretest, Parts 1 and 2 of the student activity, and the posttest are provided in the Supplemental Materials.

DISCUSSION

Field testing

We teach separate Microbiology lecture and lab courses at Cornell University. Both are introductory-level courses with prerequisites of two semesters each of biology and chemistry. We have a range of students from sophomores to seniors, some of whom have already taken biochemistry and genetics. Assessment data presented in this paper (pre- and posttest results) are from students in the spring 2013 General Microbiology Laboratory class. This group consisted of 56 students who were concurrently taking both General Microbiology Lecture and Lab classes (these students did the BLAST activity; the treatment group) and 27 students who had taken the lecture in a previous semester (these students did not do the activity; the control group). We compared the two groups to measure the effectiveness of the BLAST activity. Students in the lecture class did not receive any other instruction regarding how to do a BLAST search or how to interpret the output data.

During the first week of classes (Jan. 2013) we gave all students (control and treatment groups) a pretest to assess their prior knowledge for the course. Two of the questions related to BLAST (Appendix 6). In the third and fourth weeks of classes (Feb. 2013), the treatment group completed Part 1 and Part 2 of the activity. Forty-five student papers were chosen at random by a staff member not involved in the study, and copied for assessment. Approximately three months later (May 2013), all students (control and treatment groups) were given a posttest (Appendix 6). Three posttest questions were related to BLAST and interpretation of BLAST outputs. Also during May 2013, the treatment group (as well as other students in the General Microbiology Lecture course) were given evaluation questions, three of which pertained to the BLAST activity (Table 2).

Evidence of student learning

Both treatment and control groups had some exposure to BLAST during the semester: as part of a laboratory exercise they were given a sequence from a 16S rRNA gene and were asked to use BLAST to identify the bacterium. However, for the laboratory exercise, they were given no instruction on how to interpret the BLAST output (i.e., E-value, % query coverage, % max identity). Thus, only the treatment group had instruction on the meaning and interpretation of BLAST output parameters via the activity presented here.

TABLE 1.
Part 2, question 7^a, with answers.

| Organism | Protein Name | % Query Coverage | E-value | % Max Identity |
|----------------------|----------------------|------------------|---------|----------------|
| <i>P. aeruginosa</i> | Histidine porin OpdC | 97 | 0 | 96 |
| <i>B. subtilis</i> | Hypothetical protein | 7 | 8.8 | 34 |
| <i>E. coli</i> | Outer membrane porin | 93 | 6 e-4 | 22 |

^aQuestion 7 reads "Fill in the table with your results for the top BLAST hits for the sequence from your isolate."

TABLE 2.
Student evaluation of learning objectives.

| Question | Learning Objective | very well ^a | OK | poorly | not at all |
|--|---|------------------------|-----|--------|------------|
| "The first small group activity focused on using sequence information and bioinformatics to determine the function of a protein and to look for it in other bacteria. How well did the activity help you meet each goal (listed below)?" | 2. Determine the predicted function of a protein sequence using BLAST | 49% | 47% | 4% | 0% |
| | 3. Evaluate sequence similarity based on BLAST outputs: E-values, % query cover, and % max identity | 36% | 51% | 12% | 1% |
| | 4. Determine if a gene product is present in a specific organism using BLAST | 47% | 47% | 5% | 1% |

^an = 165 students in the General Microbiology Lecture who responded to the survey.

Learning objective 1: Students will be able to explain the basic function of BLAST. To assess the effectiveness of the activity for achieving learning objective 1, we analyzed the number of correct responses to pre- and posttest questions (question 1 on each, Appendix 6), which asked students to identify for what BLAST is used. For the class as a whole, the number of students who could identify the function of BLAST on the posttest (86%) was higher than students who could define BLAST on the pretest (37%, Table 3) though differences between the pretest (short answer) and posttest (multiple choice) questions make a direct comparison difficult. An ANOVA analysis looking at the interaction between doing the activity (or not) with pre- and posttest scores revealed that doing the BLAST activity had a significant impact ($p = 0.0275$) on student scores. That is, students who did the BLAST activity were more likely to be able to define BLAST (90%) on the posttest than those who did not do the activity (70%).

Learning objective 2: Students will be able to predict the function of a protein sequence using BLAST. We were not able to assess this objective using pre- and posttest questions because students need access to a computer to do this. However, when students were asked to evaluate how well this activity helped them learn how to predict the function of a protein sequence using BLAST, 96% students said “very well” or “OK” (Table 2).

Learning objective 3: Students will be able to evaluate sequence similarity based on BLAST outputs: E-value, % query cover, and % max identity. And Learning objective 4: Students will be able to determine if a gene product is present in a specific organism using BLAST. We assessed these objectives by a variety of measurements. First, we analyzed the number of correct responses to the multiple-choice pre- and posttest question 2 (*Which E-value would indicate a very good match for a protein sequence BLAST?*). The number of students who answered correctly on the posttest (30%) was significantly higher ($p = 0.0002$) than students who answered correctly on the pretest (4%; Table 3). In addition, an ANOVA analysis looking at the interaction between doing the activity (or not), with pre- and posttest scores revealed that doing the BLAST activity had a significant impact on student scores.

That is, students who did the BLAST activity were significantly more likely ($p = 0.014$) to choose the best E-value on the posttest (42%) than those who did not do the activity (7% Table 3).

Students answered the questions in Part 2 in groups, with a facilitator, and our goal was to have all students master the concepts during the class period. Therefore, the number of students answering these questions correctly was not an ideal tool for assessing individual student learning. It was, however, a good indicator of student misconceptions about activity content, particularly in the understanding of BLAST output values (question 9, Appendices 3 and 4). To answer these questions students had to interpret a table of BLAST values and explain why *Bacillus subtilis* had a higher % max identity than *E. coli* but a lower % query coverage. Sixty-nine percent of students correctly explained the difference between % max identity and % query coverage, 20% of students got partial credit for having some understanding and 11% received no credit. Some students were confused about what was being compared. That is, they did not seem to understand that the whole genome was being searched or that the % query coverage represented how much of the query was found in the genome. This concept seemed to be more difficult to comprehend, compared to % max identity. These challenges were reflected in the student evaluation data. While around half of the students felt the activity did “very well” to help them meet learning objectives 2 and 4, only 36% felt the same about meeting learning objective 3 (Table 2).

Because this seemed to be a difficult concept for students, we added a question about this to the posttest. We asked students to interpret similar data and explain their answers (Appendix 6). Using a paired *t*-test, we found that students who did the activity were significantly more likely ($p = 0.009$) to answer the question correctly (67%) compared to students who did not do the activity (26%). Of those in the treatment group who answered correctly, 77% were able to explain that the sequence in *P. griffinnia* aligned with a higher percentage of the queried sequence, correctly differentiating between % query coverage and % max identity. Of those in the control group who answered correctly, about half could not explain their answer. Of those who answered incorrectly, most answered *Y. gabbagabbaea*, reasoning that % max identity was most important, regardless of % query coverage.

TABLE 3.
Percentage (and number/total) of students providing correct answers on pre- and posttests.

| | | Treatment | Control | Total |
|------------|----------|-------------|-------------|-------------|
| Question 1 | Pretest | 29% (16/56) | 59% (13/22) | 37% (29/74) |
| | Posttest | 90% (47/52) | 70% (19/27) | 83% (66/79) |
| Question 2 | Pretest | 5% (3/56) | 0% (0/22) | 4% (3/74) |
| | Posttest | 42% (22/52) | 7% (2/27) | 30% (24/79) |

Modifications

Student assessment suggested that there is still room for improvement in students' understanding of BLAST terms after completing the activity. Common misconceptions included incomplete comprehension of the terms % query coverage and % max identity. Instructors should emphasize the portions of Part 1 (e.g., question 7) that illustrate these terms when conducting the in-class review prior to handing out Part 2. In addition, more questions could be added to Part 1 to test students' comprehension of the definitions of BLAST terms. Also, walking students through the process by which the BLAST algorithm functions might increase the proportion of students answering correctly. In our case we decided this level of detail was beyond the scope of the activity.

This activity has a relatively narrow scope because it was designed to fit within a single class period. Given more instruction time, it could be expanded. A more detailed discussion of the uses and limitations of BLAST would be beneficial. Discussion of uses could include additional research rationales or reasons for using BLASTp on NCBI (as this activity does) versus other algorithms and/or databases that analyze sequence data. For example, explaining the differences between BLASTp and BLASTn (nucleotide BLAST). Important limitations to discuss include the reliability of gene annotations, and the fact that similarity to a sequence in the database may predict but cannot prove a similar function for your query. A benefit of this activity is that the sequences given to students could be easily exchanged for others, making it adaptable for highlighting other class material. For example, while our activity emphasized the absence of outer membrane proteins in Gram-positive cell envelopes, a similar activity could focus on differences between transcriptional machinery in Archaea and Bacteria.

Suggested resources

Madden, T. The BLAST sequence analysis tool. In J. McEntyre and J. Ostell (ed). The NCBI Handbook [Internet], Bethesda, MD: National Center for Biotechnology Information (US); 2002. Oct 9, [Updated 2003 Aug 13]. <http://www.ncbi.nlm.nih.gov/books/NBK21097/> [PMC free article].

Madigan, M. T., J. M. Martinko, P. V. Dunlap, and D. P. Clark. 2012. Brock biology of microorganisms, 13th ed. Pearson Benjamin Cummings. Sections 3.6–3.7, 6.12–6.17, 16.17.

NCBI, Basic Local Alignment Search Tool. On-line help section: http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs

NCBI. YouTube BLAST tutorials: <http://www.youtube.com/playlist?list=PLH-TjWpFfWrtjzMCIVUe-YbrlleFQIKMq> (Oct. 2012).

Slonczewski, J. L., and J. W. Foster. 2011. Microbiology, an evolving science, 2nd ed. WW Norton & Company, Inc. Sections 3.4, 8.1–8.3, 8.7.

SUPPLEMENTAL MATERIALS

- Appendix 1: Student handout part 1
- Appendix 2: Instructor handout part 1, answers and misconceptions
- Appendix 3: Student handout part 2
- Appendix 4: Instructor handout part 2, answers and misconceptions
- Appendix 5: Part 2 sequence
- Appendix 6: Pre- and posttest questions, answers and misconceptions
- Appendix 7: BLAST tutorial

ACKNOWLEDGMENTS

This activity was developed as part of a graduate course in Teaching Microbiology (BioMi 7960) at Cornell University. Thanks to Dr. James Shapleigh for sharing his experience in teaching bioinformatics to undergraduate students and for help with the Panapto tutorial. Julie Brown, Erin Eggleston, Olga Lastovetsky, and David Sannino contributed suggestions for the improvement and evaluation of this activity. Dr. Joseph Yavitt helped with our statistical analysis. The authors declare that there are no conflicts of interest.

REFERENCES

1. **American Association for the Advancement of Science.** 2010. Vision and change in undergraduate biology education: a call to action. A report of the American Association for the Advancement of Science. [Online.] <http://visionandchange.org/>.
2. **DeHaan, R. L.** 2005. The impending revolution in undergraduate science education. *J. Sci. Educ. Teach.* **14**:253–269.
3. **Ditty J. L., et al.** 2010. Incorporating genomics and bioinformatics across the life sciences curriculum. *PLoS Biol.* **8**:e1000448. doi: 10.1371/journal.pbio.1000448.
4. **Kerfeld, C. A., and K. M. Scott.** Using BLAST to Teach “E-value-tionary” Concepts. *PLoS Biol.* **9**:e1001014.
5. **Klein, J. R., and T. Gulsvig.** 2012. Using bioinformatics to develop and test hypotheses: *E. coli*-specific virulence determinants. *J. Microbiol. Biol. Educ.* **13**:161–169.
6. **May, J. B.** 2013. Engaging students in a bioinformatics activity to introduce gene structure and function. *J. Microbiol. Biol. Educ.* **14**:107–109.
7. **National Research Council.** 2003. Bio2010: Transforming Undergraduate Education for Future Research Biologists. The National Academies Press, Washington, DC. [Online.] <http://www.nap.edu/openbook.php?isbn=0309085357>.
8. **National Research Council.** 2009. A new biology for the 21st century: ensuring the United States leads the coming biology revolution. The National Academies Press, Washington, DC. [Online.] http://www.nap.edu/catalog.php?record_id=12764.
9. **Ranganathan S.** 2005. Bioinformatics education—Perspectives and challenges. *PLoS Comput. Biol.* **1**:e52. doi: 10.1371/journal.pcbi.0010052.