Bioinformatics tools for marine biotechnology: a practical tutorial with a metagenomic approach

Ludovica Liguori1,2, Maria Monticelli3, Mariateresa Allocca1,2, Maria Vittoria Cubellis3,4* and Bruno Hay Mele3,5

From 13th Bioinformatics and Computational Biology Conference - BBCC 2018
Naples, Italy. 19-21 November 2018

Abstract

Background: Bioinformatics has pervaded all fields of biology and has become an indispensable tool for almost all research projects. Although teaching bioinformatics has been incorporated in all traditional life science curricula, practical hands-on experiences in tight combination with wet-lab experiments are needed to motivate students.

Results: We present a tutorial that starts from a practical problem: finding novel enzymes from marine environments. First, we introduce the idea of metagenomics, a recent approach that extends biotechnology to non-culturable microbes. We presuppose that a probe for the screening of metagenomic cosmid library is needed. The students start from the chemical structure of the substrate that should be acted on by the novel enzyme and end with the sequence of the probe. To attain their goal, they discover databases such as BRENDA and programs such as BLAST and Clustal Omega.

Students’ answers to a satisfaction questionnaire show that a multistep tutorial integrated into a research wet-lab project is preferable to conventional lectures illustrating bioinformatics tools.

Conclusion: Experimental biologists can better operate basic bioinformatics if a problem-solving approach is chosen.

Keywords: Marine biotechnology, Graduate education, Computer-based learning

Background

At present any biologists should capitalize on the resources, data and programs, that are available online to make their experimental plans more efficient and cost-effective [1, 2]. For this reason, it is desirable to train students using a problem-solving approach that integrates in silico work into a multidisciplinary experimental project(a
few examples [3–13]. We provide a tutorial that was administered to students of the courses of marine and environmental biology.

Biodegradation of environmental pollutants by marine prokaryotic enzymes provides the frame into which the bioinformatics tutorial is inserted. In particular, the experimental project proposed to the students, aims at finding an enzyme that is active on a scaffold commonly found among pollutants and synthetic compounds, a so-called “privileged scaffold” [14]. In this exercise, the scaffold taken into consideration is indole, an N-heterocyclic aromatic pollutant released in the aquatic environment through the industrial wastewater [15].

The students will look for an enzyme with broad specificity that is able to degrade indole and more in general aromatic compounds [16]. As it is widely known, only a small fraction of environmental microbes grow under conventional laboratory conditions [17–19]. For this reason, several authors suggested that metagenomes might be a big reservoir of novel enzymes for applications in biocatalysis, biofuels, and bioremediation (for reviews on this subject [20–23]).

The project consists of two parts: a laboratory part, that we won’t discuss in detail here, to construct a DNA clones library from the metagenomic marine sample, and a bioinformatics part that is the object of this paper.

We presuppose that DNA has been extracted from prokaryotes present in a seawater sample and that a metagenomic library has been constructed in cosmids. In principle, the screening could be carried out testing the enzymatic activity of interest directly. However, assays on plates can be cumbersome and we propose an alternative strategy that takes advantage of the simplicity of DNA colony hybridization. Only after the clone has been identified by a DNA labelled probe and isolated, the activity will be confirmed in the recombinant *E.coli* extracts.

The tutorial focuses on the design of a suitable probe for the screening. It combines information deriving from a highly annotated enzyme database such as BRENDA [24] with data on uncharacterized open reading frames (ORF) deriving from large scale metagenomic sequencing projects.

BRENDA is freely available for academic users and educational purposes [25]. Enzymes are classified according to the catalyzed reactions and well-characterized proteins from different organisms can be found in each class. BRENDA [24] can be searched with a structure-based query as well as with a text-based query. It provides a substructure search algorithm that is very useful to use a scaffold as a query. This choice is very convenient because chemical structures identify molecules uniquely whereas names are not unique. For instance, indole is also known as 1-Benzazole,1H-Benz[H]pyrrole,1H-Indol,1H-Indole, nomenclature can be even more confusing for other molecules. BRENDA can be searched for several specific information at the same time with “Advanced options”. The tutorial shows how a bacterial enzyme that is active on a scaffold commonly found among pollutants and synthetic compounds can be found. We imagine that the sequence cannot be used as such for colony hybridization of the cosmid library under stringent conditions. Hence we propose to use the enzyme found in BRENDA [24] as a query sequence to look for homologous proteins among uncharacterized ORF from marine metagenomes. This is possible because BLASTp [26] consents to search protein sequences from large environmental sequencing projects such as the Malaspina expedition [27], Global Ocean Sampling (GOS) campaign [28] and Tara Oceans expedition [29].
The tutorial was administered to 23 students with no previous knowledge of bioinformatics who filled a satisfaction test at the end of the exercise.

**Methods**

**Aims**

This tutorial is intended for laboratory biologists with no previous knowledge of bioinformatics. We envisaged a multistep bioinformatics protocol that is integrated into a project of marine biotechnology. The protocol focuses on the use of four bioinformatics web applications (BLASTp [26], Clustal Omega, Cons, Reverse Translate [30]) and two biochemical databases (UniProt [31] and BRENDA [24]). The learning goals are summarized in Table 1.

**Experiment overview**

We presuppose that the students are engaged in a project for the identification of a novel bacterial enzyme from a seawater metagenomic sample. For this purpose, we assume that a library has been constructed in cosmids and must be screened. In this tutorial, we present a bioinformatics protocol to design a probe to isolate the clone of interest (Fig. 1).

In the first step, the students will search BRENDA [24], the enzyme database [25], to find a prokaryotic enzyme that is able to use a given class of pollutants as substrates. BRENDA [24] consents looking among enzyme ligands by chemical similarity and by substructures. In the tutorial, indole is chosen as an example because it represents a privileged scaffold, i.e. its chemically active structure is common to many natural and synthetic compounds with the ability to bind different targets [14]. Unfortunately, it is not (yet) possible to carry out an advanced search drawing a substructure and for this reason, to get an enzyme class that is active on a given substrate, identified by its chemical structure, AND is expressed in bacteria, it is necessary to follow a two-phase, apparently redundant, protocol. In the first phase, the structure is used as the input to get the exact name (i.e. the one used by BRENDA that does not necessarily coincides with the IUPAC name) of the substrate of interest. In the second phase, the exact name and the class of organisms are used with advanced options. Several classes of enzymes are active on molecules that resemble indole. Naphthalene 1,2-dioxygenase has broad specificity and is involved in the degradation of many aromatic compounds [32]. By clicking on the E.C. number the students will get much information that is subdivided into different sections. The most interesting ones for the case under study are those

| Software/Database Employed | Educational Goal                                                                 |
|----------------------------|----------------------------------------------------------------------------------|
| BRENDA                    | Obtain information for a given enzyme class.                                     |
| Uniprot                   | Obtain the sequence and the annotations for a given protein.                     |
| BLASTp                     | Search sequence databases using a protein query.                                 |
| Clustal Omega              | Build phylogenetic trees and multiple alignments.                                 |
| Cons                       | Derive a Consensus sequence from a multiple alignment.                            |
| Reverse Translate          | Derive a nucleotide sequence from an amino acid sequence.                         |
concerning enzyme-ligands interactions where it is possible to learn which are all the possible substrates of this class on enzymes, those concerning organism related information and those related to enzyme structures, where the links to UniProt [https://www.uniprot.org/] [31] are found. The students will not find any Naphthalene 1,2-dioxygenases from a prokaryotic marine organism. At present (March 2019) the only well-characterized prokaryotic enzyme is from *Pseudomonas putida* [33].

In the second step, the sequence of Naphthalene 1,2-dioxygenases from *Pseudomonas putida* will be obtained from UniProt [https://www.uniprot.org/] [31]. *Pseudomonas putida* is evolutionarily distant from the marine prokaryotes that are present in the
sample of marine water used to construct the library. The students will look for homologous sequences from marine prokaryotes. They will take advantage of a large number of uncharacterized coding sequences obtained by massive genomic and metagenomic sequence projects. To carry out such an analysis, they will use BLASTp [https://blast.ncbi.nlm.nih.gov/Blast.cgi? PROGRAM = blastp&PAGE_TYPE = BlastSearch&LINK_LOC = blasthome] using Naphthalene 1,2-dioxygenases from *Pseudomonas putida* as the query and limiting the search to the metagenomic proteins deposited in databases (env_nr).

In the third step, a multiple sequence alignment will be carried out using Clustal Omega [https://www.ebi.ac.uk/Tools/msa/clustalo/] [34]. The students will identify conserved regions among the sequences homologous Naphthalene 1,2-dioxygenases...
from *Pseudomonas putida* and will derive a consensus using Cons [http://www.bioinformatics.nl/cgi-bin/emboss/cons] [35–37].

In the fourth step, the retrotranslation of the aminoacidic consensus sequence will be carried out using [https://www.bioinformatics.org/sms2/rev_trans.html] [30]. The students will learn that retrotranslation does not provide a unique DNA sequence unless the codon most frequently used for each amino acid in prokaryotes is chosen.

The output of this tutorial is a sequence of the probe for the screening by colony hybridization.
Requirements

All software used is free with a user-friendly interface available. The only requirement is a computer with an Internet connection.

The used software and databases are:

- BRENDA, https://www.brenda-enzymes.org/index.php
- BLASTP, https://blast.ncbi.nlm.nih.gov/Blast.cgi
- UniProt, http://www.uniprot.org/
- Clustal Omega, https://www.ebi.ac.uk/Tools/msa/clustalo/
- Cons, http://www.bioinformatics.nl/emboss-explorer/
- Reverse translate, http://www.bioinformatics.org/sms2/rev_trans.html

Detailed protocol

**Step 1. BRENDA: searching the enzyme database to find a microbial enzyme able to degrade indole**

Access BRENDA (The Comprehensive Enzyme Information System) at https://www.brenda-enzymes.org/index.php.

Draw the molecule of interest, a bicyclic structure, consisting of a six-membered benzene ring fused to a five-membered pyrrole ring, clicking on “Ligand Structure Search” in the BRENDA homepage (Fig. 2) to obtain the exact name of the compound. A “substructure search” with a maximal search time of 120 s restricted to “Substrates” must be selected before running the search (Fig. 3).

You obtain the exact names of several molecules containing the structure you drew, you choose “indole”. Go back to BRENDA homepage and run an advanced search (Fig. 4) filling in the kingdom (bacteria) and type (substrate) boxes and using the exact name of the molecule (indole) (Fig. 5).

BRENDA shows a list of all the enzymes in the database that use indole as a substrate. The enzyme Naphthalene 1,2-dioxygenase from *Pseudomonas* has been chosen from the results for this training. It is present in different bacterial species and has a large substrate specificity. Please click on the EC number (Fig. 6) to obtain all the biochemical data about the enzyme.

![Fig. 6](image)

Classes of enzymes acting on indole in BRENDA (inset). Naphthalene 1,2-dioxygenase can be selected among the enzyme classes acting on indole.

![Fig. 7](image)

Specific bacterial enzyme in BRENDA. The link to UniProt can be selected for Naphthalene 1,2-dioxygenases from specific proteins.
In the section “enzyme structures”, look for “AA sequences” where the links to UniProt [https://www.uniprot.org/] are found (Fig. 7). *Pseudomonas putida* has been chosen as an organism; click on the UniProt code to open the link to the proteins sequences database.

On the UniProt sheet choose Sequence and download the amino acid sequence in FASTA format (Fig. 8).

**Step 2. Protein BLAST (BLASTp) to carry out a local alignment to compare an amino acid sequence to the metagenomic proteins deposited into a database**

Open a new Internet page with BLAST at https://blast.ncbi.nlm.nih.gov/Blast.cgi and click on “Protein BLAST” to align amino acid sequences.

Paste the sequence downloaded from UniProt into the “Enter Query Sequence” field, choosing “Metagenomic proteins” as database. In the section “Algorithm parameters”, select 500 as “Max target sequences” and run BLAST (Fig. 9).

Select and download all “marine metagenome” sequences in FASTA format (Fig. 10).
Step 3. Clustal omega to perform multiple sequence alignment, helpful to predict relations and similarity among sequences

Open a new Internet page with Clustal Omega tool at https://www.ebi.ac.uk/Tools/msa/clustalo/ to perform a multiple alignment.

Paste your sequences in the blank using Pearson/FASTA as output format, leave the other parameters as default and submit the work (Fig. 11).

Step 4. Cons: an EMBOSS explorer tool to create a consensus sequence from a multiple alignment

Connect to EMBOSS Explorer [35–37] at http://www.bioinformatics.nl/emboss-explorer/ and on the left side of the web sheet click on “cons”. Copy the obtained multiple alignment in FASTA format and paste it in the blank (Fig. 12). An example of consensus sequence output is shown in Fig. 13.

Copy a region (selecting about 30 amino acid residues to obtain a ~ 100 bp DNA probe avoiding regions containing too many “Xs”).

Step 5: translate an amino acid sequence into the most probable nucleotide sequence

Search for the web tool Reverse Translate (http://www.bioinformatics.org/sms2/rev_trans.html) to retrotranslate an amino acid sequence into the most probable nucleotide sequence using the default codon usage from E.coli.

Paste the copied sequence in the blank and submit the work (Fig. 14). In Fig. 15, an example of the obtained nucleotide probe is shown.
Results

The tutorial was administered to 23 graduate students who had a bachelor’s degree in biology or natural science with a basic knowledge of biochemistry and molecular biology and no previous knowledge of bioinformatics. This tutorial can be administered to undergraduate students too if they have solid bases of biochemistry and molecular biology. Before administering the tutorial, two lectures were given to introduce the programs and the
databases reported in Table 1. All the students without the help of the supervisor concluded the tutorial successfully within one hour. After the practice, students have undergone a satisfaction questionnaire of 7 questions (Additional file 1) to evaluate the general interest and the usefulness of the multistep bioinformatics protocol. All students were satisfied although a minority admitted that they would not be able to apply the same tools to another biological project (30%). Questions indicating the student satisfaction degree are gathered in Fig. 16 and in the supplementary material. Through a computer-based approach, students have been able to search into protein and enzyme databases, performing local and multiple sequence alignments obtaining a consensus sequence and retrofetran- 
slate an amino acid sequence to obtain a DNA sequence.

We are aware of the fact that other types of questionnaires could be administered, for instance, students could be asked to explain the steps that they have performed and why.

An even more convincing proof of the usefulness of the tutorial would have been letting the students prepare the cosmid library and, after that, dividing them in two

![Fig. 12 cons interface](image)

![Fig. 13 The output file of cons](image)

>EMBOSS_001

```plaintext
XXXxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxMSxxCxxxxExLxKxxxxxxxxxx
xxSExxxxxVxKLxxxxxLxxxxxGRLxxRlVyxDQxxIYEQELEKIFARSWVFxxHESQIPNx
GDYLTxTYIGEDNVIVRDXgSiKAFNLGTHGxxLCQDXgNTXSFXCNYGHRSFxID
GxLxGmxQGxxExFQxxSxxMNKREHGLxxPXRVDSYKGGLIFANLDPDAPTELEHYLgxxRyxL
DV1xxDMxxxGLEFVGxxKSxINCNWTTxVENFILDxYHXXxTXxxSxxEXxxEXxx
xDPDNNxxVxxxGHGWxxPNxxLDMxxGxxxxEXxxRxxLxxXXxXXxXX
xxGxxRaxxIxxxxxSxSVxxIPxxFLxxxxxxxXXXXXXXXxxxxTFRxWxxPxxGxxxxEXxXX
xxxxxxxxPxEIxxxxLxKxxxxxFSxxTGxExDxxxxFXxxSXXxxXXxRxXXXXxX
xxGSRxxExxxxxxxxxxxxxxxxxTxxLPGxxVxxxxxxxExNHxxFQQWLDLMQARxxxxDxxPxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

Fig. 13 The output file of cons. The figure shows a consensus sequence for Naphthalene 1,2-dioxygenase from marine prokaryotes
groups. One group would have been asked to screen the library with an enzymatic assay on plates, the second group would have followed the tutorial, designed the probe for colony hybridization and confirmed the activity in the recombinant *E.coli* extracts.

**Discussion**

Table 1 summarizes the learning goals of the tutorial. Beyond the immediate ones, i.e. becoming acquainted with very popular bioinformatics tools (Table I) and, in particular, of BRENDA [24], there are far-reaching educational aims. In our opinion the students should recognize the major opportunity offered by big data produced by metagenomics projects and the possibility of deriving protein properties by homology.

The tutorial emphasizes the opportunity offered by metagenomic next-generation sequencing projects. They are like mines we can dig to find what is useful for our research. It is highly probable that proteins homologous to the ones we are interested in are present in the uncharacterized big data stored in databanks.

---

**Fig. 14** Reverse Translate tool web page

---

**Fig. 15** Reverse Translate results. The figure shows the sequence to be used for the screening of the cosmid library
The tutorial highlights the concept that homologous enzymes share similar functions and that homology can be found comparing sequences. It capitalizes on the principle ‘Pairwise alignments whisper while multiple alignments shout out loud’ (Arthur Lesk). In fact, the students will identify the conserved regions that are a better fit to design a probe carrying out a multiple alignment of proteins homologous to Naphthalene 1,2-dioxygenase from *Pseudomonas putida*.

Consulting BRENDA [24] permits to retrieve original research papers concerning specific classes of enzymes. This will be precious to set up enzymatic assays and check that the cosmid clone isolated by colony hybridization indeed expresses Naphthalene 1,2-dioxygenase.

**Conclusions**

Computer-based learning is an excellent method to introduce undergraduate and graduate students with biological and biotechnological background to bioinformatics. We have shown that using bioinformatics tools as steps of a research project is more useful than presenting the same tools separately in stand-alone tutorials and that having a clear experimental objective, possibly related to a “trendy” topic such as green economy [38] raises the students’ interest.

**Supplementary information**

Supplementary information accompanies this paper at https://doi.org/10.1186/s12859-020-03560-z. Additional file 1. Student Satisfaction Questionnaire.docx

**Acknowledgements**

**About this supplement**

This article has been published as part of BMC Bioinformatics Volume 21 Supplement 10, 2020: Proceedings from the 13th Bioinformatics and Computational Biology International Conference – BBCC2018. The full contents of the supplement are available online at https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-21-supplement-10 .

**Authors’ contributions**

L.L. performed the experiments, gathered the questionnaires, analyzed the data. M. M and M.A. performed the experiments and prepared the Figs. B.H.M conceived the experiments. M.V.C. and L.L. wrote the paper. The author(s) read and approved the final manuscript.
Funding
Publication costs are funded MIUR PRIN 2015 2015JHLY35 to M.V.C. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials
The programs used during the current practice are available at:
1. BLASTP, https://blast.ncbi.nlm.nih.gov/Blast.cgi Accessed 03/09/2019
2. UniProt, http://www.uniprot.org/ Accessed 03/09/2019
3. Clustal Omega, https://www.ebi.ac.uk/Tools/msa/clustalo/ Accessed 03/09/2019
4. Cons, http://www.bioinformatics.nl/emboss-explorer/ Accessed 03/09/2019
5. Reverse translate, http://www.bioinformatics.org/sms2/rev_trans.html Accessed 03/09/2019
6. Clustal W, http://www.clustal.org/ Accessed 03/09/2019

The programs used during the current practice are available at:

Availability of data and materials
The programs used during the current practice are available at:

Funding
Publication costs are funded MIUR PRIN 2015 2015JHLY35 to M.V.C. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References
1. Campbell AM. Public access for teaching genomics, proteomics, and bioinformatics. Cell Biol Educ. 2003;2(2):98–111.
2. Pevzner P, Baran R. Computing has changed biology—biology education must catch up. Science. 2009;325(5940):541–2.
3. Letchford J, Corradi H, Day T. A flexible e-learning resource promoting the critical reading of scientific papers for science undergraduates. Biochem Mol Biol Educ. 2017;45(6):483–90.
4. Cirrincione C, Liguori L, Monticelli M, Andreotti G. Citro V. E-Learning for Rare Diseases: An Example Using Fabry Disease. Int J Mol Sci. 2017;18(10):2049.
5. King MD, Phillips P, Turner MW, Katz M, Lew S, Bradburn S, Andersen T, McDougal OM. Computational exploration of a protein receptor binding space with student proposed peptide ligands. Biochem Mol Biol Educ. 2016;44(1):63–7.
6. Korcsmaros T, Dunai ZA, Vellai T, Csermely P. Teaching the bioinformatics of signaling networks: an integrated approach to facilitate multi-disciplinary learning. Brief Bioinform. 2013;14(5):618–32.
7. Ray S, Koshy NR, Reddy PJ, Srivastava S. Virtual labs in proteomics: new E-learning tools. J Proteome. 2012;75(9):2515–25.
8. Kossida S, Tahri N, Daizadeh I. Bioinformatics by example: from sequence to target. J Chem Educ. 2002;79(12):1480.
9. Blatter MC, Baille Gierretsen V, Palagi PM, Bougerier L. Xenarios I. The Metagenomic Pizza: a simple recipe to introduce bioinformatics to the layman. EMNret.journal. 2016;22:e864.
10. Hingamo P, Broich C, Tallal E, Gauthier D, Thriffty D, Herrmann C. Metagenome annotation using a distributed grid of undergraduate students. PLOS Biol. 2008;6(11)e296.
11. Gibbens BB, Scott CL, Hoff CD, Schottel JL. Exploring metagenomics in the laboratory of an introductory biology course. J Microbiol Biol Educ. 2015;16(1):34–40.
12. Edwards RA, Haggerty JM, Casman N, Busch IC, Aguiñaldo K, Chintia S, Vauphn MH, Morey R, Haraks T, Teiling C, et al. Microbes, metagenomes and marine mammals: enabling the next generation of scientist to enter the genomic era. BMC Genomics. 2013;14(1):600.
13. Quaini R, Valdes J, Jedlicki E, Holmes DS. The use of bioinformatics and genome biology to advance our understanding of bioleaching microorganisms. In: Donati E.R., Sand W. (eds). Microbial Processing of Metal Sulfides. Springer, Dordrecht. 2007:221–39.
14. Welsch ME, Snyder SA, Stockwell BR. Privileged scaffolds for library design and drug discovery. Curr Opin Chem Biol. 2010;14(3):347–61.
15. Qu Y, Shen E, Ma Q, Zhang Z, Liu Z, Shen W, Wang J, Li D, Li H, Zhou J. Biodegradation of indole by a newly isolated Cupriavidus sp. SHE. J Environ Sci (China). 2015;34:126–32.
16. Gibson DT, Parales RE. Aromatic hydrocarbon dioxygenases in environmental biotechnology. Curr Opin Biotechnol. 2000;11(3):236–43.
17. Kennedy J, Flemer B, Jackson SA, Lejon DP, Morrissey JP, O’Gara F, Dobson AD. Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. Mar Drugs. 2010;8(3):608–28.
18. Rath CM, Janto B, Earl J, Ahmed A, Hu FZ, Hillier L, Dahlgren M, Kreft R, Yu F, Wolff JF, et al. Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. ACS Chem Biol. 2011;6(11):1244–56.
19. Subramani R, Albersberg W. Marine actinomycetes: an ongoing source of novel bioactive metabolites. Microbiol Res. 2012;167(10):571–80.
20. Wadhawan A, Sindhur R, Parameswaran B, Sukumaran RK, Pandey A. Metagenome analysis: a powerful tool for enzyme bioprospecting. Appl Biochem Biotechnol. 2017;183(2):636–51.
21. Popovic A, Tchigvintsev A, Tran H, Chemikova TN, Golysheva OV, Yakimov MM, Golyshev PN, Yakunin AF: Metagenomics as a Tool for Enzyme Discovery. Hydrolytic Enzymes from Marine-Related Metagenomes. In: Prokaryotic Systems Biology. Edited by Krogan PNJ, Babu PM. Cham: Springer International Publishing; 2015: 1–Metagenomics as a Tool for Enzyme Discovery: Hydrolytic Enzymes from Marine-Related Metagenom20.
22. Almaalabadi AD, Gojobori T, Mineta K. Marine Metagenome as a resource for novel enzymes. Genomics, Proteomics Bioinformatics. 2015;13(5):290–5.
23. Ufanti L, Laville E, Duquesne S, Potocki-Veronese G. Metagenomics for the discovery of pollutant degrading enzymes. Biotechnol Adv. 2015;33(8):1845–54.
24. BRENDA: [https://www.brenda-enzymes.org] 2019. Accessed 03 Sept 2019.
25. Schomburg I, Jeske L, Ulbrich M, Placzek S, Chang A, Schomburg D. The BRENDA enzyme information system—from a database to an expert system. J Biotechnol. 2017;261:194–206.
26. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008;36(Web Server issue):W5–9.
27. Laursen L. Spain’s ship comes in. Nature. 2011;475(7354):16–7.
28. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, et al. The sorcerer II Global Ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. PloS Biol. 2007;5(3):e77.
29. Sunagawa S, Karsenti E, Bowler C, Bork P. Computational eco-systems biology in Tara oceans: translating data into knowledge. Mol Syst Biol. 2015;11(5):809.
30. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques. 2000;28(6):1102 1104.
31. The UniProt Consortium. UniProt: a worldwide hub for protein knowledge. Nucleic Acids Res. 2019;47:D506–15.
32. Resnick S, Lee K, Gibson D. Diverse reactions catalyzed by naphthalene dioxygenase fromPseudomonas sp strain NCB 9816. J Ind Microbiol. 1996;17(5–6):438–57.
33. Parales JV, Kumar A, Parales RE, Gibson DT. Cloning and sequencing of the genes encoding 2-nitrotoluene dioxygenase from Pseudomonas sp. J542. Gene. 1996;181(1–2):57–61.
34. Sievers F, Higgins DG. Clustal omega for making accurate alignments of many protein sequences. Protein Sci. 2018;27(1):135–45.
35. Mullan LJ, Blearby AJ. Short EMBOSS user guide. European molecular biology open software suite. Brief Bioinform. 2002;3(1):92–4.
36. Olson SA. EMBOSS opens up sequence analysis. European molecular biology open software suite. Brief Bioinform. 2002;3(1):87–91.
37. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000;16(6):276–7.
38. Plotka-Wasyłka J, Kurowska-Susdor A, Sajid M, de la Guardia M, Namieśnik J, Tobiszewski M. Green chemistry in higher education: state of the art, challenges, and future trends. ChemSusChem. 2018;11(7):2845–58.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.