EMBnet.journal
Bioinformatics in Action

Blatter et al. EMBnet.journal 2016, 22, e864
http://dx.doi.org/10.14806/ej.22.0.864

Reports

The Metagenomic Pizza: a simple recipe to introduce bioinformatics to the layman

Marie-Claude Blatter[1], Vivienne Baillie Gerritsen[1], Patricia Palagi[1], Lydie Bougueleret[1], Ioannis Xenarios[1 2]

[1] Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland; 2 Geneva University, Geneva, Italy.

Introduction

Bioinformatics touches many aspects of everyday life – health, nutrition, environmental care, forensics – and is a major element of modern research in the life sciences. However, it is a fairly young scientific domain and is still poorly known to the layman.

In January 2013, a widespread food contamination scandal arose regarding beef *lasagne* that contained horsemeat. This led us to imagine a workshop to explain, in a simple but engaging way, how to identify food ingredients by way of DNA and bioinformatics tools available on the Internet. 'The Metagenomic Pizza' was born.

This article describes the Metagenomic Pizza, one of several guided bioinformatics activities that are available on the 'Ateliers de Bioinformatique' website (mainly in French).

The Metagenomic Pizza Workshop: step by step

1. Biodiversity in a pizza

To get people started, participants are asked to imagine different pizza recipes and ingredients (i.e., pizza crust, mozzarella, nutella, ham, etc.), and consider the various animal and vegetable species that can be found in a pizza (i.e., wheat, buffalo, palm, cacao, hazelnuts, pig). We end up with a list of 50 different organisms, including Homo Sapiens (one of the cook's hairs), horse, and other, invisible, organisms, such as yeast, bacteria and archaea – all of which can be found on a pizza, whether they should be there or not. The question of how it is possible to identify all these species in our food (or in other samples) is raised.

2. DNA: a little theory

A brief introduction describes how food comes from living organisms, and thus contains DNA. The DNA present in an uncooked pizza can be extracted and sequenced, thanks to Next Generation Sequencing (NGS) technologies. The different ways of representing DNA – from the well-known double helix explained in the biology classroom, to a DNA sequence and its digital single-strand format – is a key step for understanding the workshop. We then introduce the concept that certain regions in DNA sequences are specific to a given organism, much as a barcode is specific to a given item in a shop. Consequently, such a region can be used as a means of identification by comparison with already known data – which is where bioinformatics comes in.

3. DNA sequence analysis using the 'BLAST approach': pen-and-pencil activity

Participants are asked to fish out several 40-nucleotide DNA-reads from a box that contains hundreds (a foretaste of what is called 'big data' in metagenomics). They then manually compare these stretches of DNA with 50 annotated DNA "entries" from a "printed knowledgebase". This is presented in the form of a booklet in which are found, page after page in alphabetical order, the 50 DNA sequences. Each page also includes information on the organism it belongs to, as well as the function of the protein it encodes (Figure 1). This step involves comparing, two-by-two, stretches of DNA about 40 nucleotides long, and looking for tiny differences ("one small difference makes all the difference"); it is a painstaking task, and the participants rapidly ask for help from the computer.

The DNA read (ttcaaaactaacaatgttaccgccaggctttttgagcgcca) is from a tomato (*Solanum lycopersicum*), and encodes a protein involved in fruit pigmentation. A link to the accession number of the corresponding entry in UniProtKB/Swiss-Prot is provided (Q9ZNU6), as well as the name of the taxonomic group to which the organism belongs (*Dicotyledon Plant*).

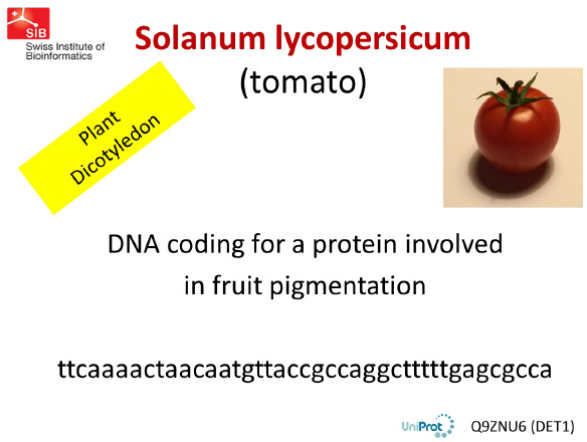

Figure 1. Example of an 'entry' out of the 50 found in the 'printed DNA knowledgebase'.

Article history
Received: 4 February 2016
Published: 16 March 2016

© 2016 Blatter *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at http://journal.embnet.org.

Blatter et al. EMBnet.journal 2016, 22, e864
http://dx.doi.org/10.14806/ej.22.0.864

4. DNA sequence analysis using BLAST: bioinformatics activity

The manually obtained results are checked on the computer. Participants perform a BLAST search (BLASTX) against the UniProtKB/Swiss-Prot knowledgebase at www.uniprot.org/blast. Little knowledge is required for someone to understand a BLAST output – the results are intuitive. Even school children easily grasp the fact that the computer is comparing their DNA sequence with all the sequences stored in an electronic knowledgebase. They are informed that the knowledgebase has 550,000 sequences, representing 12,000 species, and that the sequence most similar to theirs will be the first to appear in the list of results. To give the exercise a flavour of reality, some of our 40-nucleotide DNA-reads are not discriminative enough when using BLASTX against UniProtKB/Swiss-Prot. This provides an opportunity to discuss the fact that scientists sometimes need to use longer or other stretches of DNA for the unambiguous identification of (food) contaminants.

5. Universality of DNA and species classification

Once the taxonomic origin of the pizza DNA sequences has been identified, the participants map the DNA sequences onto a printed reference species tree. This is a nice way of illustrating the biodiversity found within a pizza, and the fact that DNA is found in every living organism.

6. The gender of the cook

Using yet another bioinformatics tool – *i.e.*, the BLAST-like alignment tool (Blat) of the UCSC Genome Browser – the participants are asked to check whether a DNA sequence discovered in the cook's hair bulb (atgcaatcatatgcttctgctatgttaagcgtattc) is located on chromosome Y or not. If it is, then the cook is a man.

From this point onwards, participants are left to their own devices, to extend their knowledge by using the skills they have acquired during the workshop. They can, for instance, type out a random DNA sequence of about 40 nucleotides and check whether this sequence actually exists or not, and whether it belongs to the human genome, or to the genome of another species. They can also try to see whether the previously identified tomato DNA sequence exists somewhere in the human genome.

Discussion

Since 2013, the workshop has been successfully offered to more than 2,000 people, from the age of nine years upwards, and from different backgrounds. These events took place in classrooms, during science fairs, university open houses, bioinformatics labs at the SIB Swiss Institute of Bioinformatics or during high-school teacher training courses. It is also one of the workshops given by (R)amène ta Science, a concept developed by Geneva University. This involves academic experts who train students – future 'ambassadors' – how to conduct the workshop. In turn, these ambassadors conduct the workshop at their own school.

Figure 2. A participant mapping the DNA sequences onto a printed reference species tree.

What are the advantages of such a workshop? It is highly adaptable in time (20 to 90 minutes), content and level of difficulty, and is thus convenient for all kinds of participant. Our main objective is to engage the layman in activities that are similar to authentic scientific research practice, and not to get lost in the technical know-how (Landhuis, 2015; Form and Lewitter, 2011). This way, the participants manipulate 'real' DNA sequences, either manually or with the help of bioinformatics tools used on a daily basis by scientists. It is an ideal way for them to understand the key role played by bioinformatics in the life sciences today. A few applications of current research in metagenomics are also discussed, such as the study of DNA preserved in the teeth of 1,000-year old skeletons. Participants learn how such studies are capable of identifying bacteria and food remains (Warinner *et al.*, 2014). They also discover how DNA derived from microbes extracted from agricultural soil, ocean surface water, or deep-sea whale bone (von Mering *et al.*, 2007), for example, helps to define new species or biological functions.

Acknowledgements

We thank Sandrine Pilbout for the original idea of the 'taxonomic' pizza recipe, and all the workshop participants, whether volunteers or not.

References

1. Form D and Lewitter F (2011) Ten simple rules for teaching Bioinformatics at the High School level *PLOS Computational Biology* **7** (10):e1002243. DOI: http://dx.doi.org/10.1371/journal.pcbi.1002243
2. Landhuis E (2015) Early BLAST OFF: bringing bioinformatics to secondary schools. Biomedical Computation Review: http://biomedicalcomputationreview.org/content/early-blast
3. von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, Ward N, Bork P (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**, 1126-1130. DOI: http://dx.doi.org/10.1126/science.1133420
4. Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, Radini A, Hancock Y, Tito RY, Fiddyment S, Speller C, Hendy J, Charlton S, Luder HU, Salazar-Garcia DC, Eppler E, Seiler R, Hansen LH, Castruita JA, Barkow-Oesterreicher S, Teoh KY, Kelstrup CD, Olsen JV, Nanni P, Kawai T, Willerslev E, von Mering C, Lewis CM Jr., Collins MJ, Gilbert MT, Ruhli F, Cappellini E (2014) Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336-344. DOI: http://dx.doi.org/10.1038/ng.2906