

# The bioinformaticist's/computational biologist's laboratory

## Opinion

### Setting the stage

My first official contact with bioinformatics was as an internship student in Anthony Ting's laboratory in the Institute of Molecular and Cell Biology (Singapore) back in late 1997. There, my task was to clone rSec6 and rSec8 proteins<sup>1</sup> for yeast 2-hybrid system. The eventual goal was to find out which proteins encoded by a mouse brain cDNA library can bind to these proteins. The main feature of these proteins is the presence of helical coil structures and Anthony introduced me to use COILS,<sup>2</sup> one of the first bioinformatics tool, developed to predict which part of the protein can form helical coil structures. You can still run COILS today at [http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html). Nevertheless, I was amazed by the concept of using computer programs to help in biological research.

A bit of my background is required here. I had been very interested in computers back in my high school and I could not make up choice for my tertiary education. In the end, I chose to go with a Diploma in Biotechnology on the reasoning that I will need a laboratory to learn about molecular biology but I can probably learn computing by myself. By the end of my first semester, I enrolled in Diploma in Computer Studies (offered by University of Cambridge Local Examinations Syndicate, via distance learning through a private education provider in Singapore) for part-time studies in 1996. I was pursuing both diplomas at the same time not knowing how biology and computing can merge eventually, as none of my biology lecturers back then demonstrated any better computing skills. In fact, my biology lecturers were almost rejecting computers. It was based on a conviction that eventually these 2 seemingly parallel lines can come together. Hence, when Anthony showed me how COILS can be useful, I saw light and I was hooked. These 2 lines got increasingly entangled all the way from my undergraduate days to amalgamate into possibly the first doctoral graduate in bioinformatics, coming from biochemistry major in undergraduate; in The University of Melbourne (I had written up a themed autobiography those who are interested in my journey, [http://maurice.vodien.com/monographs/SixYearsOfMelbourne\\_1stEd.pdf](http://maurice.vodien.com/monographs/SixYearsOfMelbourne_1stEd.pdf)).

### Bioinformatics versus computational biologist

Is there a difference between bioinformaticist and computer biologist? Strictly, speaking, there is. The NIH working definition<sup>4</sup> of a bioinformatics is the "research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data" while computational biology is "the development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems". Hence, NIH takes the view that the root word for "bioinformatics" is "informatics", and the focus is on "informatics" – the research and development of computational tools and approaches. Computational biology, on the other hand, is

Volume 3 Issue 1 - 2016

Maurice HT Ling<sup>1,2</sup>

<sup>1</sup>Colossus Technologies LLP, Republic of Singapore, Singapore

<sup>2</sup>School of BioSciences, The University of Melbourne, Australia

**Correspondence:** Maurice HT Ling, Colossus Technologies LLP, 8 Burns Road, Trivex, Singapore 369977, Republic of Singapore, Tel +6596669233, Email [mauriceling@colossus-tech.com](mailto:mauriceling@colossus-tech.com)

**Received:** January 27, 2016 | **Published:** February 01, 2016

focused on biology. Notwithstanding the semantics of the terms, I will generally see a bioinformatics as the "research and development of computational tools for use in biology" while computational biology is the "exploration and use of computational tools towards the understanding of biology".

Strictly speaking, a computer scientist venturing into the realm of biology will be a bioinformatics while a biologist using computational tools to study biology will be a computational biologist. In practice, this distinction can be blurred. It does not mean that a computer scientist/bioinformatics cannot learn enough biology to be a computational biologist or even cross the chasm to be a biologist, and vice versa. Take myself as an example, I have baccalaureates in both computing and biology; what does that make me? Schizophreniac, perhaps, before I can take this argument and distinction to ad absurdum.

### Bioinformaticist's / computational biologist's laboratory

What is then, the basic piece of equipment for either a bioinformatics or a computational biologist? It will be the computer. In fact, a computer, a regular laptop, can be the entire laboratory for a bioinformatics, as much as a laptop is the entire development hardware for a software engineer. Every software or developmental platform installed represents a piece of laboratory equipment. For a computational biologist, it can be a bit trickier. It does depend on the nature of the research and development work. Very often, a computational biologist will need to have access to a physical experimental laboratory (the "wet" laboratory) to validate the computational results. Nevertheless, the computer (representing the "dry" laboratory) is as crucial, if not more, to the computational biologist. In fact, the very moment a biologist calls himself or herself a computational biologist, he/she is welded to the computer, as much as a mathematical biologist can be separated from his/her mathematics.

However, it can be possible that a laptop forms the entire research hardware for a computational biologist but this depends much on the area of study. There are many areas in biology that wet laboratory experiments can be very difficult if not impossible. Examples of such will be evolution and epidemiology. It is obvious that much of epidemiological research is based on data collection from field sources rather than experimental sources. Data collection from field

sources is an important aspect of epidemiology.<sup>3</sup> We are not able and should never be allowed to infect humans or animals to trace routes of infections or virulence. From these collected field data, simulations are performed.<sup>5</sup>

Similarly, studying evolution experimentally is both difficult and expensive.<sup>6</sup> Although a number of experimental evolution studies had been carried out,<sup>7–10</sup> they are restricted to evolution of microorganisms which has a much shorter time generation time compared to eukaryotes. For example, it is impossible to study human evolution in an experimental setting, both in terms of ethics and in duration. Computer simulations of virtual organisms (commonly known as “digital organisms” or “artificial life”) had been used instead<sup>11</sup> and may provide some insights into human evolution.<sup>12</sup> Digital organism simulations had also been used to provide insights into areas that are ethically impossible to access, such as antibiotics resistance,<sup>13,14</sup> as it is not ethical to willfully induce antibiotics resistance.

In the current world of biological research, availability and access to equipment is an important aspect. However, the purchase and maintenance of such equipment comes with a heavy price tag and this adds to the un-levelness of playing field between nations and even within various research groups in the same institution. It usually ends up with the funded gets more funding, and funding comes with good research results to show. This can easily degrade into a competition for funding and prestigious appointments, creating a breeding ground for scientific fraud.<sup>15</sup> In the end, everyone, from tax payers to institutions to perpetrators, loses.<sup>16</sup>

However, in the realm of bioinformatics and computational biology, a single piece of computer can provide a rather leveled playing field when it comes to equipment and maintenance funding. A significant number of my research work, which resulted in publications,<sup>12–14,17–21</sup> started off as pet projects that I engage out of my salaried jobs; hence, received no funding, using no prior experimental data that I had generated and have to pay for article processing fees levied by journal if I cannot obtain a full waiver. Despite so, with a laptop, I am able to do something. With a laptop in hand, I am limited but at least, I am not disabled; for the humble laptop is my laboratory.

I see the humble laptop to a bioinformatics or computational biologist as what pen and paper is to a pure mathematician or a theoretical physicist. At the same time, I do understand that this road as a Bioinformatics or computational biologist or both may not be suitable for everyone. Nevertheless, it may be a potential option in a resource or funding limited setting. At the end of the day, we have enormous volumes of experimental data made publically available for re-use. At this moment (27th January 2016), National Center for Biotechnology Information (NCBI) has more gene expression data of more than 1.7million biological samples in its Gene Expression Omnibus (GEO) database and almost 300 genomes sequenced in its genome database, and this is just a tiny fraction of the available data resource available.

It should be noted that even with such massive volume of data publically available, it may be possible to reach a conclusion whereby the data needed is not yet available. Then, it is both a personal and professional choice to make. In spite of this, I believe that there is an enormous open field to wander by using available data, and with an innocuous and relatively affordable laptop as a mobile laboratory, what is there to lose? I enjoy doing my research while taking a bus ride or while drinking coffee or even, churning purposefully while I am sleeping. Give it a try.

## Acknowledgements

None.

## Conflict of interest

The author declares no conflict of interest.

## References

- Hsu SC, Ting AE, Hazuka CD, et al. The Mammalian Brain rsec6/8 Complex. *Neuron*. 1996;17(6):1209–1219.
- Lupas A, Van Dyke M, Stock J. Predicting Coiled Coils from Protein Sequences. *Science*. 1991;252(5009):1162–1164.
- Buring JE. Primary Data Collection: What Should Well-Trained Epidemiology Doctoral Students be Able to Do? *Epidemiology*. 2008;19(2):347–349.
- Huerta M, Haseltine F, Liu Y, et al. NIH Working Definition of Bioinformatics and Computational Biology. 2000.
- Bellan SE, Pulliam JR, Scott JC, et al. How to Make Epidemiological Training Infectious. *PLoS Biol*. 2012;10(4):e1001295.
- Batut B, Parsons DP, Fischer S, et al. *In Silico* Experimental Evolution: A Tool to Test Evolutionary Scenarios. *BMC Bioinformatics*. 2013;14(Suppl 15):S11.
- Lee CH, Oon JSH, Lee KC, et al. *Escherichia coli* ATCC 8739 Adapts to the Presence of Sodium Chloride, Monosodium Glutamate, and Benzoic Acid After Extended Culture. *ISRN Microbiology*. 2012:965356.
- Goh DJW, How JA, Lim JZR, et al. Gradual and Step-Wise Halophilization Enables *Escherichia coli* ATCC 8739 to Adapt to 11% NaCl. *Electronic Physician*. 2012;4:527–535.
- How JA, Lim JZR, Goh DJW, et al. Adaptation of *Escherichia coli* ATCC 8739 to 11% NaCl. *Dataset Papers in Biology*. 2013:219095.
- Loo BZL, Low SXZ, Aw ZQ, et al. *Escherichia coli* ATCC 8739 Adapts Specifically to Sodium Chloride, Monosodium Glutamate, and Benzoic Acid after Prolonged Stress. *Asia Pacific Journal of Life Sciences*. 2014;7(3):243–258.
- Ling MHT. Applications of Artificial Life and Digital Organisms in the Study of Genetic Evolution. *ACSIJ*. 2014;3(4):107–112.
- Castillo CFG, Ling MHT. Digital Organism Simulation Environment (DOSE): A Library for Ecologically-Based *In Silico* Experimental Evolution. *ACSIJ*. 2014;3(1):44–50.
- Castillo CFG, Ling MHT. Resistant Traits in Digital Organisms Do Not Revert Preselection Status despite Extended Deselection: Implications to Microbial Antibiotics Resistance. *Biomed Res Int*. 2014;2014:648389.
- Castillo CFG, Chay ZE, Ling MHT. Resistance Maintained in Digital Organisms Despite Guanine/Cytosine-Based Fitness Cost and Extended De-Selection: Implications to Microbial Antibiotics Resistance. *MOJ Proteomics Bioinform*. 2015;2(2):00039.
- Lisberger SG. Sound the Alarm: Fraud in Neuroscience. *Cerebrum*. 2013;2013:6.
- Freckelton I. Criminalising research fraud. *J Law Med*. 2014;22(2):241–254.
- Chan OYW, Keng BMH, Ling MHT. Correlation and Variation Based Method for Reference Genes Identification from Large Datasets. *Electron Physician*. 2014;6(1):719–727.
- Chia CY, Lim CWX, Leong WT, et al. High Expression Stability of Microtubule Affinity Regulating Kinase 3(MARK3) Makes It a Reliable Reference Gene. *IUBMB Life*. 2010;62(3):200–203.

19. Heng SSJ, Chan OYW, Keng BMH, et al. Glucan biosynthesis protein G (mdoG) is a Suitable Reference Gene in *Escherichia coli* K-12. *ISRN Microbiol.* 2011;2011:469053.
20. Keng BMH, Chan OYW, Heng SSJ, et al. Transcriptome Analysis of *Spermophilus lateralis* and *Spermophilus tridecemlineatus* Liver Does Not Suggest the Presence of Spermophilus–liver–specific Reference Genes. *ISRN Bioinformatics.* 2013;2013:1–8.
21. Too IHK, Ling MHT. Signal Peptidase Complex Subunit 1 (SPCS1) and Hydroxyacyl–CoA Dehydrogenase Beta Subunit (HADHB) are Suitable Reference Genes in Human Lungs. *ISRN Bioinformatics.* 2012;2012:790452.