

Biological experiments become transparent – anywhere, any time

February 14 2017

Biological experiments are generating increasingly large and complex sets of data. This has made it difficult to reproduce experiments at other research laboratories in order to confirm – or refute – the results. The difficulty lies not only in the complexity of the data, but also in the elaborate computer programmes and systems needed to analyse them. Scientists from the Luxembourg Centre for Systems Biomedicine (LCSB) of the University of Luxembourg have now developed a new bioinformatics tool that will make the analysis of biological and biomedical experiments more transparent and reproducible.

The tool was developed under the direction of Prof. Paul Wilmes, head of the LCSB group Eco-Systems Biology, in close cooperation with the LCSB Bioinformatics Core. A paper describing the tool has been published in the highly ranked open access journal *Genome Biology*. The new bioinformatics tool, called IMP, is also available to researchers online (r3lab.uni.lu/web/imp/).

Biological and biomedical research is being inundated with a flood of data as new studies delve into increasingly complex subjects – like the entire microbiome of the gut – using faster automated techniques allowing so-called high-throughput experiments. Experiments that not long ago had to be carried out laboriously by hand can now be repeated swiftly and systematically almost as often as needed. Analytical methods for interpreting this data have yet to catch up with the trend. "Each time you use a different method to analyse these complex systems, something different comes out of it," says Paul Wilmes. Every laboratory uses its

own computational programs, and these are often kept secret. The computational methods also frequently change, sometimes simply due to a new operating system. "So it is extremely difficult, and often even impossible, to reproduce certain results at a different lab," Wilmes explains. "Yet, that is the very foundation of science: an experiment must be reproducible anywhere, any time, and must lead to the same results. Otherwise, we couldn't draw any meaningful conclusions from it."

The scientists at LCSB are now helping to rectify this situation. An initiative has been launched at the LCSB Bioinformatics Core, called "R3 – Reproducible Research Results". "With R3, we want to enable scientists around the world to increase the reproducibility and transparency of their research – through systematic training, through the development of methods and tools, and through establishing the necessary infrastructure," says Dr. Reinhard Schneider, head of the Bioinformatics Core.

The insights from the R3 initiative are then used in projects such as IMP. "IMP is a reproducible pipeline for the analysis of highly complex data," says Dr. Shaman Narayanasamy. As co-author of the study, he has just completed his doctor's degree on this subject in Paul Wilmes' group. "We preserve computer programs in the very state in which they delivered certain experimental data. From this quasi frozen state, we can later thaw the programs out again if the data ever need reprocessing, or if new data need to be analysed in the same way." The scientists also aggregate different components of the analytical software into so-called containers. These can be combined in different ways without risking interference between different program parts.

"The subprograms in the containers can be strunged together in series as needed," says the first author of the study, Yohan Jarosz of the Bioinformatics Core. This creates a pipeline for the data to flow through.

Because the computational operators are frozen in containers, one does not need reference data to know the conditions – e.g. type of operating system or computer processor – under which to perform the analysis. "The whole process remains entirely open and transparent," says Jarosz. Every scientist can thus modify any step of the program – of course diligently recording every part of the process in a logbook to ensure full traceability.

Paul Wilmes is especially interested in using this method to analyse metagenomic and metatranscriptomic data. Such data are produced, for example, when researching entire bacterial communities in the human gut or in wastewater treatment plants. By knowing the full complement of DNA in the sample and all the gene products, they can determine what bacterial species are present and active in the gut or treatment plant. What is more, the scientists can also tell how big the population of each bacterial species is, what substances they produce at a given point in time, and what influences the organisms have on one another.

The catch, until recently, was that researchers at other laboratories have had a hard time reproducing the experimental results. With IMP, that has now changed, Wilmes continues: "We have already put [data](#) from other laboratories through the first tests with IMP. The results are clear: We can reproduce them – and our computations in IMP bring far more details to light than came forth in the original study, for example identifying genes that play a crucial role in the metabolism of bacterial communities."

"Thanks to IMP, only standardised and reproducible methods are now used in microbiome research at LCSB – from the wet lab, where the experiments are done, to the dry lab, where above all computer simulations and models are run. We have an internationally pioneering role in this," says Wilmes. "Thanks to R3, IMP also sets standards which other institutes, not only LCSB, will surely be interested to apply," adds

Reinhard Schneider of the Bioinformatics Core. "We therefore make the technology of other researchers openly available – the standard ought to be quickly adopted. Only reproducible analyses of results will advance biomedicine in the long term."

More information: Shaman Narayanasamy et al. IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses, *Genome Biology* (2016). [DOI: 10.1186/s13059-016-1116-8](https://doi.org/10.1186/s13059-016-1116-8)

Provided by University of Luxembourg

Citation: Biological experiments become transparent – anywhere, any time (2017, February 14) retrieved 25 April 2024 from <https://phys.org/news/2017-02-biological-transparent.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.