

Dr. Marnix H. Medema  
Assistant Professor

(+31) 317484706  
marnix.medema@wur.nl

Bioinformatics Group  
Wageningen University  
Droevendaalsesteeg 1  
6708PB Wageningen  
The Netherlands

November 30<sup>th</sup>, 2018

To whom it may concern,

The PhD thesis of Alexey Gurevich addresses two interesting and highly relevant challenges in the field of computational metabologenomics:

1. Quality control of genomic and metagenomic sequences.
2. Dereplication of peptidic natural products and identification of novel variants thereof.

These challenges are highly relevant, as recent technological advances now provide large-scale paired genomic and metabolomic data that allow a new understanding of chemical interactions in microbiomes, as well as effective new strategies for natural product discovery. Quality control and dereplication are crucial steps within these procedures. On both fronts, Alexey has made a tremendous contribution that effectively sets the standard for the field for years to come.

The QUASt and metaQUAST tools for quality assessment of (meta)genome assemblies provide a carefully chosen set of well-chosen statistical parameters that make it possible to rapidly assess assembly contiguity and correctness. The tools are very user-friendly and can be run by non-experts. Also, they are flexible, in the sense that they provide options to assess quality based on (a) reference genome(s) or in the absence of any reference. These tools have made an enormous impact on the field, which can also be seen from the fact that the QUASt paper has accumulated more than a thousand citations: it has become the de facto standard for a crucial step in all genome-based scientific projects.

The Dereplicator and VarQuest tools provide rapid means to identify known peptidic natural products in massive mass spectrometry datasets, as well as to find novel variants of these important molecules. The use of theoretical spectra and the new strategies for constructing PNP decoy databases provide a smart solution that allow accurate and scalable dereplication. Additionally, the discovery (made using VarQuest) that thousands of hitherto unknown variants of known natural products naturally exist, and that, consequently, most PNP biosynthetic pathways in fact produce 'libraries' of multiple molecules, changes the way the natural product field looks at these pathways and is likely to have a large impact on how research in this area is being performed.

---

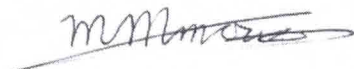
09/2 - 254 omz 30.11.18

---

Compared to other PhD theses I have seen, the quality and impact of Alexey's work are absolutely remarkable. The methods are innovative, the results are of high impact, and the chapters are very clear and well-written. Also, all chapters have been published in highly esteemed journals (including Bioinformatics, Nature Chemical Biology and Nature Microbiology) and have been very well received by the scientific community.

Beyond a few grammatical errors (mainly in the introduction section, highlighted in the attached file), I have no concerns about the quality of the work, nor on the appropriateness/correctness of the methodology. I look forward to an exciting scientific discussion with the soon-to-be Dr. Gurevich on December 18th.

Yours sincerely,

A handwritten signature in black ink, appearing to read 'Marnix H. Medema', with a horizontal line drawn underneath it.

Dr. Marnix H. Medema

---