REVIEW

Of the Doctoral Dissertation of Gurevich Alexey Alexandrovich "Computational methods for analysis of error-prone metabologenomic data" submitted for defense of the degree of candidate of physico-mathematical sciences, specialization 03.01.09 – Mathematical biology, bioinformatics.

Relevance

The dissertation is divided into two main themes: 1) quality assessment of genome and metagenome assemblies, and 2) identification and dereplication of peptidic natural products (PNPs) from mass spectra data. The relevance of both themes is high.

Genome assemblies are the foundation of modern genomics. Because genomes cannot yet be read in a single, continuous pass, modern sequencing technologies must read many shorter fragments that are then computationally assembled into the original genome. This is an error-prone process, and errors in the assembly can propagate to all downstream analyses, possibly leading to false conclusions and failed experiments. Thus, validating the quality of genome assemblies is crucial to ensuring the accuracy of all downstream analyses, making this area of study highly relevant. This relevance is evident from the large community impact and high number of citations achieved by these publications. For example, Google Scholar reports >1,000 citations for the QUAST quality assessment tool described in chapter 1 of the dissertation, which is remarkable. A majority of these citations are from users who have used QUAST to measure and improve the quality of their own genome assemblies.

The second theme of the dissertation focuses on computational methods for high-throughput identification of known and novel PNPs from mass spectra data. This class of compounds includes many potential antibiotics, but they are difficult to predict and study due to their production via non-ribosomal peptide synthetases and other post-translational processes. Identifying these compounds directly from mass spectra data is especially challenging because many form nonlinear peptides (i.e. cyclic) and can include non-standard amino acids. There is an urgent need to discover and understand the action of novel antibiotics and other naturally occurring metabolites, so this area of study is also highly relevant.

Validity

The validity of the work presented in the dissertation is sound. Both methods and results are presented in a clear and understandable manner, and appropriate validation has been performed to demonstrate the accuracy of the methods. The English portion of the defense is well written overall and contains only a small number of minor grammatical errors (e.g. missing/extraneous articles) that can be resolved by a thorough proofing. I am unable to assess the Russian portion of the dissertation and am thankful that my review is not also required in Russian. As for the developed software, I have personally used a number of the described packages and found them to be very useful and of high quality.

Novelty

The primary contribution of the presented work is a collection of tools that work well for the problems they address. The assembly validation software programs are conglomerations of known methods but combined and presented in a unique way. The PNP methods address
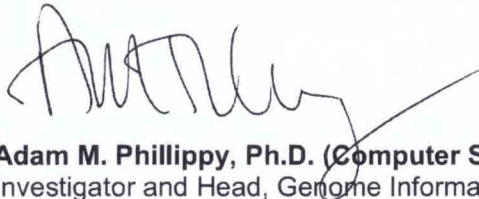
comparatively more challenging problems and demonstrate algorithmic novelty. Such well-designed tools are incredibly powerful for advancing research. They enable discovery and they amplify the impact of the dissertation, because they do not simply answer a single hypothesis; they enable others to answer countless hypotheses.

Conclusion

The dissertation of Alexey Gurevich includes work from four papers published in highly regarded and indexed journals. Both the published papers and the dissertation itself demonstrate excellent scholarship and cite the relevant prior works. As detailed above, the presented work is highly relevant, valid, novel, and accurately presented. Thus, the dissertation of Alexey Gurevich meets the requirements necessary for the granting of the degree of candidate of physico-mathematical sciences, specialty 03.01.09, and I support this action without hesitation. I commend the candidate on the quality of this work, and I thank him for the contributions he has made to these fields of research.

30 November 2018, serving in my personal capacity,

Adam M. Phillippy, Ph.D. (Computer Science)
Investigator and Head, Genome Informatics Section
Computational and Statistical Genomics Branch
National Human Genome Research Institute
National Institutes of Health
Bethesda, MD USA