

On “Bioinformatics in Russia: history and present-day landscape” by M.A. Nawaz, I.E. Pamirsky, and K.S. Golokhvast

Dear Editors,

As an author of several papers in *Briefings in Bioinformatics* and a reviewer for the journal, I've been surprised by the paper “Bioinformatics in Russia: history and present-day landscape” by M.A. Nawaz, I.E. Pamirsky, and K.S. Golokhvast (vol. 25, no. 6, p. bbae513). While the topic is clearly important, the paper does a disservice both to readers of *Briefings in Bioinformatics* and to Russian bioinformaticians. It provides random information (e.g. the entire chapter about the market state of IT, pharmaceutical, agrotechnology, and biotechnology industries), while giving no real insight into the history and problems of Russian bioinformatics, both applied (e.g., related to medical genetics) and fundamental (that is, molecular evolution, about which nothing is said at all). The choice of scientific topics to discuss seems to follow lists of publications at Internet sites of several universities and research institutes with no comprehensive overview.

As I've been involved in several of the activities mentioned by Nawaz *et al.*, and as I'm a co-author of more than dozen papers they reference (12, 39-43, 60-62, 144, 184, 186, 192-193), with even more referenced papers by my former students, I think I'm in a good position to note specific errors and omissions. This is not an exhaustive review, as I'll mainly restrict myself to the facts, events, and results about which I have a first-hand knowledge.

The history narrated by Nawaz *et al.* is rather incomplete. They mention (correctly) the Institute of Mathematical Biology in Pushchino (where the group of Alexei Kondrashov worked) and the Institute of Cytology and Genetics in Novosibirsk (where bioinformatics was pioneered by Vadim Ratner), but neglect early contributions by groups that ceased to exist or changed location, in particular, labs of Andrei Mironov at the State Research Institute of Genetics and Selection of Industrial Microorganisms (now at the Lomonosov Moscow State University) and Aleksandr Aleksandrov at the Institute of Molecular Genetics. Nothing is said about the Bioinformatics section of the Russian Human Genome Program whose conferences in the 90's collected contributions of almost all active groups. A crucial role in integrating the bioinformatics community was played by international conferences, *Bioinformatics of Genome Structure and Regulation* (BGRS, even years since 1998, Novosibirsk) and *Moscow Conference on Computational Molecular Biology* (MCCMB, odd years since 2003), and the regular *Moscow Bioinformatics Seminar* (biweekly since 1993 until early 10's, then less regular; I believe this seminar had been a unique phenomenon not only in Russian, but in the World bioinformatics). Important but not mentioned were contacts with the Russian bioinformatics diaspora, in particular (but by no means limited to) Eugene Koonin (NCBI), Leonid Mirny (MIT), Shamil Sunyaev, Vadim Gladyshev, and Peter

Kharchenko (Harvard Medical School), Pavel Pevzner (UCSD), Oleg Gusev (Juntendo); the latter four at some point had led labs in Russia.

I'm flattered that the entire "Information Box II" is dedicated to the work of the Research and Training Center "Bioinformatics" (RTCB) of the Institute for Information Transmission Problems that I have created in 2003 and led until this year, but the choice of papers and results to mention is random: too many minor papers are mentioned while really important ones (e.g. on riboswitches) are missing; the listed Internet resources have been mainly developed by my former students in the US, not in Russia; and Ref. 190 is not related to RTCB, nor even to Russia, as all its authors work in France and none is of the Russian descent.

The Genomics Core Facility of the Skolkovo Institute of Science and Technology (SkolTech) performs sequencing, but does not provide bioinformatics service, contrary to what is claimed by Nawaz *et al.* As in the previous case, the choice of results by Skoltech researchers is capricious, with entire successful research directions completely omitted (e.g. comparative systems biology of the brain or large-scale metagenomics). On the other hand, Skoltech is not involved in the project to sequence 100 000 human genomes aside of the fact that the latter is led by a former SkolTech professor Konstantin Severinov.

"Institute of Protein of the USSR Academy of Sciences" and "Protein Institute of the RAS" is commonly known as the "Institute of Protein Research". Hence it is not surprising that, as reported by Nawaz *et al.*, their "literature search produced limited results"; in particular they have not noticed fundamental contributions by the protein structure school created by Oleg Ptitsyn and maintained by Alexei Finkelstein.

Nawaz *et al.* mention bioinformatics programs in a number of universities, but merely list the flagship Faculty of Bioengineering and Bioinformatics at the Lomonosov Moscow State University (established in 2002) and Department of Information Biology at the Novosibirsk State University in the Appendix.

To that I add a number of other observations. The most interesting one is "the founding of the Academy of Sciences of the USSR (abbreviated as AH CCCP in Russian) by Pei Kib Breat in 1725". This mysterious person is the Russian tsar Peter the Great, he would had been surprised to learn back in 1725 about the coming of the Soviet Union in about two hundred years. (It has been brought to my attention by colleagues that the strange phrase originates from a text recognition error in the electronic copy of a *Nature* note published in 1946 [1].) The authors wrote, "George Gamow optimistically proposed a fairly precise genetic code in a letter to Watson and Crick [33]" — in fact, he did not; Gamow's major contribution had been understanding that a code linking DNA and proteins should exist, while the code he had suggested was very far from reality, as explicitly described in the cited Ref. 33. Ref. 8 has an interesting author "Union IT" and an editor "Center GI". Ref. 125 completely and Ref. 155 partially are given in Russian Cyrillic (it is not clear how they have survived the editorial process). Similarly, "the Scientific Research Computing Centre of the AH CCCP" is, of course, "... of the USSR Academy of Sciences". Three out of four first references (refs. 1, 2, and 4) are

not relevant to the papers' topic; this is just one small example. "Information Box – 1" does not deal with bioinformatics at all, while important papers on the SARS-CoV-2 evolution by Georgii Bazykin's group are not mentioned. The authors mention that large-scale projects on sequencing chickpea, soybean, and rice varieties are not realized yet — but how does that relate to their topic, Russian bioinformatics? — the references are given to the international projects.

Nawaz *et al.* notice an increase in the number of publications since 2014 and assign it to the establishment of the Russian Science Foundation. A more likely explanation is one of the 2012 decrees of then just re-elected president Vladimir Putin calling for an increase of the Russian share in the international bibliometrics. That percolated downstream and, in many universities, led to establishment of so-called *publication bonuses* of considerable value, hence inflating the publication rates. Similarly, Nawaz *et al.* write that "it is somewhat understandable from the literature survey that relatively more focus has been given to prokaryotic genome sequencing rather than eukaryotes". Indeed, the former is orders of magnitude cheaper than the latter, which is an important, yet unmentioned factor in Russia.

It is surprising that this paper has passed a strict reviewing process in a respected, high-impact journal.

[1] Tomkeieff S. I. The Academy of sciences of the U.S.S.R. *Nature* 1946; 158: 315. <https://www.nature.com/articles/158315a0>

Mikhail S. Gelfand