

August 09, 2012 | By Ioana Patringenaru

U.S.-Russian Collaboration Develops New Method for Sequencing Dark Matter of Life from a Single Cell



Pevzner

An international team of researchers led by computer scientist Pavel Pevzner, from the University of California, San Diego, have developed a new algorithm to sequence organisms' genomes from a single cell faster and more accurately. The new algorithm, called SPAdes, can be used to sequence bacteria that can't be submitted to standard cloning techniques—what researchers refer to as the dark matter of life, from pathogens found in hospitals, to bacteria living deep in ocean or in the human gut.

Ultimately, the researchers hope to apply this algorithm to cancer cells to monitor early stages of the disease when normal cells first turn into malignant ones. Pevzner and colleagues published their findings in the May issue of the *Journal of Computational Biology*. They released SPAdes Aug. 8.

Last fall, Pevzner's group, in collaboration with single-cell sequencing pioneer Roger Lasken at the J. Craig Venter Institute and researchers at Illumina Inc., developed the first software capable of handling single-cell sequencing. Researchers published those findings in *Nature Biotechnology* in September 2011. The fact that a new sequencing algorithm was developed in just months reflects the frenetic pace of progress in single-cell sequencing, one of the fastest-growing, and most important, areas in modern genomics.

Pevzner's group, which includes scientists from the Jacobs School of Engineering at UC San Diego and the Russian Academy of Sciences, along with Lasken's team, are now using SPAdes to sequence the bacterial dark matter of life and human pathogens.

The international collaboration is part of an ambitious “megagrant” initiative launched by then Russian President Dmitry Medvedev, including an invitation to 40 world-class scientists to help jumpstart Russian science, which has been faltering since the fall of the Soviet Union.

Megagrants brought to Russia experts in various fields, including some Nobel Prize and Fields Medal winners. Pevzner was the only researcher at the intersection of modern biology and computer science in the group.

He agreed to start DNA and protein sequencing projects in Russia at the Saint Petersburg Academic University, an elite graduate school headed by Nobel Prize Laureate Zhores Alferov. This was no easy task. There wasn’t a single computer science expert in the area of DNA and protein sequencing in the entire country. The only experts that were on hand in large numbers for this project were mathematicians, thanks to Russia’s strong tradition of educational excellence in the exact sciences.

But Pevzner still took a gamble and started a Laboratory for Algorithmic Biology (LAB) with a dozen of 20-something mathematicians and computer scientists, some of them undergraduate students, who knew nothing about DNA sequencing. They went through a series of grueling bioinformatics boot camps and just two months later started working on the SPAdes genomes assembler. Sergey Nurk was one of them.

“As an undergraduate student, I had been working as a programmer for some time and was almost ready to sell my soul to industry,” Nurk said. “Now, as a graduate student on Pavel’s team, I learned that looking for simpler, more elegant solutions is important. I also was reminded of the value of time—and the need to use it wisely.”

Six months later, working closely with Pevzner’s students and colleagues at the Jacobs School of Engineering and Professor Max Alekseyev at the University of South Carolina, the Russian team developed their new, extremely accurate assembler.

“Fragment assembly is not unlike assembling a puzzle from a billion pieces and it is often viewed as one of the most sophisticated problems in bioinformatics. A new assembler may take years to develop even by seasoned bioinformatics experts. The fact that young Russian researchers without prior bioinformatics experience developed SPAdes so quickly and improved on state-of-the-art assemblers in just half a year is remarkable,” Pevzner said.

The research was partially supported by grants from the National Institutes of Health and the Russian Megagrant Initiative.

UC San Diego's Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <https://ucsdnews.ucsd.edu/media-resources/faculty-experts>.