On the Wild Frontier

I have been lucky to find great puzzle solvers to work with me. I look for people who think in nontraditional ways, who can pursue a problem without an apparent answer. Sometimes students and postdocs want to look for problems they can solve with a technique they learned in school. That will not work in my lab.

I am interested in questions for which common sense has failed to find an answer. That is what I get excited about. It is always risky to tackle such puzzles, because you never know how to assemble the pieces. And it is difficult to write a grant for this type of work. For the most part I have succeeded, but there have been times I spent a year without any progress to show for it. That is why it is important to be able to evaluate the worthwhile problems.

A good example is our work on genomic rearrangements. For over 30 years, people believed that changes in the genomic architecture occur randomly. Three years ago Glenn Tesler and myself rebutted that model by providing evidence for fragile sites with a much higher propensity for rearrangements than the rest of the genome.

It was the first case I am aware of in which a biological theory fell because of a sophisticated computational theory. We did not generate any new data; we used a new algorithm to look at the mouse genome data. We could not say how many fragile regions are in the genome or where they are located, but nonetheless showed that they exist.

I make the analogy to the discovery of the planet Pluto. Pluto was first predicted to exist by American astronomer Percival Lowell, who immediately started to search for the new planet. However, no one knew where it was, and some scientists were skeptical of its existence. Pluto was found 20 years later by Lowell's successor Clyde Tombaugh. In the case of fragile regions, shortly after our prediction, scientists began finding them on a case-by-case basis. In July 2005, we published a paper identifying a large number of fragile regions, but the search is by no means complete.

We are now developing algorithms to predict where to find all the fragile sites. We think about them as footprints that have been left by evolution. If you were to study earthquakes, you would see that they use the same fault lines many times over, and the same elements keep getting destroyed. In the same way, we think that if we look at many evolutionary lineages, we will see that the same positions in the genome keep being broken and rearranged.

I started working in the field of bioinformatics in 1985. As a graduate student at the Moscow Institute of Physics and Technology, I read an article by Michael Waterman at the University of Southern California describing an algorithm for gene comparisons. I wrote a letter to him asking whether I could join his lab. He was kind enough to let a crazy Russian in.

At the time, there were so few people doing bioinformatics that we all knew each other and what we were working on. It is an amazing story to me how the discipline exploded in such a short time. Given the expansion of the field, it was clear that we should have a conference to bring people together, and in 1997, Michael Waterman, Sorin Istrail, and myself co-founded the International Conference on Research in Computation (RECOMB). It is still the best place to hear about algorithmic biology.

My web site carries a photograph of me dressed like a cowboy. The picture was taken when I was vacationing with my children in Yellowstone National Park. One of my colleagues says it is symbolic of bioinformatics being the wild frontier. To be honest, I was not trying to make a statement. I just liked the photo. But I also like my colleague's explanation.

Pavel Pevzner, Ph.D.

*Ronald R. Taylor Chair
in Computer Science,
University of California,
San Diego, CA*

*As told to Laura Bonetta,
a science writer based in Bethesda, MD.*