

Comment

The ninth wave

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.
Email: petsko@brandeis.edu

Published: 30 June 2006

Genome Biology 2006, **7**:109 (doi:10.1186/gb-2006-7-6-109)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2006/7/6/109>

© 2006 BioMed Central Ltd

Surfboard riders, borrowing an old sailor's expression, often speak of a 'ninth wave'. It means a single wave larger than all the others. Colossal, unexpected - ninth waves are the stuff of legend. It is said that nothing can withstand their power. I'm fascinated by all-pervading technologies that seem to spring up overnight. They are the ninth waves that wash over our culture with sudden, transforming power. The personal computer is most assuredly not one: it seemed to take forever before it made its way from business offices into most homes. But video rental is a great example of one. Didn't we all wake up one morning and find a video rental store in every shopping center, looking as though it had been there forever? In biology, the polymerase chain reaction most assuredly fits the definition: because of PCR, almost overnight, cloning went from something that was hard even for experts to something anyone could do, so everyone started doing it. Email is not one. It took years for email to replace telephone calls and regular mail as the main form of personal correspondence in business and academia, and it still hasn't done so outside of those venues.

We could argue whether the internet is a ninth wave (my personal opinion is that it isn't, since I'm old enough to remember that for a long time it was just a useful, but cumbersome, data-exchange mechanism that was restricted to government labs, the military, and a few universities). But surfing the web definitely is. Even before Google, which works so well that it has become a verb, just like Xerox (another example), the invention of the web browser changed the way we think about information, and did so with astonishing speed. For thousands of years, information was the property of a privileged elite, whose value depended on the fact that data were hard to come by, and they had access. Once, these high priests of knowledge were priests in fact: monasteries were the repository of learning for centuries. In the modern world, it is money that has tended to define who has access. As a luxury item, information - and the education needed to understand it and the technology required to obtain it - has tended to be found primarily in

the developed world. If people in developing nations wanted access, they usually had to emigrate to a developed country for education, and once there, they most often stayed. A taste for information, once acquired, is not easily forsaken.

Surfing the web has changed all that, and changed it in a heartbeat. With so much information so easy to come by, and with most of it available free of charge, information has suddenly become a commodity. No longer can it be hoarded. And the barriers to entry into the world of information have, equally suddenly, become very low. All that is needed is a computer, and they are relatively cheap. To meet the demand for access from those without their own computers, the internet café has sprung up almost as rapidly, and spread as widely, as the video rental store once did. Young people all over the world are now accustomed to virtually unlimited access to a virtually unlimited store of knowledge.

The social consequences of this have been profound, no less so in science than in other aspects of life. Because the ability to access information became widely available at the same time that the genomics revolution was producing a flood of data (the ninth wave of biology, at least for this generation), people who could organize and make sense of the data had enormous value. And such people could not only come from anywhere; they could work anywhere.

I haven't seen any statistics to support my contention, but I believe that, once their training is completed, graduate students and postdocs today return to their own countries much more often than they used to. Partly this is because other countries are investing more in science and technology, so facilities and opportunities are better. But a large part of it is access to information. Remember how often we in the developed nations used to gather up our old journals and send them off to less-developed countries where they had none? I haven't had a request like that in quite some time. Internet surfing, combined with the rapid rise of open access publishing, has made many scientific articles accessible anywhere in

the world. I think this is a very healthy trend. Countries like India and China have not had the infrastructure to compete with the West scientifically, but they've never lacked for brainpower. Thanks to the widespread availability of information, that brainpower can now be used to tackle many of the questions that genomics, in particular, has raised.

But this trend also presents a great danger. A PubMed search for the term 'bioinformatics' produces 12,657 scientific articles, not one of which is older than 1993. In fact, over 12,000 of these articles - more than 95% - were published since 1999. This explosive growth is fueled by a number of factors: widespread data access thanks to the internet; an armada of computationally savvy people thanks to a decade of surfing that same internet; the relatively low cost of setting up the research program of a newly hired faculty member in bioinformatics; and above all, a desperate need on the part of biologists to make sense of the flood of data produced by genomics. And it is this demand for analysis, combined with the reciprocal demand from bioinformatics for more data to analyze, that constitutes the danger, because heavy demand is rarely associated with high quality.

People often complain that there is nothing good on television. That is simply nonsense. There are many first-rate programs. But there are many more bad ones. The reason is simple: with the advent of hundreds of cable channels there is an insatiable demand for content to fill the enormous number of programming hours, and there aren't enough quality offerings to make much of a dent in that huge demand. Quality programming, like quality research, is a pretty fixed, relatively rare quantity, and its frequency is largely independent of demand. Increased demand does bring some additional high-level offerings, but mostly the extra slots just get filled with mediocrity.

Bioinformatics and genomics are creating a huge demand for data and data analysis, neither of which should be confused with greater understanding of how the world works. To analyze something is not *de facto* to understand it. In my experience, correlations are interesting but causality can only be proven by carefully designed experiments. Yet analysis is cheap and seems useful, data gathering is popular and easy to justify because it produces reams of tangible results - and there seems to be less and less room for hypothesis-driven, experimental research. Properly designed, clever experiments are hard to do and don't always yield clear-cut answers. Computational analysis of someone else's data, on the other hand, always produces results, and all too often no one but the cognoscenti can tell if these results mean anything.

Funding agencies feel the need to learn something from the mass of information their genomics and genomics-enabled projects are generating. Given the choice between lengthy, difficult, expensive individual-investigator-initiated experiments and inexpensive, flashy computational studies that

are guaranteed to produce something quickly, it seems pretty obvious which they are likely to prefer. The fact that such studies can be done anywhere in the world only adds to their popularity. And while I have learned a lot from some bioinformatics papers, I still much prefer, and have been taught much more about the world by, a good experimental study - which, I fear, may be in danger of going the way of the dodo.

It isn't as bleak as it seems, though. Hidden in the results of genomics and proteomics and structural genomics and metabolomics and transcriptomics and god-knows-what-other-omics studies are a wealth of hypotheses waiting to be formulated and tested experimentally. Bioinformatics can help find them. We need to demand that it do that, and, if it doesn't, we need to harness its tools ourselves and use them to do that. Some biologists are already taking that approach. If more did, we might all be able to ride this wave together.

There's a painting by the 19th century Russian artist Ivan Aivazovsky called 'The Ninth Wave'. (You can read about him at http://center.rusmuseum.ru/inetbook/gaivazan_pict_eng.htm.) It depicts a huge wave about to crash down on the survivors of a shipwreck, who are desperately clinging to the broken mast. The question for all of us, as biology - driven by the combined forces of genomics and bioinformatics - seems about to become an information science, is whether hypothesis-driven, individual-investigator-initiated experimentation is about to suffer the fate of the people in the painting. If it does, we will all be poorer for it. Or is it possible that, like surfers who actually wait for the ninth wave, hypothesis-driven research will somehow manage to climb to the crest of this trend and use its enormous energy? One thing seems clear: if that happens, it will be a heck of a ride.