

## Debates over credit for the annotation of genomes

**Researchers who devote themselves to sequencing genomes often lack the time to interpret their results. Others don't. The tensions that can result reflect the need for a rethink of sequencers' priorities or a change in approach to collaboration.**

It has become a well-established tradition among the publicly funded genome-sequencing community to place sequence data on public and freely accessible databases as the sequences are generated. But there is nothing to stop others from using those data to do good science. Indeed, what else would be the point of a policy of prompt and open access?

But how should credit be attributed? What rights do those doing the sequencing have over subsequent dependent research? And do people putting data and analysis freely on the web prejudice their chances of publication?

To take the last question first: not in the case of *Nature* (or, for that matter, other *Nature* journals). We believe that genomics databases, like preprint servers and conferences, represent a form of intra-community networking from which all researchers benefit. *Nature* does not count them as prior publications. If, exceptionally, an exciting result gets picked up from such a source by the media, that, too, will not necessarily disqualify a paper from consideration, as long as researchers have not pre-empted peer review or publication by encouraging prior publicity. (For a fuller statement of policy, see <http://www.nature.com/nature/author/embargo.html>.) As always, *Nature* will apply its judgement on the significance of a submitted sequence, recognizing that formal peer-reviewed publication of genome maps and sequences represents a necessary culmination of years of research, allowing authors to communicate results and commentary on their significance to the wider world and to gain due credit for their efforts.

This policy applies not only to raw sequence data but also to 'annotations': proposals for the functions of the genes in the database. These often represent substantial pieces of research in themselves. If the results have been added as annotations to a recognized database but have not been subject to the process of peer-reviewed publication, their inclusion on the database is not considered as prior publication by *Nature*.

### Publication rights

So far, so good. But the problem, from the point of view of those doing the sequencing, occurs on occasions when they are getting on with their sequencing while others, perhaps better placed to annotate the sequence, are free to use it to publish biologically useful information. What rights of first publication do the sequencers then have? As the discoverers of the sequence, they surely deserve some credit in the subsequent elucidation of function — credit that should extend beyond a simple reference to the database website. Yet, once someone else has annotated their sequence data, the sequencers' own ambitions in that direction have effectively been pre-empted.

These genomes are not mere scientific curiosities. For example, the sequences of the 14 chromosomes of the protozoan *Plasmodium falciparum* and the 11 of *Trypanosoma brucei* represent keys to significant advances in the treatment of malaria and sleeping sickness,

respectively. It would be self-evidently undesirable to allow the accumulating data, and those of other pathogens, to sit uninterpreted on open databases while sequencers concentrate on reaching 99.9 per cent completeness in projects that typically extend over years.

One could argue that those involved in sequencing cannot have both prompt openness and self-protection and that, once the data are publicly available, it is open season. Others might propose the opposite extreme: that the sequencers have the right even to co-authorship on papers that spring directly from their efforts. That would seem to amount to an extension of publishing practice in biology to recognize the fact that distinct but interdependent roles — in effect, collaboration — are played by those who produce basic information and those who accomplish interpretation.

### Pragmatic solutions

Such relationships are already commonplace in the physical sciences: for example, in astronomy, the developer of a new detector is sometimes, by agreement, included on initial papers that emerge from the application of that device. In high-energy physics, huge collaborations are necessary, and everyone gets due credit by a listing of the hundreds involved. While these particular examples may not translate directly to biology, they are suggestive of a way forward.

That pragmatism seems further than some are prepared to go (see, for example, *Nature* 405, 601; 2000). Some sequencers firmly resist the idea that others can 'cream off' the data they so generously display. Some argue that people are rushing to annotate prematurely, and that waiting for more extensive and accurate sequence coverage is often advisable. It is also true that sequencing is sometimes perceived as scientifically narrow — some senior biologists, in response to media hype about genomics, have emphasized the scientific limitations of a genome sequence, highlighting all the work needed to reveal its meaning. It must be recognized that the scientists involved have dedicated themselves to a systematic project at some expense of their freedom to look more deeply at what they are uncovering. Perhaps not surprisingly, they may resist those whom they perceive to be taking such opportunities away from them.

On the other hand, researchers outside the sequencing collaborations who wish to annotate those data feel not only that the world should not have to wait, but also that the results of annotation should get maximum exposure through publication. Simply adding annotations to a database doesn't do justice to the significance of that work, nor to the need to make that insight available to the biomedical community in an intelligible and convenient form.

There is no simple or generic answer to such issues. Looking at them in a positive light, one can be thankful that there is so much pressure to reveal both the sequence and the function of these genomes as promptly as possible. But more time on annotation by sequencing researchers, and more collaborative goodwill between them and others seeking to use their sequences, would seem to be in order. ■