



**Mapped out**  
The human genome is unveiled in all its glory  
p747



**Crop control**  
Experts call for tighter regulations on transgenics  
p749



**Ray of hope**  
Cut-price anti-AIDS drugs offered to developing world  
p751



**Food for thought**  
Public consortium anticipates early finish of rice genome  
p752

# Publication of human genomes sparks fresh sequence debate



**Declan Butler**

This week's publication of the human genome sequence by both Celera Genomics of Rockville, Maryland, and the publicly funded international

Human Genome Project (HGP) has reignited the debate over the relative merits of the two teams' different strategies.

The two groups published their work simultaneously, as promised last summer, and held a cordial joint press conference in Washington on Monday to advertise the fact. At five more press conferences around the world, participants in the public project celebrated their achievement, which is published in *Nature* (see pages 860–921).

But in the run-up to these meetings, leading members of both teams had been working hard in an attempt to ensure that history — or at least the media — would judge them to have made the more important contribution.

Celera, which embarked on its sequencing effort only three years ago, needs to convince customers who will pay for access to its database that what they are getting is superior to the freely available public data. Members of



Cordial: Celera's Craig Venter (left) and the HGP's Francis Collins at Monday's press conference.

the HGP, having fought off what they regard as an effort by Celera to undermine their project, are now arguing that Celera's assembly, published in *Science* (291, 1304–1351; 2001), could not have been completed without drawing heavily on their own work.

The public project adopted a 'clone-by-clone' approach. In this, the entire genome is chopped into chunks up to several hundred thousand base pairs long, and inserted into synthetic chromosomes known as bacterial artificial chromosomes (BACs).

The key to the HGP's strategy is the subsequent 'mapping' step in which the BACs are each positioned on the genome's chromosomes by looking for distinctive marker sequences, called sequence tagged sites (STSs), whose location has already been pinpointed. In this way, the BACs provide a high-resolution map of the entire genome.

Clones of the BACs are then shattered into tiny fragments in a process known as shotgunning. Each fragment is sequenced and computer algorithms that recognize matching sequence information from overlapping fragments are used to reconstruct the complete sequence inserted into each BAC.

But in 1997, Gene Myers, now vice-president of informatics research with Celera, and James Weber of the Marshfield Medical Research Foundation in Wisconsin argued that the mapping step was unnecessary. They said that algorithms used to reassemble shotgunned DNA fragments could be applied to cloned random fragments taken from the genome as a whole (*Genome Res.* 7, 401–409; ▶

## And now for the proteome...

**Alison Abbott**

In an announcement timed to coincide with the publication of the human genome sequence, a group of top-level proteomics researchers has launched a global Human Proteome Organisation (HUPO).

HUPO's founders see it as a post-genomic analogue of the Human Genome Organisation (HUGO). Its mission will be to increase awareness of, and support for, large-scale protein analysis, in scientific, political and financial circles.

HUGO was created in 1988 by publicly funded researchers who wanted to coordinate global efforts to sequence the genome. Now that the draft human genome sequence has been published, researchers are turning their attention to identifying the functions and expression patterns of the proteins encoded

by the genes. It is generally believed that patterns of protein production — the proteome — will correlate with disease states, which may lead to new treatments.

"Proteins are central to our understanding of cellular function and disease processes, and without a concerted effort in proteomics, the fruits of genomics will go unrealized," says Ian Humphery-Smith of the University of Utrecht in the Netherlands, one of HUPO's founder members.

So far, the embryonic HUPO has created a Global Advisory Council to foster international cooperation, and two regional task-forces in Europe and Japan. An inaugural meeting will take place in the spring to define detailed objectives and to elect a president.

1997). In this 'whole genome shotgun' strategy, fragments are first assembled by algorithms into larger scaffolds. The correct position of these scaffolds on the genome is then worked out using STSs.

Although the whole genome shotgun strategy had been successfully applied to small, simple genomes — such as those of viruses and bacteria — critics argued that it would not work for the human genome, which contains millions of repetitive DNA sequences. The critics expected these repeats to confound the algorithms, making a complete genome assembly impossible.

Nonetheless, Celera was founded with the mission of solving the human genome sequence in short order using a whole genome shotgun approach. The public project's criticisms of the Celera paper have focused on the company's alleged failure to meet this goal.

### Data disputes

Eric Lander, director of the genome centre at the Whitehead Institute for Biomedical Research at the Massachusetts Institute of Technology, asserts that the Celera project would have been unable to find locations for much of its sequence without reference to the public project's genome map.

Celera's paper actually describes two genome assemblies, one put together using the whole genome shotgun approach and a second, 'compartmentalized' assembly. Both were done using Celera sequencing data in which each base of DNA had been sequenced, on average, 5.1 times (5.1X coverage), plus, the paper says, a further 2.9X coverage taken from the HGP's publicly available data.

But members of the public project say that this description is misleading. They argue that the 2.9X coverage is not a random selection of the HGP data. Instead, they say that the data were carefully chosen to cover the entire genome, giving few gaps and retaining maximal mapping information. As a result, they argue, Celera actually obtained the equivalent of 7.5X coverage from the HGP's data.

Lander also notes that Celera's whole genome shotgun sequence contains almost 119,000 scaffolds, rather than the 5,000 that the company had originally predicted. It is impossible to position this many scaffolds on the genome using STS markers, he argues. The majority of Celera's scaffolds are very small, claims Lander, and represent a "tossed genome salad".

Celera's compartmentalized assembly put the genome together region by region, making some use of the HGP's mapping information. In theory, this combination of both projects' sequencing data should produce a better sequence than the HGP data alone. But leaders of the public project argue that there is actually very little in it. "Remarkably, this product is very similar to ours," says John Sulston, former director of Britain's Sanger Centre near Cambridge.



Shooting from the hip: Celera claims its assembly methods have produced a more accurate genome.

"For three years the public project was told that we were inefficient, slow and pointless for proceeding in a careful fashion, and that a whole genome shotgun obviated the need for all these 'wasteful' steps," Lander says. "At the end of the day, it has transpired that we have been the ones who have guaranteed that there is a human genome sequence. We have saved the day." Sulston adds: "They [Celera] failed by their own standards."

Myers dismisses the criticisms from the HGP members as "pure speculation". The whole genome shotgun approach "worked extremely well", he says. "We got tremendous continuity and order and orientation across the genome."

Myers agrees that the method produced a large number of scaffolds in total, but says that more than 90% of the genome sequence is contained in less than 3,000 scaffolds. And he disputes the argument that the contribution from the public project was worth more than 2.9X coverage. "If we had had 3X more Celera sequence data, we could have done it completely on our own," he says.

Originally, Celera had planned to generate its own sequence with 10X coverage. Craig Venter, Celera's president, says that the company decided to make use of the public data, once it became clear that it would be available in time, "instead of spending six months and [another] \$60 million".

### Strategic sequences

Myers points out that the publicly funded mouse genome project is now planning to use whole genome shotgun techniques, in addition to generating a BAC map. Thanks to Celera's efforts, says Myers, strategies for sequencing large genomes have been revised. "I'm proud of being a part of that," he says.

Celera has also responded with a critique of the HGP's sequence. At a press briefing last week, members of the Celera team outlined comparisons of the two papers which, they argued, showed that the company's sequence is more accurate than that produced by the HGP. Myers described analyses of 'mate pairs', which correspond to sequences from either end of an individual cloned DNA segment.

The length of these segments is known. So if mate pairs subsequently end up the wrong distance apart within the final sequence, it suggests a problem with the assembly. Myers pointed out that, for certain chromosomes, the HGP's sequence contains many more of these 'break points' than does Celera's. For chromosome 1, for example, he said, there are "about 35 times more break points in the public assembly".

But Richard Durbin, deputy director of the Sanger Centre, responds that mate-pair analysis is an integral part of Celera's assembly strategy, so one would expect it to make the Celera sequence look good. Nevertheless, Durbin acknowledges that, at present, the detailed ordering in parts of the Celera's sequence may be superior — especially in the case of chromosomes such as chromosome 1, which is still largely in draft form.

The dispute over the success or otherwise of the respective approaches is set to resonate for months, or even years. It will be fanned by the recognition that Nobel prizes may be at stake, and its resolution hampered by the fact that the intricacies of genome assembly are fully understood by only a small community of experts — most of whom have a foot firmly in either the public or the Celera camp.

But Ari Patrinos, head of biological and environmental research at the US Department of Energy, who over the past two years has been a tireless peacemaker between the two rival projects, hopes that publication of the two sequences will lead to a more considered and constructive analysis of the different methodologies by competent experts.

To aid this process, Patrinos intends to organize a joint Celera-HGP workshop, probably to be held in Washington on 3 April. The plan is for the meeting to be chaired jointly by Myers and David Haussler, an expert in computational biology at the University of California at Santa Cruz. Its goal, says Patrinos, is to establish "what can we learn from the experiences of both sides, and what in the future should be the optimum approach to sequencing mammalian genomes". ■

Additional reporting by Peter Aldhous in London and Colin Macilwain in Washington.