

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

## Refining the genome

ArticleInfo		
ArticleID	:	5011
ArticleDOI	:	10.1186/gb-spotlight-20041022-01
ArticleCitationID	:	spotlight-20041022-01
ArticleSequenceNumber	:	74
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	3
ArticleHistory	:	RegistrationDate : 2004-10-22 OnlineDate : 2004-10-22
ArticleCopyright	:	BioMed Central Ltd2004
ArticleGrants	:	
ArticleContext	:	130595511

**Cathy Holding**

**Email:** [cathy.holding@absw.org.uk](mailto:cathy.holding@absw.org.uk)

---

A comparison of human genome sequences produced using different approaches by the International Human Genome Sequencing Consortium and the private firm Celera reveals that whole genome shotgun sequencing (WGS), as used by [Celera](#), fails miserably on genome segmental duplications.

Writing in [Nature](#) this week, [Evan Eichler](#), from Washington University, and colleagues show that WGS does very well for 95% of euchromatin, but falls down in the very large duplications that are greater than 98% identical and longer than 20 kb in size. "It's not that bad unless you really care about those large identical duplications," he told us. The technique "fails and fails miserably" in that regard, he said.

For the current study, Eichler teamed up with previous collaborators Granger G. Sutton and Aaron L. Halpern, both now at the [J. Craig Venter Institute](#) in Rockville, Md., in an apparent revisiting of results [published](#) by *PNAS* in February this year. They conclude that breaking up the sequencing of whole genomes into two phases - WGS followed by clone-order-based sequencing in bacterial artificial chromosomes - is the way forward.

Eichler told us he had wanted to return to the comparison because his work for the first paper was completed in just 2 weeks and had been reduced to three sentences of actual data in the *PNAS* paper. "At that point, I thought that I would do a much more rigorous analysis; that I'm going to broadcast this on my own," he said.

Eichler and Sutton both felt Celera hadn't published in more detail because of *PNAS* page constraints. But, said Halpern, "it wasn't just page constraints that limited us back in February - we just didn't have the results at the time."

Eichler had only had time to do a preliminary analysis in the first place, Sutton and Halpern said. "So, for instance, in this [latest] one, the discussion is about the effect of the length of the repeat and the percent identity - those are entirely new to this paper," Halpern said.

"The take-home message was that strict application of WGS is going to miss [segmental duplication] regions of the genome," Eichler added. "In fact, it's one of those things that without a complete finished sequence of the human genome, we would not understand the architecture of these regions."

Those sentiments are echoed [elsewhere](#) in *Nature* this week, where [Francis Collins](#), Adam Felsenfeld, and colleagues of the International Human Genome Sequencing Consortium report the final finishing of the human genome sequence.

"If we want to finish the genome sequences for other organisms besides the human, one cannot just count on the shotgun method to do that correctly, at least not in its current form," Collins, director of the National Human Genome Research Institute (NHGRI), told us.

"That's a lesson that we had pretty much resigned ourselves to," Collins said. "Now that you have a really final beautifully finished sequence, you can compare it to that of other organisms, particularly the mouse, the rat, and the chimpanzee," Collins said.

Still, it is the paucity of genes in the human genome - a figure revised downward to fewer than 25,000 - that continues to "blow our socks off," according to Collins. It seems like an awfully short list to account for the biological properties of a human being, he said.

Felsenfeld, also at NHGRI, coordinated the teams involved in the sequencing project. He said the main surprises had only become apparent with the passage of time. "One of them is the real inhomogeneity of the genome. The most obvious manifestation of that is there is some small number of regions that we just can't sequence," he said.

"From here on out, it will require individual ingenuity to try and close any of those remaining gaps," said Collins. But, he noted, there are probably few, if any, genes left in those unsequenceable regions.

## References

1. Holding C: Celera defends human sequence *Genome Biology* February 20, 2004., [[http://genomebiology.com/researchnews/default.asp?arx\\_id=gb-spotlight-20040220-01](http://genomebiology.com/researchnews/default.asp?arx_id=gb-spotlight-20040220-01)]
2. *Nature*, [<http://www.nature.com/>]
3. Evan Eichler, [<http://www.gs.washington.edu/faculty/eichler.htm>]
4. J. Craig Venter Institute, [<http://www.venterinstitution.org/>]
5. Whole-genome shotgun assembly and comparison of human genome assemblies
6. Finishing the euchromatic sequence of the human genome
7. Francis S. Collins, [<http://www.genome.gov/10000779>]