

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

New bugs in the human genome draft

ArticleInfo		
ArticleID	:	3894
ArticleDOI	:	10.1186/gb-2001-2-11-reports0041
ArticleCitationID	:	reports0041
ArticleSequenceNumber	:	19
ArticleCategory	:	Paper report
ArticleFirstPage	:	1
ArticleLastPage	:	3
ArticleHistory	:	RegistrationDate : 2001-9-11 Received : 2001-9-11 OnlineDate : 2001-10-26
ArticleCopyright	:	BioMed Central Ltd2001
ArticleGrants	:	

Reiner Veitia

Abstract

A critical evaluation of the quality of different versions of the draft human genome sequence has revealed discrepancies between experimental mapping results and data available online.

Significance and context

The release of the draft of the human genome sequence can be considered as one of the major breakthroughs in biology of the past few years. Finishing the sequence, and above all its full annotation, will still take some time. Although the importance of the draft human genome is undeniable, there may be inaccuracies in the published information, an issue of major importance for geneticists working on positional cloning projects that rely on fine-mapping data to localize genes responsible for human diseases.

Key results

Katsanis *et al.* assembled 925 expressed sequence tag (EST) clusters to evaluate the draft genome sequence coverage and mapping fidelity. To estimate the degree of completion of the draft they assessed EST sequence representation by comparing ESTs, using BLAST (basic alignment search tool), with several versions of the draft released last year. To avoid problems of conflicting mapping due to chimerism, they used only one 3'-end EST sequence per EST cluster. Statistically significant differences were found between the observed and expected numbers of ESTs that matched the genomic sequence. As ESTs were represented less than expected, the authors propose that the redundancy of the draft may be greater than expected, that the sequence available is biased towards gene-poor regions, that the size of the genome is greater than current estimates, and/or that a component of the genome is repeated.

To ascertain mapping accuracy of the segments of the draft, 138 ESTs (from different clusters out of the 925) were mapped by PCR using radiation and monochromosomal hybrid panels. In 137 cases the location of the EST coincided with its annotated location (in the EST data sheet). When comparing the experimental locations to the annotated locations of bacterial artificial chromosomes (BACs) containing most of the mapped ESTs, however, about one-third of the positions were discordant.

The authors noticed a modest improvement in mapping accuracy in subsequent releases of the draft. It seems, however, that this is due to the higher quality of the new input sequence and not to correction of the previous versions. Analysis of the sequence recently assembled into scaffolds (March and April 2001) has shown that mapping accuracy has improved by a factor of two. The authors noticed, however, that in some instances single-copy ESTs were represented several times within the assembled segment, suggesting artifactual 'electronic' duplications. As a corollary, the authors suggest that it would be better to finish the sequence of the human genome before generating drafts for other mammals.

Reporter's comments

The results reported, although extremely simple, are important in the context of the human genome project. In fact, the main role of this kind of sampling work (as an isolated effort) is not to help correct the draft but to attract the attention of the genetics community to possible inaccuracies in the published data. At the same time, it points out the need for more cross-talk between the annotation of EST data and of genomic sequence. It is clear that mapping data for ESTs is much more accurate than for the genomic segments of the draft. In addition, for each EST cluster, mapping data are redundant, providing a statistical 'cartographical' consensus. Greater efforts should therefore be made in the annotation of the draft, to take into account EST information, which may help save time and money.

Table of links

[Nature Genetics](#)

References

1. Katsanis N, Worley KC, Lupski JR: An evaluation of the draft human genome sequence. *Nat Genet.* 2001, 29: 88-91. 1061-4036