NEWS AND VIEWS

Wrong Sequences in Databases: Whose Fault??

Devi Lal · Rup Lal

With the development of newer sequencing techniques the databases have been flooded with enormous amount of sequencing data. This saw emergence of yet another field in biology- "bioinformatics" which is principally based on sequence information from DNA, RNA or protein molecules. Sequence databases provide massive data-sets that serve as basic raw material for any bioinformatics research. Novel biological finding are driven by comparison of huge data-sets [1], therefore a bioinformatician exploits these databases for such data-sets and comes up with logical interpretations. However, an increase in the flow of sequence data in the databases has in turn raised issues concerning the accuracy of the submitted data. It has been seen that the sequence data submitted to the databases are not free from errors [2] and the use of such data can lead to erroneous interpretations and even illogical conclusions.

The errors in any sequence can erupt from two sources: (i) the sequence deposited in the database itself can be wrong and (ii) there can be error in the annotations. While it is difficult to check the flow of wrong sequences into the databases, some measures need to be taken to establish the validity of the sequences before public release. It has been suggested that this can be improved by including the third party annotation system [3]. The onus should also be on individuals to submit only valid sequence data for subsequent use by others. The current era belongs to the modern sequencing techniques and bioinformatics. Therefore valid sequence data will provide additional bonus and will subsequently help in making bioinformatics study a logical field in biology.

References

1. Jensen LJ, Bork P (2010) Ontologies in quantitative biology: a basis for comparison, integration, and discovery. PLoS Biol 8:e1000374
2. Harris DJ (2003) Can you bank on GenBank? Trends Ecol Evol 18:317–319
3. Bidartondo M (2008) Preserving accuracy in GenBank. Science 319:1616

D. Lal · R. Lal (✉)
Department of Zoology, University of Delhi,
Delhi 110007, India
e-mail: ruplal@gmail.com