

On the Multifaceted Aspects of Bioinformatics in the Next Generation Era: The Run that must keep the Quality

Maria Luisa Chiusano*

Department of Agricultural Sciences, University of Naples Federico II, via University, Portici (NA), Italy

Editorial

The well-established role of bioinformatics in support of structural and functional genomics does not need extensive presentation, nowadays. This multidisciplinary research area, evolving preliminarily for biological data management and distribution, after the initial delivery of nucleic acid and protein structure, was soon revealed to be also appropriate in providing tools for data analysis, integration and modeling. Indeed, bioinformatics is suitable not only for detecting value added information from omics data, but also for paving the way for the prediction of emergent properties from holistic investigations of massive, heterogeneous collections.

Methods in bioinformatics are required to face the challenges due to data complexity. This is associated not only to dimensionality, but also to the nature of the represented molecules, even when basic efforts, such as data organization or string comparisons, are considered. Bioinformatics also provides the most appropriate approaches in biology in support of comprehensive data views. Indeed, useful information can be derived from the integration of heterogeneous collections concerning specific levels (eg: transcriptome analyses based on technologies endowed with different specificity and sensitivity) or varied levels of cell functionality (eg: epigenomics, transcriptomics, proteomics, metabolomics).

The spreading of Next Generation Sequencing (NGS) technologies did not change this scenario in terms of achievable information. What strongly had impact on biology and associated research is indeed the incredible amount of data these technologies provide, favoring unexpected resolutions, though further challenging bioinformatics. On the other hand, the growth of interest that cost-affordable, fast and accessible experimental technologies raised in all the scientific community was even more striking. Two main effects were determined: an incredible number of collections based on NGS and an unexpected increase in computational demand. This also enhanced the need for suitable strategies for centralized data maintenance and for specialized, user friendly, stabilized tools and resources for data analysis and integration. As a direct consequence, this caused a major revolution in bioinformatics.

The aim here is to highlight the requirements that bioinformatics should fulfill in this framework. Major trends in biology are becoming evident. High throughput scientific challenges, mainly aiming to solve the primary structure of genomes from different species, genotypes, single cells, in diverse biological fields, including medicine, agriculture and ecology, are being supported by leading international consortiums and/or fully equipped sequencing centers accompanied by highly specialized and "service dedicated" bioinformatics facilities. The results of these efforts are being newly sequenced genomes or their scaffolds, endowed with a preliminary gene annotation that, though still draft, are often published in high impact journals. However, the high production rate is not always permitting to reach adequate quality standards and, often, data are not immediately profitable to support associated analysis (like RNAseq) and or "non bioinformatics expert" users. Specifically, bioinformatics tools and omics resources are being produced too fast, though still hardly following the evolution of technologies. This

affects the offer of stable, comparable, integrated reference resources. As an example, frequent updates of genome sequence data, as well as the constant release of improved annotations, accompanied by novel RNAseq collections, are associated with the flourishing of heterogeneous databases, causing difficulty in fixing common, comparable, good quality and well-tested references in terms of software and datasets. The effect is heterogeneity in similar resources, like those presenting different gene annotation versions of the same genome or misalignments with collections from reference datasets. A recurring problem is also the flourishing of community specific databases not associated with information from worldwide accepted reference resources and therefore not easily accessible by all scientists. This may perilously represent a step backward in a field that owes its success to the sharing and the distribution of biological resources worldwide. To bridge the opening gap between "advanced" and "end users" bioinformatics, two major efforts are required. It is indeed necessary to move efficiently from draft to curated resources, to make results exploitable and the end user efforts appropriately supported to make their efforts reasonable and coherent. The need of a "tailoring" bioinformatics for dedicated efforts that could improve the quality of draft results and make them accessible, usable and reconciled with related resources, is fundamental to reducing the risk that fast data production could compromise the sharing of reliable and updated information. Project reviewers and publication policies should contribute a relevant role in facing these issues, since they can drive towards appropriate cross referencing, data sharing and high quality standards in both software and resources. In a world where data production rate is dramatically increasing, the primary goal of making the data usable and exploitable should not be passed over since it is required and compulsory in science. We should keep focused on favoring and prioritizing the criteria that have played a major role in the scientific revolution we are now experiencing.

***Corresponding author:** Maria Luisa Chiusano, Department of Agricultural Sciences, University of Naples Federico II, via University, Portici (NA), Italy, Tel: +39 081 253 11; E-mail: chiusano@unina.it

Received November 16, 2015; **Accepted** November 18, 2015; **Published** November 20, 2015

Citation: Chiusano ML (2015) On the Multifaceted Aspects of Bioinformatics in the Next Generation Era: The Run that must keep the Quality. Next Generat Sequenc & Applic 2: e106. doi:10.4172/2469-9853.1000e106

Copyright: © 2015 Chiusano ML. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.