

Editorial

Bioinformatics through the history

The term “Bioinformatics” was first used by Paulien Hogeweg and Ben Hesper in the beginning of 1970s, defining it as “the study of informatic processes in biotic systems”. Although they have proposed the definition as above in article in Dutch language that is not generally accessible [1] but various public sources trace the origin of the term to publications by Paulien Hogeweg and Ben Hesper that appeared in 1978 [2, 3]. Their main aim was to combine pattern analysis and dynamic modeling and apply them to the challenge of unraveling pattern generation and informatic processes in biotic systems at multiple scales but now a days meaning of the term has been superseded as denoting the development and use of computational methods for comparative analysis of genome data. The long term goals they set for bioinformatics in the 1970s, were termed by them as the “horse part” and the “elephant”. The horse part is the “modeling morphogenesis, through the use of cell based models that incorporate some of the physical properties of cells [4]. Second but important part i.e. the elephant is “Constructive models of evolution”, are generally being created from studies on the evolutionary consequences of non-linear physical mapping includes both genotype and phenotype mapping [5-8]. Metabolic networks [9-10], regulatory networks [11-14] and chromosome organization [15-17] are also providing useful information in the above model construction.

Many of the basic pattern analysis methods which are now being used in bioinformatics very often, were pioneered in the 1960s and further developed in the 1970s. A very notable work was of Margaret Dayhoff, who developed the first ever biological database known as “atlas of protein sequences and structure” [18]. It was built as a collection of sequences for investigating evolutionary relationships among proteins. In the 1970s and 1980s, novel modeling formalisms were actively explored and developed along with the development of pattern analysis methods.

With help of the development in the *in-silico* technologies and wetlab methods sequence data was generated exponentially in the late 1980s and 1990s. It was the time when the term bioinformatics became mainstream, coming to mean the development and use of computational methods for data management and data analysis of sequence data, protein structure determination, homology-based function prediction, and phylogeny. But the information which is provided by the massive sequencing projects, and the related bioinformatics analysis to unravel the relationship in between function and evolution, is not really the “roots of bioinformatics”, but rather they are the “trunk of bioinformatics” [19],

Currently we are having a large set of fully sequenced genomes which also include human and it is expanding at very fast rate. This is providing a massive high-

throughput “omics” data, which is further available for comparative research and is presenting great challenge for bioinformatics. Recently, 2012 Nobel laureate Paul Nurse has emphasized about self-organization and the resulting counter-intuitive results, he argues that the next “quantum leap” in biology will come through studying information processing in biological systems. Also similarly Walter Gilbert another Nobel laureate has said-Most of the biological investigations in 21st century will be *in silico*.

It is apparent that during the last 5 years, we have moved into a new stage, that can define our future strategy. By 2007, things had become more intelligent: text mining were used for decision making [20], ontology growth was manifold into every aspect of computing [21], and bioinformatics was distributive in the life sciences, for example, extending to biodiversity conservation planning [22] or synthetic biology [23]. Besides the more theoretical aspects of network biology [24], exemplified by gene and protein interaction networks, pressure mounted for support of translational medicine, ranging from structural variation [25] to cancer bioinformatics [26]. At the same time, new problems were emerging, related to next-generation sequencing efforts, ranging from re-sequencing to metagenomics [27]. The prediction in 2008 was that in 10 years, we will possess an adequate infrastructure for biological research [28], in a fusion of disciplines [29]. At present, we are facing an expansion of difficulty, ranging from genome assembly [30], protein design [31], or metagenomics [32] to genomic medicine [33], infectious disease [34], and phenotyping [35].

It can be concluded that whether bioinformatics in the wider sense of studying information processing in biotic systems is a quirk or a quantum leap, it is still an interesting area to understand and work in! There is lots and lots of data and software available and lot more need to be developed. More genomes, more database, more tools and analysis of all this available data and biology, physics and mathematics behind all this, which motivated us to start this new initiative to serve humanity in the coming decades and centuries.

Dr. Gulshan Wadhwa

Department of Biotechnology, Ministry of Science & Technology,
Government of India.
New Delhi, India

References

1. Hesper B, Hogeweg P (1970) Bioinformatica: een werkconcept. *Kameleon* 1(6): 28–29. (In Dutch.) Leiden: Leidse Biologen Club.
2. Hogeweg P, Hesper B (1978) Interactive instruction on population interactions. *Comput Biol Med* **8**: 319–327.
3. Hogeweg P (1978) Simulating the growth of cellular forms. *Simulation* **31**: 90–96.
4. Anderson A, Chaplain M, Rejniak K, Fozard J (2008) Single-cell-based models in biology and medicine. *Basel*: Birkhauser Verlag.
5. Schuster P, Fontana W, Stadler P, Hofacker I (1994) From sequences to shapes and back: a case study in RNA secondary structures. *Proc Biol Sci* **255**: 279–284.
6. Huynen M, Stadler P, Fontana W (1996) Smoothness within ruggedness: the role of neutrality in adaptation. *Proc Natl Acad Sci U S A* **93**: 397–401.

7. van Nimwegen E, Crutchfield J, Huynen M (1999) Neutral evolution of mutational robustness. *Proc Natl Acad Sci U S A* **96**: 9716–9720.
8. Huynen M (1996) Exploring phenotype space through neutral evolution. *J Mol Evol* **43**: 165–169.
9. Kacser H, Beeby R (1984) Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection. *J Mol Evol* **20**: 38–51.
10. Soyer O, Pfeiffer T (2010) Evolution under fluctuating environments explains observed robustness in metabolic networks. *PLoS Comput Biol* **6**: e1000907. doi:10.1371/journal.Pcbi.1000907.
11. Crombach A, Hogeweg P (2008) Evolution of evolvability in gene regulatory networks. *PLoS Comput Biol* **4**: e1000112. doi:10.1371/journal.Pcbi.1000112.
12. Draghi J, Wagner G (2009) The evolutionary dynamics of evolvability in a gene network model. *J Evol Biol* **22**: 599–611.
13. Wagner A (2008) Robustness and evolvability: a paradox resolved. *Proc Biol Sci* **275**: 91–100.
14. Draghi J, Parsons T, Wagner G, Plotkin J (2010) Mutational robustness can facilitate adaptation. *Nature* **463**: 353–355.
15. Crombach A, Hogeweg P (2007) Chromosome rearrangements and the evolution of genome structuring and adaptability. *Mol Biol Evol* **24**: 1130–1139.
16. Hurst L, Pa'l C, Lercher M (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* **5**: 299–310.
17. Batada N, Hurst L (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* **39**: 945–949.
18. Dayhoff, M. (1969) Atlas of Protein Sequence and Structure 1969. **vol. 4**, National Biomedical Research Foundation, Silver Spring, Maryland.
19. Hogeweg P (2011) The Roots of Bioinformatics in Theoretical Biology. *PLoS Comput Biol* **7(3)**: e1002021.
20. Perez-Iratxeta C, Andrade-Navarro MA, Wren JD (2007) Evolving research trends in bioinformatics. *Brief Bioinform* **8**: 88–95.
21. Brewster C, O'Hara K (2007) Knowledge representation with ontologies: present challenges—future possibilities. *International Journal of Human- Computer Studies* **65**: 563–568.
22. Faith DP, Baker AM (2007) Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evolutionary bioinformatics* 2006.
23. Serrano L (2007) Synthetic biology: promises and challenges. *Mol Syst Biol* **3**: 158.
24. Gatenby RA, Frieden BR (2007) Information theory in living systems, methods, applications, and challenges. *Bull Math Biol* **69**: 635–657.
25. Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, *et al.* (2007) Challenges and standards in integrating surveys of structural variation. *Nat Genet* **39**: S7–15.
26. Hanauer DA, Rhodes DR, Sinha-Kumar C, Chinnaiyan AM (2007) Bioinformatics approaches in the study of cancer. *Curr Mol Med* **7**: 133–141.
27. Pop M, Salzberg SL (2008) Bioinformatics challenges of new sequencing technology. *Trends Genet* **24**: 142–149.
28. Stein LD (2008) Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. *Nat Rev Genet* **9**: 678–688.
29. Smith TF (2008) Computational biology: its challenges past, present, and future. Proceedings of the 12th annual international conference on research in computational molecular biology. Singapore: *Springer-Verlag*. Pp 1–2.
30. Pop M (2009) Genome assembly reborn: recent computational challenges. *Brief Bioinform* **10**: 354–366.

31. Suarez M, Jaramillo A (2009) Challenges in the computational design of proteins. *J R Soc Interface* **6 Suppl 4**: S477–491.
32. Kyrpides NC (2009) Fifteen years of microbial genomics: meeting the challenges and fulfilling the dream. *Nat Biotechnol* **27**: 627–632.
33. Auffray C, Chen Z, Hood L (2009) Systems medicine: the future of medical genomics and healthcare. *Genome Med* **1**: 2.
34. Berglund EC, Nystedt B, Andersson SG (2009) Computational resources in infectious disease: limitations and challenges. *PLoS Comput Biol* **5**: e1000481.
35. Thorisson GA, Muilu J, Brookes AJ (2009) Genotype-phenotype databases: challenges and solutions for the post-genomic era. *Nat Rev Genet* **10**: 9–18.