Capacities vs. jobs in bioinformatics and biotechnology: a few points to the attention of current students & job-seekers

Author:
Kshitish Acharya K, *PhD juneindia@gmail.com*

Author information: A total 22 yrs of research experience with about 6 yrs in industry, 4 yrs in USA. At IBAB (www.ibab.ac.in) for 11 yrs & also the founder director of Shodhaka (www.shodhaka.com) since 2009. Conceived & organized a unique & successful 14-months 'laboratory course in bio-techniques' (about 95% placement record in 8 batches). Experience in screening more than 500 candidates for several positions; he has also placed students in various organizations. Gave several talks, wrote articles & conducted workshops on topics related to HR development and career planning.

Background:

Bioinformatics & biotechnology are multidisciplinary in nature. Special methodologies are characteristic of both domains. Activities in these areas have existed since several decades. In the recent past, however, bioinformatics & biotechnology have emerged as a major subject of study, particularly in graduate (BSc, BE/BTech) & post-graduate (MSc, ME/MTech) programs. The true value of these subjects was masked by the 'hype' created around the time of reporting of the first human genomic projects. Bioinformatics & molecular biology are essential for the basic & applied biology*. Even though many had unrealistic expectations in terms of jobs/career, the actual 'scope' in these areas is likely to be there as long as human-life persists.

Like other sectors, there are more degree/diploma holders in life sciences, but real employable youngsters are less. This is mainly because, the quality of related courses at many educational organizations in India has not been of good standards. Hence, there is a largely unmet need for graduates & post-graduates with 'real' capacities. If the students are careful in selecting the organization for their study, they are more likely to find good jobs &, more importantly, grow quickly in their career.

This article intends to provide a quick review of various aspects related to capacity development & career opportunities.

What is my strength vs. weakness, & what type of job suits me?

Finding an answer to such questions requires constant study of options out there as well as self-assessment. There is a need to identify the strengths & weaknesses across subjects and find out the type of job one might suit most. This can be done with the help of the nation-wide online job-suitability-assessment system (SOTS-JSA; www.shodhaka.com/sotsjsa). The profiles of students/test-takers will also be made available for potential employers (companies & academic groups) so that they can be screened for job-interviews.

Who can study these subjects?

Bioinformatics and biotechnology courses need to include substantial amount of teaching in biochemistry, cell biology & molecular biology. However, it is a misconception that bioinformatics suits biologists better than other science students. Core expertises (see next sections & table 1) in bioinformatics require knowledge & skills from other science & technology areas. Hence, biologists* & non-biologists perform equally well; many times, non-biologists do better than biologists in bioinformatics (observation by the author during his 11 years of teaching). It is sad that many students from statistics, physics, mathematics, computer-science & chemistry streams are not aware of the opportunities in bioinformatics.

User-end bioinformatics (UEB), molecular biology & core bioinformatics:

While plenty of articles/books[@] have elaborated on type of biotechnology activities, bioinformatics domain has not been explained well that much. It is convenient to consider a bioinformatician as someone with a specialization of one or more of the following expertises (also see the table 1, below for more specific activities & corresponding bioinformatics capacities):

A. Creating & managing databases as well as software tools. These types of activities are best carried out by bioinformaticians with specific expertises. But a group can pool multiple types of expertise to develop databases and software.

The expertise required in understanding biological experiments & the research process are particularly high when the database creation involves data-curation (also called biocuration[#]). Curated data can be inputs for databases[2], but biocurator(s) need not know much about databases or programming.

B. Developing new algorithms or computational methods for bioinformatics. Development of a rapid method to compare a sequence with millions of other sequences[3] is a good example of such activity. Other examples include developing new methods of predicting genes, specialized sequence analysis, genome-wide or system-wide analysis of molecules & their interactions & rapid analysis of large scale data-sets from novel methods such as next generation sequencing (NGS) & mass spectroscopy.

C. Biological research using the knowledge & skills of computational, biological & other science subjects. Common bioinformatics research areas include computational structural biology[4], cheminformatics[5] & mathematical modelling of molecular networks (e.g., pathways or all or most interactions within a complete cell[6]. Drug designing[7] involves computational structural biology. There are also many bioinformatics approaches employed by scientists in basic or applied research.

The above-listed expertises are often integrated; for example, a new algorithm may be used to develop new software, which is integrated into a database[2]. And, the actual purposed of all this could be to eventually address a research-problem such as understanding the regulation of gene expression. In fact, the next type of activity is also never a stand-alone expertise.

D. UEB is mainly about accessing & analyzing biological data. These activities require a general understanding of the process of research in life sciences & acquaintance with at least the most commonly used databases & software (there are too many of these resources; see www.startbioinfo.com). Despite being the simplest & most widely required type of activity, this perhaps is often taken too lightly in many of our teaching programs, including MPhil & course-works for PhD, across biological domains. Most biology[*] students seem to lack the ability to pick the apt database or software for different type of research problems, & interpret the output of these resources. Analysis of mass-scale data is particularly crucial for genomics, transcriptomics & proteomics - which pervade many biological problem-solving cases in the recent years.

Biology[*] students/researchers will benefit better if they have at least a broad understanding of the other core bioinformatics types, in addition to UEB. Similarly, bioinformatics of any type would be more complete if the students know the basic biotechniques & understand how the data/information is obtained in molecular biology. Many have stressed the need of incorporating bioinformatics in different subjects at different levels[8]. Actually, UEB and training for molecular biology techniques should be an essential part of any bio-course, but often neglected.

Degrees & capacities:

As with many other subjects, though there are lots of degree/diploma holders in the subject, the number of good quality employable youngsters seems to be low in the recent years. It is important to realize the significance of type of capacities (domain knowledge & skills, see table 1) & the required level of expertise. Unfortunately, the quality of bioinformatics courses at many educational organizations in India has not been of good standards. Students should not be carried away by the external glitter/glamour of organizations in India/abroad. In fact, many students who have done MS-abroad also may not be really learning much. Trying an MS degree abroad is a good idea only if the University/Institute is a reputed one & a scholarship is offered by the same.

One way to find the right place for higher education is to look for departments or specialized institutes with good research publications in peer-reviewed journals having good impact factor (IF**). This is because, a good research work requires good facilities & knowledge, & they are also the key aspects for good learning. This is true for most higher-studies/courses in any science subject. Students can use pubmed## to get publications from a specific institute & names of the faculty members (using 'affiliation' & 'author' fields while searching). Students should be also cautious about low-quality-journals; they accept any article submitted to them, & colleges publishing in such journals are not perceived to be doing good quality research.

Obtaining PhD, & often a post doctoral research experience (usually 2-4 years), is a must if one intends to lead research project(s). The extent of independence in choosing research topic by a scientist is often severely limited when working in companies (where salaries can be higher), while in most academic organizations it is compounded with multiple other responsibilities such as teaching & other associated responsibilities. Many other allied jobs (such as teaching, technical support, marketing, scientific-unit-management, IPR management etc) may also sometimes require a doctorate degree. Fortunately, several government-schemes offer financial support for bright candidates while studying PhD. Mere doctorate degree & post doctoral experience will not help. It is also important to have authorship in good research publications in at least moderate IF-journals. In other words, one has to attain best capacities for research, with proof.

Table 1. A broad classification of job-responsibilities & corresponding type of capacities required for them.

Serial no.	Job-responsibility	Capacities
1.	Creating &/or maintaining software for biologists	Independent programming in two or more languages (Perl, Bioperl, Python, Java, R, PHP, C, C++, html); basic operations in linux; good knowledge in mathematics & statistics.
2.	Creating &/or maintaining databases	Command over MySQL, Oracle & decent/moderate capacities as mentioned in row 1 will also be required.
3.	Creating molecular networks & mathematically modelling or simulating pathways or cells as a system (systems biology)	Same capacities mentioned above, in row 1, will help. Moderate to intense knowledge in biochemistry, particularly enzymology, & high-throughput data analysis will be required too. Depth of knowledge required in mathematics and statistics is even higher.
4.	Studying structures of macromolecules (such as DNA or proteins) &/or their interactions, via computers	Clear knowledge in at least basic chemistry & physics, & the capacities mentioned in row 1.
5.	Studying structures & chemical properties of small molecules (cheminformatics or chemoinformatics)	Almost same capacities as mentioned in the above row. But advanced knowledge of chemistry, and some understanding of pharmacology, will help.
6.	Data analysis	Ability to select appropriate databases & software from a biologist's perspective, use them & interpret the results. Moderate to intense knowledge in cell & molecular biology are often essential. Linux operations

		and familiarity with programming can help.
7.	Biocuration	A good understanding of the research process, particularly the relationship between biotechniques, their applications & data interpretations will be needed. Experience in reading research papers. Moderate to intense knowledge in cell & molecular biology are often essential. Command over English language & ability to quickly understand research papers will help.
8	Data entry & management	A basic understanding of biological concepts & databases. Training in a specialized domain of requirement is likely to be provided by the employer.
9.	Clinical data analysis & management	Multiple types of jobs that require medical &/or other biological domain knowledge. Some of the types of responsibilities need moderate to very reliable expertise in programming &/or statistics.

Important notes:

a) Multiple job-handling & multiple capacities are always better, but not at the cost of at least one type of reliable expertise.
b) Clear conceptual understanding of fundamentals biochemistry, genetics, cell biology, molecular biology, physiology, microbiology & immunology are essential for bioinformatics & biotechnology.
c) At least a moderate & independent programming capacity is needed for most bioinformatics jobs.

Where do I find jobs?

There have been many fluctuations in bioinformatics & biotech employers - globally. It should be noted that the rapid changes in bio-technologies (such as microarray & NGS), focus on biosimilars (macromolecules used for disease-treatments) etc have also influenced the rise & fall of companies, departments &/or their strategies.

While there are many biotech-companies of various nature, there are fewer companies in India that can be identified with mainly bioinformatics activities. The article can only give examples. Strand Life Sciences (http://www.strandls.com) is known for production of commercial software useful for data analysis. Biobase (http://www.biobase-international.com), a multinational company - with a branch in Bengaluru & 'Molecular Connections' (www.molecularconnections.com) have been active in biocuration & programming domains. Several small companies have started recently. For example, Shodhaka Life Sciences Pvt. Ltd. (www.shodhaka.com) is active in the areas of service providing (data analysis) for other scientists as well as training.

But many jobs are available outside specialized organizations as most biological research today requires bioinformatics &/or molecular biology expertise. Such requirements frequently occur in R&D centres of biotech & related companies as well as in academic research groups (listed elsewhere in the book).

Summary statements

Bioinformaticians & biotechnologists have roles to play in different types of companies & academic research divisions. Hence, even if the number of pure bioinformatics & biotechnology companies does not increase considerably in India or abroad in near future, bioinformaticians & biotechnologists will most probably continue to be in demand. However, the type of expertise required will be different depending on the type of the organization &/or responsibilities of the job. From the point of learning & capacity-building, it is essential to have an excellent conceptual understanding of biology subjects (e.g., biochemistry, genetics, molecular biology), & skills for bio-techniques & UEB.

By the time one completes the education (BE, BTech, MSc, MTech) he/she should be able to take up any responsibility, or to at least start executing the assigned tasks, with minimum help/guidance. To attain the right capacities one has to take apt courses. If main courses fall short for any reasons, additional short courses. Internships/project-works can be extremely useful, even after completion of formal courses. The real expertise comes only with years of work. Most importantly, making the best use of all such learning opportunities can help to be a ‘beginner with reasonable capacity & confidence’.

REFERENCES / NOTES:

*Biology &/or life sciences, in this article, refers to any/all of these: biotechnology, molecular biology, genetics, bioinformatics, microbiology, biochemistry, zoology, botany, biotechnology, biomedical engineering, agriculture, veterinary science, pharmacology, health sciences & similar/related subjects.

@ FURTHER READING:

- Career opportunities in biotechnology and drug development, Toby Freedman, Cold Spring Harbor Laboratory Press, 2008.
- Career development in bioengineering and biotechnology, G. Madhavan, B. Oakley & L. Kun, Springer, 2008.

** IF is a crude measure of utility or popularity of published articles; IF of most reputed journals ranges from approximately 0.9 (e.g., Current Science) to 32 (Cell).

#Biocuration is the process of capturing the gist of sections from research papers, about experimental results & conclusions; it might also involve assessing &/or interpreting the reliability of scientific data in the context of specific biological events.

##www.ncbi.nlm.nih.gov/pubmed

EXAMPLE-ARTICLES:

2. A novel tissue-specific meta-analysis approach for gene expression predictions, initiated with a mammalian gene expression testis database. PMID: 20699007, http://www.biomedcentral.com/1471-2164/11/467.
3. Basic local alignment search tool. http://www.ncbi.nlm.nih.gov/pubmed/2231712.
4. Designing artificial enzymes by intuition & computation. PMID: 21124375; http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3443871.
5. Computational databases, pathway & cheminformatics tools for tuberculosis drug discovery. PMID: 21129975; http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3034835
6. When ubiquitination meets phosphorylation: a systems biology perspective of EGFR/MAPK signalling. PMID: 23902637; http://www.biosignaling.com/content/11/1/52.
7. Structure-based ligand design & the promise held for antiprotozoan drug discovery. PMID: 19103598; http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2673241.
8. Educating biologists in the 21st century: bioinformatics scientists versus bioinformatics technicians. Bioinformatics 20 (14) p. 2159-2161. PMID: 15073013 http://bioinformatics.oxfordjournals.org/content/20/14/2159.full.pdf+html?sid=1f9f8fd1-ffeb-4762-b2b4-1d041eeaaa3f