

Characteristics of bioinformatics employment advertisements

Jennifer I. Hill, W. John MacMullen*, and Carole Palmer

Graduate School of Library & Information Science, University of Illinois at Urbana-Champaign, 501 E. Daniel St., MC-493, Champaign, IL 61820, USA. Email: wjohn@illinois.edu

*corresponding author

Current trends in the bioinformatics job market were assessed using a sample of 1,996 online employment advertisements from the 6-year period of January 2003 through December 2008. Job postings were classified by employer type, job role, and location, and a content analysis of the text of a subset of 404 of the posts was performed to identify detailed characteristics associated with master's degree level positions, such as educational requirements and preferred scientific and technical skills. Consistent with previous studies, academic institutions, corporations, and research institutes are the primary employers of bioinformaticists. In the U.S. only three states, California, Maryland, and Massachusetts, provided the majority of opportunities. Graduates from all levels of education are needed in the field, although those with a bachelor's degree in computer science or a Ph.D. in biology or bioinformatics are especially in demand. Perl programming is the most frequently requested skill across advertised positions, and experience using bioinformatics software tools (such as BLAST, CLUSTAL, and HMMER) is the bioinformatics skill in greatest demand. A small number of positions specifically requesting librarians and LIS-trained bioinformaticists was observed, but the significant bioinformatics activities in biomedical libraries over the same time period are not reflected in this snapshot of the bioinformatics workforce.

Introduction

As administrators of a growing master's-level biological informatics training program (BIS, 2009), with a second cohort of students nearing graduation, we are in the process of assessing the current employment environment for master's-level informatics graduates. What types of institutions are employing biological informatics specialists? What types of positions are available, and what duties are typical? What knowledge and skills are employers seeking? What educational backgrounds are requested? How might these factors influence curriculum evolution and internship objectives?

To begin answering these questions, the first phase of analysis presented here examined positions advertised in standard bioinformatics job resources. The results will inform our program development efforts aimed at preparing information professionals to make needed contributions in librarianship and biological research. The report should also be of interest to directors of informatics programs with other foci (such as computer science and biological sciences), to higher-level academic administrators, and to students.

Context

Recent years have shown a dramatic increase in the volume of biological data that is generated, a result of new technologies and innovations in the field. However, biological scientists often lack the technical knowledge and expertise required to maintain and analyze these large data sets. In contrast, computer scientists often lack the biological knowledge necessary to fully develop technical solutions. The problem is further exacerbated by the many

difficulties involved in organizing, storing, and preserving large amounts of data; neither group usually has training or experience in information management. Apart from data, there has also been corresponding growth in the scientific literature, which is the locus for storing scientific knowledge. Intellectual and technical tools to assist in the collection, organization, processing, synthesis, presentation, and dissemination of scientific knowledge have also proliferated. These issues all call for the involvement of professionals knowledgeable not only in the computer science and biological domains, but also in library and information science concepts, techniques, and tools. (Boguski, 1998; Heidorn, Palmer, & Wright, 2007; Howe, et al., 2008).

In response to the growing need for professionals to address these issues, many bioinformatics degree programs have been developed in recent years (Black & Stephan, 2004; Black & Stephan, 2005; Hemminger, Losi, & Bauers, 2005). These programs are typically established in biological or computer science departments. Occasionally, library and information science departments are also involved. The approaches of these different programs and even their definitions of “bioinformatics” often differ greatly (Sahinidis, et al., 2005). The variety of educational programs available and the interdisciplinary nature of the field make it difficult to judge the skills and education most valued in the current employment environment. In this paper we explore the state of the current bioinformatics job market in an effort to discover which bioinformatics skills are most in demand at this time. We also investigate employer types and locations. We address these questions by taking an in-depth look at the data derived from a sample of bioinformatics positions advertised between January 2003 and December 2008.

Prior work

The studies most aligned with our approach were conducted by Black and Stephan (2004), beginning with a series of interviews with industry representatives to assess the hiring practices of their firms. They found that when hiring bioinformaticists, individuals trained in both biology and computer science were preferred. Specifically, training in mathematics, statistics, systems, structural, and molecular biology, and strong computational skills were emphasized as important. They also found that private industry was the most frequent employer of graduates from bioinformatics programs, with the academic sector a close second. An analysis of job ads in the journal *Science* between 1996 and 2002 showed that over time, ads for jobs in industry have become less frequent, while ads for jobs in academia have become more frequent. The rise of academic jobs corresponds with an increased demand for graduates with doctoral degrees, with very few jobs available to graduates with a master’s or bachelor’s degree. Black and Stephan (2005) built on this work by analyzing bioinformatics positions advertised online through BioPlanet.com and Monster.com between 2002 and 2004, again finding that academic positions were rising, industry positions were falling, and that demand for graduates with doctoral degrees was increasing.

Methods

Data Sources

Bioinformatics-themed jobs were collected from multiple online sources for the 2-year period of January 2007 through December 2008. In addition, jobs dating back to January 2003 were collected from the archives of the bioinformatics.org job forums. Jobs were classified as “bioinformatics-themed” if they used the term “bioinformatics” in the job title or description, or if they required a mixture of computational and biological skills or education. A short description for each of the online sources used for this collection follows.

Bioinformatics.org

The Bioinformatics Organization, Inc. (<http://www.bioinformatics.org/>) has provided resources and opportunities for collaboration for bioinformatics professionals since its establishment in 1998. The website provides a number of tools, databases, and forums free for public use. One of these forums is the Jobs Forum, where postings of job offering in bioinformatics are archived dating back to October of 2001. This study made use of these archives to analyze jobs from January 2007 through December 2008. In addition, jobs dating back to January 2003 were analyzed in brief to show trends over time and to provide context for the in-depth analysis.

BioSpace

BioSpace (<http://www.biospace.com/>) is an online community that provides news and career information for life science professionals. BioSpace specializes in the biotech and pharmaceutical industries and maintains a job board for opportunities in those fields. This study used jobs posted in BioSpace from March 2008 through June 2008 which were retrieved by a simple search of the keyword “bioinformatics”.

Biohealthmatics

Biohealthmatics (<http://www.biohealthmatics.com/>) is a “career networking portal” that specializes in biotechnology and healthcare IT fields. In addition to networking opportunities the website maintains a job search portal where employers may post opportunities and users can post resumes and browse for jobs. This study used jobs that were retrieved in a search of the keyword “bioinformatics” and appeared between March and June of 2008.

ISCB

The International Society for Computational Biology (ISCB) (<http://www.iscb.org/>) is a non-profit organization that hosts conferences, writes publications, provides education and training, and offers many other opportunities with the goal of promoting understanding in the field of computational biology. ISCB also maintains a job board, and jobs posted on the board between February and June of 2008 were used in this analysis.

Science Careers

Science Careers (<http://nextwave.sciencemag.org/>) is the career component of the journal *Science*, and the website acts as a career portal for scientists looking for jobs in industry, academia, and government. This study used jobs posted in *Science Careers* between February and June of 2008 which were retrieved from a keyword search of “bioinformatics”.

Bioinformatics (journal) website

The journal *Bioinformatics* (<http://bioinformatics.oxfordjournals.org/>) publishes scientific papers and review articles which focus on genome bioinformatics and computational biology. The *Bioinformatics* website occasionally provides career announcements as a reader service, and these announcements were collected from the website between March and June of 2008.

Other Sources

Some job postings were not retrieved through a job search service but were discovered through biological mailing lists or directly from the employer’s website. The majority of these jobs were posted on the Taxacom (<http://mailman.nhm.ku.edu/mailman/listinfo/taxacom>) or TDWG (<http://www.tdwg.org/>) mailing lists.

Data Preparation

We collected employment advertisements from the sources listed above into two spreadsheets. The first was compiled from the entire content of the bioinformatics.org jobs

archive in the 6-year period from January 2003 through December 2008. Reposts of unfilled jobs were eliminated when identification of these was possible. Posts with more than one position per advertisement were separated into multiple spreadsheet rows when the actual numbers of positions were stated. All entries were coded into four variables: date, location, employer type, and job role. 'Date' was coded as month and year, and 'location' was a binary variable used to assess whether jobs were based in the United States or elsewhere. 'Employer type' distinguished 8 broad categories of employers we observed in the data (Academia, Corporate, Government, Hospital, Non-governmental organization (NGO), Research Institute, Resource Provider, and Unknown). Since many ads are posted by recruiters or government contractors, we attempted to identify the actual venue for each position (e.g., coding a GenBank curator position at NIH that was advertised through a corporate recruiter as 'Resource Provider' rather than 'Corporate'). 'Job role' was coded into 18 categories, which, like 'employer type', we observed in analyzing the data (Consulting, Curator, Customer Service, Data Analyst, Faculty, Librarian, Manager, Other, Postdoc, Programmer, Sales/Marketing, Scientist, Statistician, Technical Writer, Technician, Trainee, Trainer, and Unknown). This data set was used to characterize the larger context of bioinformatics jobs and trends over time.

A second spreadsheet was used to compile more detailed job description data from all of the venues described in the Data Sources section for the period of January 2007 through December 2008. In order to focus on jobs that a graduate of a bioinformatics master's degree program might be qualified for, we removed advertisements for positions requiring a Ph.D. degree (e.g., faculty and post-doctoral positions), and for Ph.D. traineeships. We also removed positions that were specifically focused on biostatistics rather than bioinformatics, as well as research assistantships and internships. Finally, we removed positions that were located outside of the United States. As with the larger data set, posts with more than one position per advertisement were separated into multiple spreadsheet rows, and duplicate posts for the same position were removed when possible.

The full texts of the resulting set of position advertisements were coded into several facets of interest, including date, location, and employer type as described above, as well as level of education requested, science and technical experience requested, and specific programming languages, biological protocols, and bioinformatics skills requested for each job. We also analyzed location at the state level in order to examine geographic distribution of jobs more granularly.

Results

The total number of employment advertisements obtained and analyzed from the bioinformatics.org archive for the January 2003 through December 2008 period was 1,996. The subset of advertisements from 2007-2008 obtained from all sources and analyzed in more detail totaled 404 (about 40% of all 2007-2008 advertisements).

Table 1 shows that the number of opportunities posted to bioinformatics.org has increased substantially over the 6-year period studied, but declined slightly in 2008 compared with 2007. Of the 1,996 positions, 1,175 (59%) were U.S.-based jobs. The top 3 types of employers (academic, corporate, and research institutes) accounted for more than 82% of all jobs during the period, as shown in Table 2. For the more selective set of 404 job posts explored for relevancy to master's graduates, the top three employer types are the same as the larger data set, but are ordered differently, with Corporate, Research Institute, and Academic jobs in the top three positions, respectively. Figure 1 illustrates that these three categories exhibited the most growth over the period as well, with other types remaining relatively flat.

Table 1. Distribution of job posts to bioinformatics.org, Jan. 2003 – Dec. 2008

| Year | Job posts |
|--------------|--------------|
| 2003 | 174 |
| 2004 | 162 |
| 2005 | 278 |
| 2006 | 360 |
| 2007 | 540 |
| 2008 | 482 |
| Total | 1,996 |

Table 2. Distribution of job posts by employer type (bioinformatics.org)

| Employer Type | 2003-2008 Job Posts (% of Total) | 2007-2008 Job Posts (% of Total) |
|--------------------|--|--|
| Academic | 705 (35.3) | 53 (13.7) |
| Corporate | 541 (27.1) | 169 (43.8) |
| Research Institute | 398 (19.9) | 94 (24.4) |
| Unknown | 109 (5.5) | - |
| Government | 79 (4.0) | 19 (4.9) |
| Resource Provider | 65 (3.3) | 32 (8.3) |
| Hospital | 50 (2.5) | 14 (3.6) |
| NGO | 49 (2.5) | 5 (1.3) |
| Total | 1,996 (100) | 386 (100) |

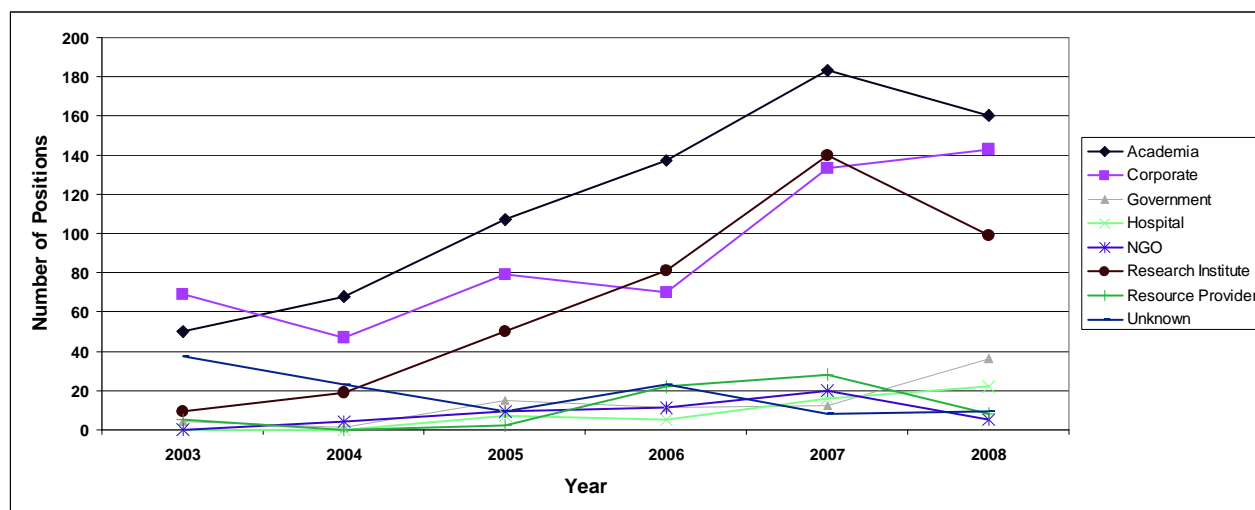


Figure 1. Trends in posted positions by employer type (2003-2008)

Table 3 shows that half of the 18 identified job roles accounted for more than 96% of all positions advertised over the period.

Table 3. Distribution of positions by job role (top 9 of 18 for 2003-2008)

| Job Role | Positions (% of Total) | Job Role | Positions (% of Total) |
|------------|---------------------------|--------------|---------------------------|
| Programmer | 612 (30.7) | Manager | 121 (6.1) |
| Postdoc | 406 (20.3) | Statistician | 101 (5.1) |
| Scientist | 240 (12.0) | Data Analyst | 93 (4.7) |
| Trainee | 163 (8.2) | Curator | 65 (3.3) |
| Faculty | 123 (6.2) | Other | 16 (0.8) |

We explored geographic location in more detail for the 2007-2008 data, and found that 31 of the 50 U.S. states were represented, but the bioinformatics positions were not distributed evenly

across them. The three states with the highest number of advertisements, California, Maryland, and Massachusetts, accounted for more than 50% of all advertisements. The top 10 states accounted for over 75% of the positions. Seven states had only 1 advertisement each.

In California, the state with the highest number of advertisements, most jobs are with corporations. Maryland has a disproportionate number of resource provider positions because NCBI, a division of NIH, is based there, and Massachusetts is biased towards research institutes because it houses the Broad Institute of MIT and Harvard.

The two most frequent advertisers of bioinformatics positions were the National Institutes of Health (NIH) (25 jobs) and the Broad Institute at MIT and Harvard (22 jobs). Many employers advertised multiple jobs during the 18 month period of this analysis, and in particular large corporations such as the Monsanto Company and SAIC-Frederick tended to post multiple opportunities. Although many employers had more than one advertisement, only NIH and the Broad Institute each exceeded 10 job postings since January 2007.

Education attributes

In the selected 2007-2008 data, the vast majority (91%) of advertisements explicitly required some type of university degree. While 43% of posts asked for a Bachelor's degree at minimum, 24% required a Master's, and 23% required a Ph.D. Only 2 positions asked for an associate's degree. A variety of disciplines were requested, but typically the required degrees were in computer science (CS), biology, or bioinformatics. At the Bachelor's level, 82% of positions asked for a CS degree, while 74% asked for bioinformatics or biology degrees. At the Master's level, 53% of positions asked for CS degrees and 87% asked for bioinformatics or biology degrees. Finally, at the Ph.D. level, only 41% of positions asked for CS degrees while 96% of positions (all but 4) asked for a bioinformatics or biology degree. (These percentages do not sum to 100 because the CS and bioinformatics/biology categories are not mutually exclusive: many positions cited multiple fields as possibilities for a degree type.)

Experience

The type of work experience required in advertisements can also be divided into categories based on discipline: computer science, biology, and bioinformatics. Each category contains knowledge and skills that are frequently repeated.

Biology/Bioinformatics Experience

Table 4 lists the most common biology and bioinformatics skills cited by employers. Experience with bioinformatics tools appears most often, and by a good margin. This category refers chiefly to sequence analysis tools such as BLAST and CLUSTAL, but it can apply to any software used for processing and/or analysis of bioinformatics data. Education and research experience in genetics and genomics follows as the second highest. The third skill listed is sequence analysis, and refers to experience with genetic sequence analysis software.

Experience using bioinformatics data resources is listed fourth. This category refers to experience manipulating, managing, and retrieving data from public bioinformatics resources, such as the Protein Data Bank (PDB), Genbank, or the Gene Ontology. Knowledge of molecular biology is another fairly common request, along with skills in microarray analysis. Biomedical informatics, experience in pharmaceutical research, and experience with clinical research are also cited multiple times, showing that there are recurring clinical medicine features in the bioinformatics job market.

Computer Science Experience

The most often requested computer science skill is the ability to program in one or more programming languages (Table 5). Perl is the most frequently asked for language, closely followed by Java, SQL, and finally C/C++. Other languages are mentioned much less frequently. Table 6 lists the major languages and the number of times they were cited by an employer as a required or desired skill.

After programming skills, the ability to create and maintain relational databases is the next most frequently required knowledge skill, followed by software development experience. Experience with web development technologies and knowledge of UNIX/LINUX are also frequently mentioned skills.

Table 4. Most frequently requested biology and bioinformatics skills

| Skill | # of Positions |
|--|-----------------------|
| Bioinformatics tools | 93 |
| Genetics/Genomics | 90 |
| Sequence analysis | 73 |
| Bioinfo data resources | 57 |
| Molecular Biology | 41 |
| Microarray analysis | 35 |
| Biomedical Informatics | 26 |
| Lab Experience | 24 |
| Proteomics | 18 |
| Pharmaceutical | 17 |
| Clinical research | 15 |
| Phylogenetic Tools | 8 |
| LIMS (laboratory information management systems) | 8 |
| Annotation of genomes | 6 |

Table 5. Most frequently requested computer science skills

| Skill | # of Positions |
|-------------------------|-----------------------|
| Programming | 287 |
| Databases | 202 |
| Software development | 115 |
| Web technologies | 97 |
| Unix | 96 |
| Linux | 90 |
| Oracle | 43 |
| XML | 37 |
| Algorithm development | 35 |
| version control systems | 21 |
| Mac OS | 19 |
| Data Mining | 13 |
| web services | 10 |

Table 6. Programming languages that are a required or desired skill in multiple jobs

| Language | Count | Language | Count |
|-----------------|--------------|-----------------|--------------|
| Perl | 178 | SAS | 15 |
| Java | 146 | Ruby/RoR | 13 |
| SQL/mysql | 128 | CGI | 13 |
| C/C++ | 122 | AJAX | 13 |
| Python | 68 | CSS | 12 |
| R | 54 | S-Plus | 10 |
| HTML | 40 | ASP | 8 |
| PHP | 29 | bioPerl | 7 |
| Javascript | 24 | VB/VB.NET | 7 |
| Matlab | 22 | Jython | 5 |
| C#/C#.NET | 19 | | |

Discussion

Employer

The differences in job type are very clear in this analysis. Corporate positions are clearly in the majority, and they far outnumber jobs in academia. This is in direct conflict to results from Black and Stephan's 2004 report on the bioinformatics job market. They saw industry jobs decreasing rapidly while academia positions increased. If that trend had continued we would expect to see the results for academia and corporate switched in our analysis, with academic positions dominating the job market.

One possible explanation for this conflict is the fact that we removed postdoctoral and faculty positions from our focused data set. Both job types are typically in academia, so it is possible their addition would change the look of our results entirely. To test this, the entire archive of the bioinformatics.org forum between January 2007 and December 2008 was analyzed for job location (U.S. and non-U.S.) and employer category. This data set includes all the job types disregarded in the main study, most notably postdoctoral positions, which constitute approximately 20% of the entire data set.

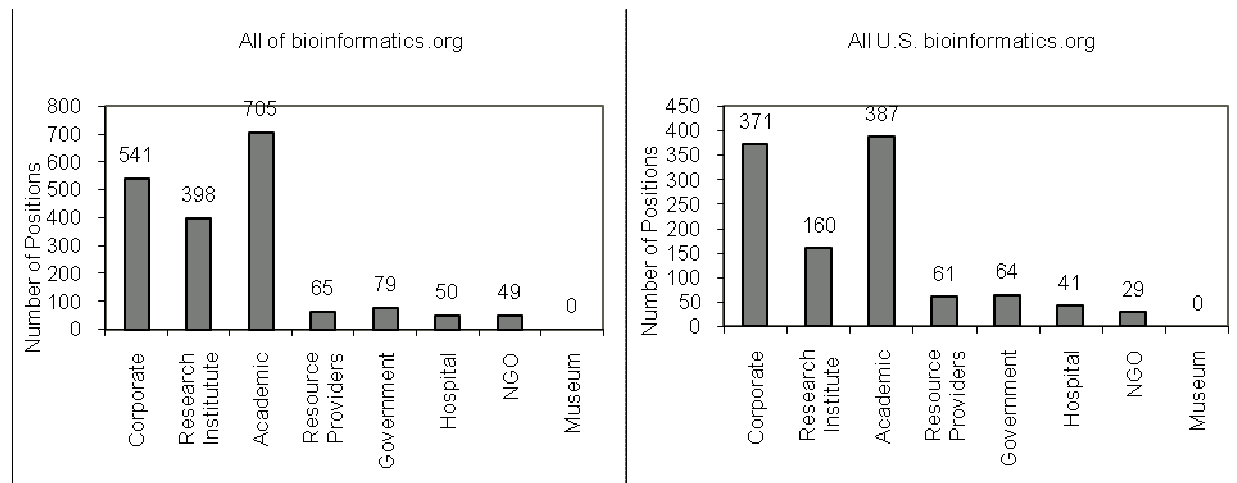


Figure 2. Bioinformatics positions divided by employer category. Depicts all positions posted in the bioinformatics.org job forums between January '07 and December '08 worldwide (left) and US only (right).

The results (Fig. 2) show that when all types of jobs are included, academia is the category with the most jobs, as indicated by Stephan and Black. When all ads worldwide are considered, academia is the most frequent by a very large margin; however, when only ads from the U.S. are considered, the gap between corporate and academic jobs decreases significantly. In both cases, research institutes also represent a large portion of open positions.

Education

When comparing the level and type of degree requested, it is interesting to note that computer science is seen more often as a degree requirement at the Bachelor's level, while biology and bioinformatics are featured more often at the Master's and Ph.D. level. These results imply that in bioinformatics lower levels of computer science training are acceptable, while biological or bioinformatics training needs to be at a Master's level or higher.

Another interesting trend conflicts with the 2004 Black and Stephan report in terms of degrees in highest demand. They found that Ph.D. level positions were the most numerous, followed by

Master's and then Bachelor's. In addition, they found that very few or no positions were at the Bachelor's level. This study found that positions with a minimum requirement of a Bachelor's degree were most numerous, followed by Ph.D. and then by Master's. Again this is certainly due to the exclusion of postdoctoral and faculty positions from this study. Including these would raise the count for Ph.D. positions much higher than any of the others. However, unlike Black and Stephan, we identify an apparent market for graduates with Bachelor's degrees.

Experience

The results from this study allow us to make some observations about what skills are essential for a bioinformaticist to develop. The minimum technical experience is certainly programming in Perl. Next would be an understanding of relational databases, including knowledge of Oracle and programming in SQL. Finally, experience with object-oriented programming in either Java or C++ is also in demand. The minimum biological experience is a solid background in molecular biology and genetics. Bioinformatics requirements include experience with sequence alignment and other bioinformatics tools, and also experience using public resources such as the Protein Data Bank and the Gene Ontology.

Trends Over Time

Although academic jobs outnumber corporate and research institute jobs, the gap between these categories has remained consistent for several years and does not seem to be widening. This trend may indicate a stabilization of the market, with academic jobs at the top and corporate and research institute positions in nearly equal numbers right below it. All other categories except government jobs have increased in frequency only minimally and do not constitute a significant portion of the work force. Government jobs show a large rise in 2008, perhaps reflecting growth of government resources via NCBI and other NIH initiatives.

Conclusion

To conclude we turn to the original goal of this paper, which was to profile the range of bioinformaticist expertise that is most in demand. We found that the majority of bioinformatics positions are available in corporations, research institutes, and academic institutions, and California, Maryland, and Massachusetts are the three states that dominate the labor market. While we expected to see an abundance of what might be considered more traditional bioinformatics positions, we also thought new trends in the information professions might be better represented in the skills being sought or in the scope of the positions advertised. This was not the case.

Fundamental Skill Sets

We have shown that graduates from all levels of education are needed in the field. Those with a Bachelor's degree in a relevant discipline, especially computer science, are especially in demand, as well as graduates with a Ph.D. in biology or bioinformatics. Finally, we know that programming in Perl is the most frequently cited knowledge skill out of all advertised positions, and experience with the use of bioinformatics tools is the bioinformatics skill in most demand. A combination of the above characteristics would make an applicant more than qualified for the majority of bioinformatics jobs in the current market.

Roles for Librarians

Library and information science was represented only minimally in this data set. Only 17 positions reference library science skills as desirable. Only 7 positions asked specifically for data management and curation skills. The remainder were oriented toward science librarianship or librarians as a resource provider. This was not surprising, as other venues more commonly list available jobs associated with library-oriented bioinformatics resources and services, such

as the MEDLIB-L email list, and the employment opportunities sections of the Medical Library Association's (MLA) web site and monthly newsletter, *MLA News*. There has been a focus since the early 2000s on developing bioinformatics roles in academic health sciences libraries (see, e.g., Helms, et al., 2004; Osterbur, et al., 2006), and we expect this trend to continue. The preceding analysis shows that the focus of most bioinformatics positions at this time is in gathering and interpreting data, not in organizing, preserving and providing access to it. This is an unfortunate trend because many biological data sets are irreplaceable, and without proper curation a great deal of information will be, and certainly has been, lost. As the volume of data continues to grow, this problem will simply become more urgent as time passes. It thus seems likely that the value of information professionals and librarians in bioinformatics will be recognized, but it remains unclear how and when the need for these skills will be represented in the mainstream job market. The 'Curator' role, which is conceptually closest to an LIS perspective, accounted for only 3.3% of advertisements in the 2003-2008 period (Table 3).

Supplementary Data

Additional data and analysis is available online at: <http://www.isrl.illinois.edu/~wjohn/asist>

Acknowledgements

This work was performed as part of NSF award # IIS-0534567. We are grateful to Mijung Yoon for assistance with supplementary data analysis.

References

- BIS, 2009. Master of Science in Biological Informatics degree program, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Retrieved 2009-06-11, from: <http://www.lis.illinois.edu/programs/ms-bioinformatics.html>
- Black, G. C., & Stephan, P. E. (2004). *Bioinformatics: Recent Trends in Programs, Placements and Job Opportunities*. Report to the Alfred P. Sloan Foundation. Retrieved 2009-06-11, from: http://www2.gsu.edu/~ecopes/reports/Bioinformatics%20Final%20Report_Black%20and%20Stephan_June%202004.pdf
- Black, G. C., & Stephan, P. E. (2005). Bioinformatics Training Programs are Hot but the Labor Market is Not. *Biochemistry and Molecular Biology Education*, 33(1), 58-62.
- Boguski, M. (1998). Bioinformatics – a new era. *Trends in Biotechnology*, 16(suppl. 1), 1-3.
- Heidorn, P. B., Palmer, C. L., & Wright, D. (2007). Biological Information Specialists for Biological Informatics. *Journal of Biomedical Discovery and Collaboration* 2(1).
- Helms, A. J., Bradford, K. D., Warren, N. J., & Schwartz, D. G. (2004). Bioinformatics opportunities for health sciences librarians and information professionals. *Journal of the Medical Library Association*, 92(4):489-93.
- Hemminger, B., Losi, T., & Bauers, A. (2005). Survey of Bioinformatics Programs in the United States. *Journal of the American Society for Information Science and Technology*, 56(5), 529–537.
- Howe, D., Costanzo, M., Fey, P., et al. (2008). The future of biocuration. *Nature*, 455(4), 47-50.
- Osterbur, D. L., Alpi, K., Canevari, C., Corley, P. M., Devare, M., Gaedeke, N., Jacobs, D. K., Kirlew, P., Ohles, J. A., Vaughan, K. T., Wang, L., Wu, Y., & Geer, R. C. (2006). Vignettes: Diverse library staff offering diverse bioinformatics services. *Journal of the Medical Library Association*, 94(3):306, E188-91.
- Sahinidis, N. et al. (2005). Establishing a master's degree programme in Bioinformatics: challenges and opportunities. *Systems Biology, IEEE Proceedings*, 152(4), 269-275.

Supplementary Data and Analysis

Hill, J. I.; MacMullen, W. J.; & Palmer, C. L. Characteristics of bioinformatics employment advertisements. *Proceedings of the 2009 Annual Meeting of the American Society for Information Science & Technology*.

This file provides additional data and analysis for which space was not available in the published paper. The authors thank Mijung Yoon for her assistance.

Tables S1-S3. Skills requested for master's level by domain in each of 3 main employment settings (selected 2007-2008 data (n=404))

Table S1. CS skills

| CS Skills | Major Employment Setting | | | Total |
|--------------------------|--------------------------|-----------|--------------------|-------|
| | Academic | Corporate | Research Institute | |
| Programming | 6 | 26 | 15 | 47 |
| Oracle+Databases | 6 | 22 | 15 | 43 |
| Unix+Linux | 6 | 15 | 8 | 29 |
| Web | 5 | 4 | 10 | 19 |
| Software development | 1 | 11 | 5 | 17 |
| Data mining | 1 | 1 | 0 | 2 |
| Algorithms | 1 | 6 | 1 | 8 |
| XML | 1 | 0 | 3 | 4 |
| Version control systems | 0 | 1 | 0 | 1 |
| Web service architecture | 0 | 0 | 2 | 2 |
| Mac OS | 0 | 1 | 0 | 1 |
| No tech skills specified | 6 | 7 | 1 | 14 |

Table S2. Bio/info skills

| Bio/info skills | Major Employment Setting | | | Total |
|--------------------|--------------------------|-----------|--------------------|-------|
| | Academic | Corporate | Research Institute | |
| BI tools | 4 | 13 | 7 | 24 |
| BI resources | 4 | 9 | 4 | 17 |
| Sequence analysis | 2 | 11 | 5 | 18 |
| Microarray | 1 | 3 | 3 | 7 |
| Genetics/genomics | 2 | 9 | 1 | 12 |
| Molecular bio | 1 | 6 | 0 | 7 |
| Lab experience | 1 | 0 | 0 | 1 |
| Pharma | 0 | 3 | 0 | 3 |
| Proteomics | 0 | 2 | 0 | 2 |
| Clinical research | 0 | 3 | 0 | 3 |
| Biomed | 0 | 5 | 1 | 6 |
| Phylogenetic tools | 1 | 1 | 0 | 2 |
| Genome annotation | 0 | 3 | 0 | 3 |
| Statistics | 2 | 16 | 1 | 19 |

Table S3. All skills

| All Skills | Major Employment Setting | | | Total | Average |
|--------------------------|--------------------------|-----------|--------------------|-------|---------|
| | Academic | Corporate | Research Institute | | |
| Programming | 6 | 26 | 15 | 47 | 15.7 |
| Oracle+Databases | 6 | 22 | 15 | 43 | 14.3 |
| Unix+Linux | 6 | 15 | 8 | 29 | 9.7 |
| BI tools | 4 | 13 | 7 | 24 | 8.0 |
| Web | 5 | 4 | 10 | 19 | 6.3 |
| Statistics | 2 | 16 | 1 | 19 | 6.3 |
| Sequence analysis | 2 | 11 | 5 | 18 | 6.0 |
| Software development | 1 | 11 | 5 | 17 | 5.7 |
| BI resources | 4 | 9 | 4 | 17 | 5.7 |
| No tech skills specified | 6 | 7 | 1 | 14 | 4.7 |
| Genetics/genomics | 2 | 9 | 1 | 12 | 4.0 |
| Algorithms | 1 | 6 | 1 | 8 | 2.7 |
| Microarray | 1 | 3 | 3 | 7 | 2.3 |
| Molecular bio | 1 | 6 | 0 | 7 | 2.3 |
| Biomed | 0 | 5 | 1 | 6 | 2.0 |
| XML | 1 | 0 | 3 | 4 | 1.3 |
| Pharma | 0 | 3 | 0 | 3 | 1.0 |
| Clinical research | 0 | 3 | 0 | 3 | 1.0 |
| Genome annotation | 0 | 3 | 0 | 3 | 1.0 |
| Data mining | 1 | 1 | 0 | 2 | 0.7 |
| Web service architecture | 0 | 0 | 2 | 2 | 0.7 |
| Proteomics | 0 | 2 | 0 | 2 | 0.7 |
| Phylogenetic tools | 1 | 1 | 0 | 2 | 0.7 |
| Version control systems | 0 | 1 | 0 | 1 | 0.3 |
| Mac OS | 0 | 1 | 0 | 1 | 0.3 |
| Lab experience | 1 | 0 | 0 | 1 | 0.3 |
| LIMS | 0 | 0 | 0 | 0 | 0.0 |

* Skills are listed by the number of listings for all three settings from high to low.

Figures S1-3. Skill set profiles for master's level in each of 3 main employment settings (selected 2007-2008 data (n=404))

Figure S1. Academic setting

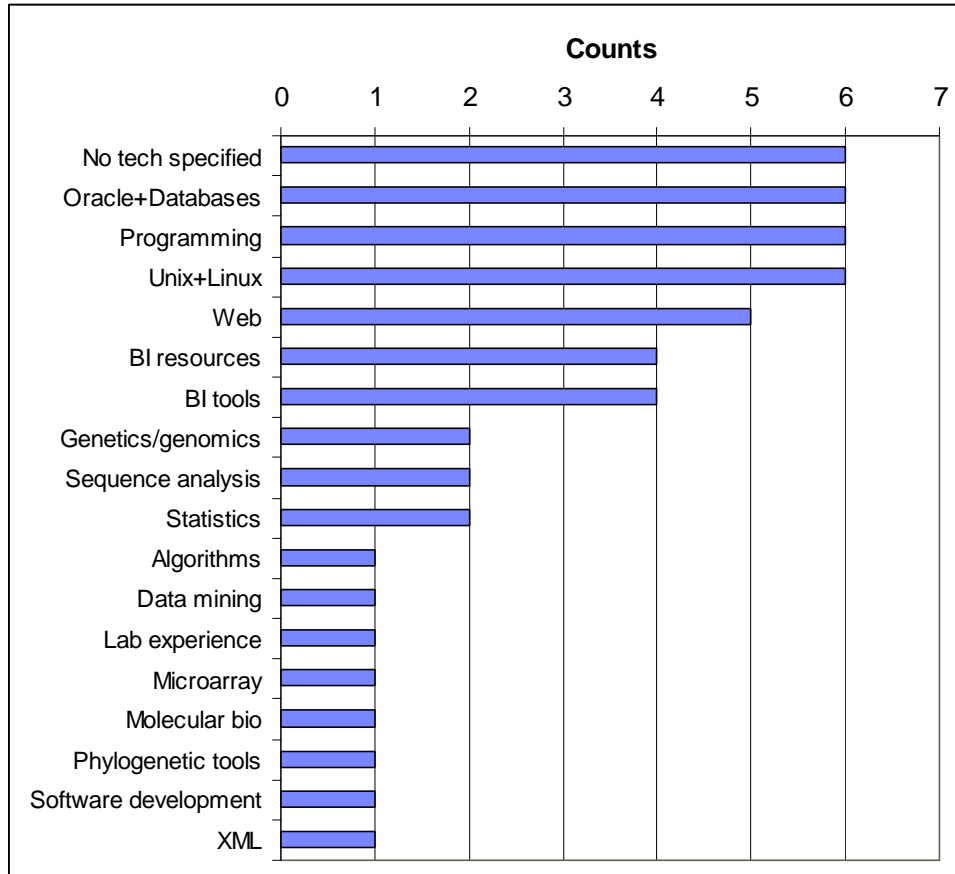


Figure S2. Corporate setting

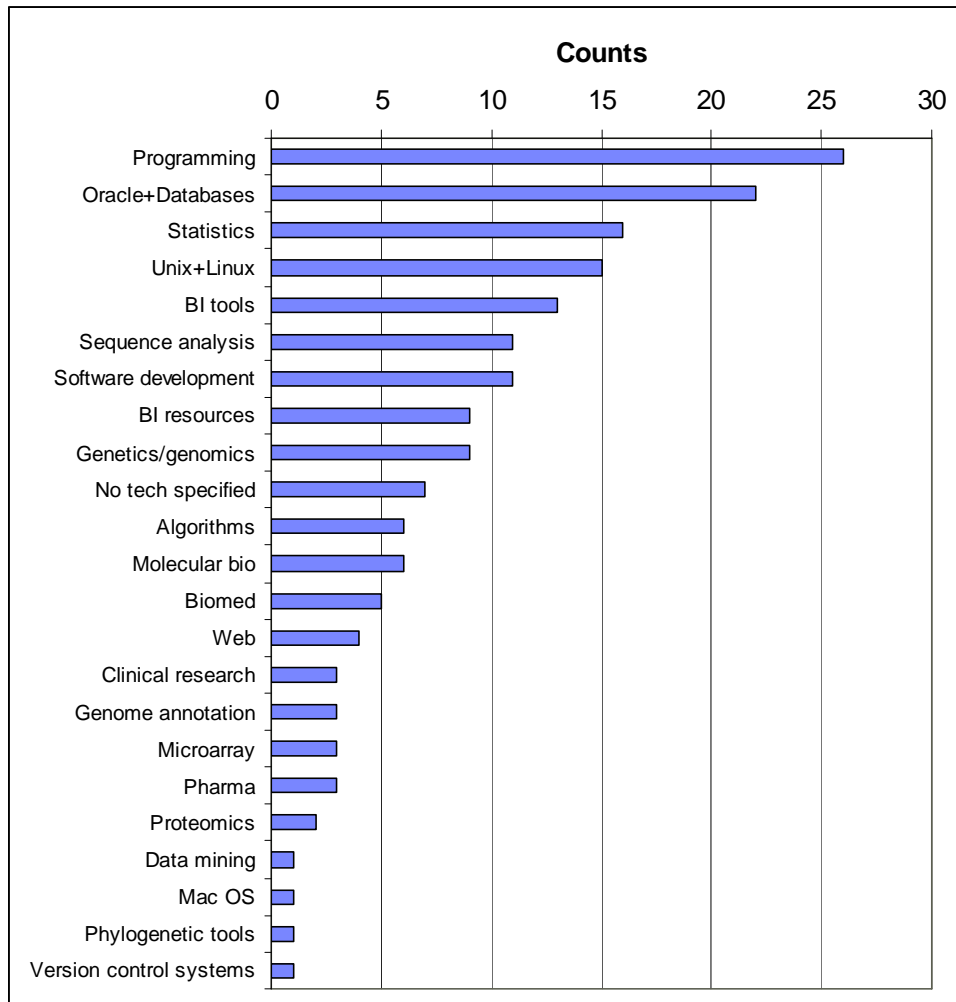
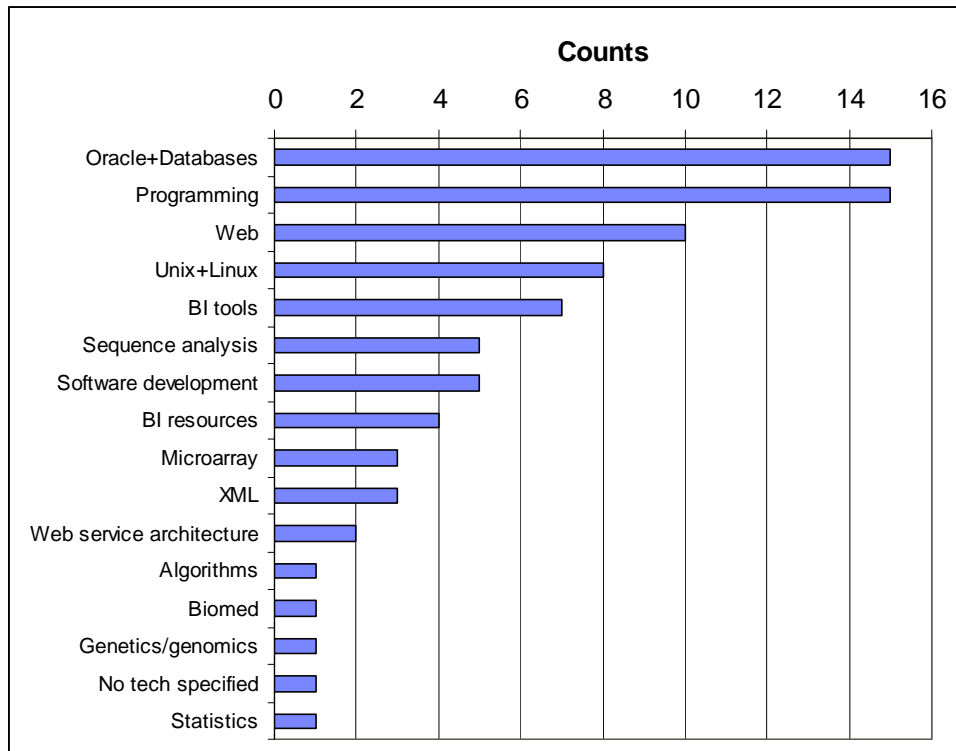


Figure S3. Research Institute setting



Summary of Position Descriptions: Programmer

There were 66 positions in the 2007-2008 data set (n=404). The most common responsibility associated with a programmer position was to develop and support bioinformatics software and databases for the institution-wide use or for specific research projects, often working closely with researchers and scientists. The programmer positions were often titled as Bioinformatics Programmer, Scientific Programmer, Programmer Analyst, or Programmer. While most of the programmer positions shared the common responsibility of computational programming support for the institution or projects, programmer analyst positions sometimes included additional responsibilities of data transfer and modeling. The following is a summary of the 66 programmer positions:

1. Development, modification, and maintenance of bioinformatics databases and software.
 - Developing tools for browsing, managing and integrating research/clinical data or databases, including:
 - Designing and writing databases and software for biological and clinical data management (e.g., defining architecture of customized databases for the institution's infrastructure);
 - Testing the software by performing edit checks in compliance with standard operating procedures and regulatory agency guidelines, in coordination with data management;
 - Documenting and reporting data management process and outcome, and producing dictionary-coding programs and miscellaneous data management listings;
 - Creating database conversion programs;
 - Providing support for implementation and maintenance of custom or third-party bioinformatics software or databases, including:
 - Operating and upgrading software and databases for data management or analysis (e.g., web- or report-based sequence mapping tools);
 - Troubleshooting software problems;
 - (often titled as a scientific/research programmer) providing programming support for scientists in architecture, design, implementation, and testing of software application suites used for collecting, analyzing, reporting, and sharing (genetic and genomic) data, including
 - Maintaining close communication regarding data management programming activities with researcher(s);
 - Implementing tools for data modeling, analysis, and interpretation;
 - Developing and applying statistical techniques to data analysis (e.g., identification of genes for diseases);
 - Implementing middleware;
2. (often with the term 'systems' in a position title) building and maintaining infrastructure for computational biology, including: (1) development of a hardware platform; (2) configuration, maintenance, and integration of servers, workstations, clusters, and advanced storage under multiple OSes; (3) interfacing with network configuration and

- security; (4) developing reliable procedures for data backup and restoration; (5) helping define and implement user policies; and (5) implementing new system tools/processes.
3. (often titled as programmer analyst or programmer/data analyst) performing data analysis (e.g., microarray data analysis, sequence mapping);
 4. Developing web resources, improving websites and managing information portals
 5. (research programmer) data modeling, database design, middleware implementation, implementation of web and data services.

Terms That Were Frequently Associated with 'Programmer'
in Job Titles in the 2007-2008 data (n=404)

| Terms | Frequency |
|---|-----------|
| Bioinformatics, computer, or no additional terms | 23 |
| Analyst | 16 |
| Scientist/scientific/research | 12 |
| Systems /systems administrator/ systems analyst | 4 |
| Database administrator/application | 4 |
| Specific areas (e.g., immunology, plant genomics, clinical) | 4 |
| specific software (e.g., perl, oracle, java) | 4 |
| Computational Biology | 3 |
| Data analyst | 2 |
| Biostatistics | 1 |
| Web | 1 |
| TOTAL | 74 |

* The total doesn't equal to the number of positions due to several listings that had multiple terms that were associated with 'programmer'.

Summary of Position Descriptions: Curator

This description is based on the ten curator positions that were listed between January 2007 and June 2008. The descriptive terms listed in the titles along with the term 'curator' included: scientific, genome, microarray, genotype, phenotype, and name of a database (e.g., dbGaP, dbSNP). Qualifications sought for the position included: education (e.g., molecular biologist), background/expertise (e.g., life science, genotype analysis). Most of the positions listed responsibilities related to curation, review, and processing records for analysis. Less common responsibilities included literature review, data analysis, and communication with investigators. The detailed summary of responsibilities is as follows:

1. Ensuring data quality
 - a. Define data (e.g., genomic sequence records) standards
2. Data curation and processing (e.g., perform "advanced curation" of data into comparable datasets)
 - a. Reviewing and processing incoming data/records submissions
 - b. Developing infrastructure for data analysis (e.g., applications to process data into databases and CGI)
 - c. Using specialized databases and bioinformatics software
 - d. Providing functional annotation with records and companion resources
3. Perform data analysis
 - a. Computational analysis using various bioinformatics and programming tools
 - b. Scholarly interpretation of results
4. Documentation
 - a. Annotation and integration of study meta data (e.g., protocols and data collection forms)

Examples of qualifications listed for curator positions:

Computational biologist with background in genotype analysis

Biologist with expertise in life sciences

Genomic scientist with background in sequence analysis and positional cloning

Molecular biologist/biochemist