A decade of web server updates at the bioinformatics links directory: 2003–2012

Michelle D. Brazas¹, David Yim¹, Winston Yeung¹ and B. F. Francis Ouellette¹,²,*

¹Ontario Institute for Cancer Research, 101 College St., Suite 800, Toronto, Ontario, Canada M5G 0A3 and ²Department of Cell and Systems Biology, 25 Harbord Street, University of Toronto, Toronto, Ontario, Canada M5S 3G5

Received May 28, 2012; Revised June 4, 2012; Accepted June 5, 2012

ABSTRACT

The 2012 Bioinformatics Links Directory update marks the 10th special Web Server issue from Nucleic Acids Research. Beginning with content from their 2003 publication, the Bioinformatics Links Directory in collaboration with Nucleic Acids Research has compiled and published a comprehensive list of freely accessible, online tools, databases and resource materials for the bioinformatics and life science research communities. The past decade has exhibited significant growth and change in the types of tools, databases and resources being put forth, reflecting both technology changes and the nature of research over that time. With the addition of 90 web server tools and 12 updates from the July 2012 Web Server issue of Nucleic Acids Research, the Bioinformatics Links Directory at http://bioinformatics.ca/links_directory/ now contains an impressive 134 resources, 455 databases and 1205 web server tools, mirroring the continued activity and efforts of our field.

COMMENTARY

‘Because the number of artifacts has increased greatly, it is impossible for many bioinformatics researchers to track tools, databases, and methods in the field—or even perhaps within their own specialty area. More critically, however, biologist users and scientists approaching the field do not have a comprehensive index of bioinformatics algorithms, databases, and literature annotated with information about their context and appropriate use.’ (1)

The need to not only share bioinformatics databases, tools and resources but also compile a bioinformatics ‘resourceome’ has been evident for some time (1). In dedicating a special issue each year since 1993 to new databases or updates and a second issue each year since 2003 to new web server tools or updates, Nucleic Acids Research has successfully created a viable and recognizable dissemination mechanism. Each year since 2005 and extending back to 2003, the Bioinformatics Links Directory has coordinated with Nucleic Acids Research (NAR) to update their community driven ‘resourceome’ with all of NAR’s web server tools (2–8). This coordinated effort has resulted in a comprehensive public repository of bioinformatics tools and databases, annotated with contextual information and organized by functional relevance (http://bioinformatics.ca/links_directory). Users requiring a bioinformatics tool or database for their research may search the Bioinformatics Links Directory using any number of its features, from browsing the biological categories and their drop-down subcategories on the main page, to computing a keyword search from the search box or selecting related links. Content within the Directory is organized by biological subject, with subcategories of common tasks relevant to each subject listed under the corresponding category (Figure 1). A keyword search from the search box (Figure 2) returns links containing the entered keywords found within either the link’s title, description or tag terms—a search for ‘gene set enrichment’ for example returns links containing a Boolean combination of gene AND set AND enrichment, with these keywords highlighted in yellow where they appear in the returned list (Figure 3). Other specific searches may also be computed such as exact word matches using quotation marks. Returned links from a given search may be refined to one or more of the link types within the Links Directory (tools, resources or databases), using the ‘Search Options’ feature (Figure 4). From a link of interest, related or similar links within the Links Directory are retrievable by selecting the ‘Related Links’ tab, where related links are chosen based on the PubMed-related citations feature for the primary
link. To aid with determining the usefulness and currency of a link, all links within the Links Directory with one or more citations have a per PubMed Identifier (PMID) and cumulative Google Scholar (9) citation count, as well as a Google+ and Twitter sharing count (10) (Figure 5). A keyword search will return links in descending order of cumulative citation count. Detailed explanations and examples of searching within the Links Directory are provided on the tutorials page at http://www.bioinformatics.ca/links_directory/tutorial. These annotations and search features provide the ‘biologist user with information about a link’s context’ and usefulness so to narrow down the algorithm or database within the Bioinformatics Links Directory ‘resourceome’ best suited to their research need (1).

**SHIFTING TRENDS**

The past decade of *NAR* Web Server releases has seen significant growth and change in the types of tools, databases and resources being put forth. Although each Web Server issue has seen a steady number of new web servers and updates, the functional utility classification of these web servers has differed greatly over the decade. The Bioinformatics Links Directory in 2006 compiled a graph highlighting the trends in five subcategories of links published by *NAR* between 2003 and 2006 (3). An update of that graph (Figure 6) shows the exceptional changes in new bioinformatics web servers within these same selected subcategories; Table 1 lists how substantial these changes have been across the entire field of
bioinformatics as well (Table 1). Subcategories with active contributions of new web servers in 2006 differ from the subcategories with active contributions of new web servers in 2012 (Figure 6 and Table 1). For example subcategories such as ‘DNA: Tools for the Bench’ subcategory, which contains new web servers for conducting generic DNA analyses such as primer design, exhibited strong growth over 2003–2006 but shows very little growth in 2012 as tools in this category remain relevant over time (Figure 6). In contrast, other subcategories such as Protein: 3D Structural Features have remained active in web server development over time, as the introduction of better technologies and algorithms has permitted a deeper appreciation of proteins and their structural features in 3D (Figure 6).

Indeed, many of these trends in subcategory activity reflect the introduction of a new research technology and/or an underlying research need. As originally discussed in 2009, yet still evident throughout Table 1, there are observable sharp spikes in the number ofinformatics tools available in a given biological area or specific to a particular task (6). Although between 2006 and 2007, there was a spike in ‘DNA: Structure and Sequence Feature Detection’, ‘Expression: Gene Regulation’ and ‘Expression: Transcript Expression and Microarrays’ most likely in response to the pervasive use of expression platforms and transcript analyses at that time, current spikes in ‘Human Genome: Genomics’, ‘Sequence Comparison: Comparative Genomics’ and ‘Sequence Comparison: Pairwise Sequence Alignments’ between 2010 and 2012 may be reflective of the introduction of next-generation sequencing technologies. Table 1 highlights all the trends between 2005 and 2012 across all subcategories of the Bioinformatics Links Directory (Table 1) (individual calendar year data are not available for 2003–2005 as the Bioinformatics Links Directory collaboration with NAR began in 2005 for all data before that). Of note are the aggressive coverage of databases, which began in 2011 and the need for additional subcategories such as ‘Other Molecules: Enzymes and Peptides’ in response to the growing diversity and depth in the field.

CONSTANT GROWTH

The number of bioinformatics links produced by the community has nonetheless continued to accumulate over time. When first released in 2002 under Bioinformatics.ca, there were only 385 links in the database. By 2005, when the Bioinformatics Links Directory began its collaboration with NAR, the Directory already contained 700+ web servers and resources (2). The Bioinformatics Links Directory has
Figure 3. Keyword searches such as for gene set enrichment return matching categories and links, which contain the requested keywords highlighted in yellow.
A returned keyword search may be refined using the Search Options, where returned links may be refined to include or exclude all tools, resources and databases.
continued to grow (and diminish in size as web servers become extinct) with each successive year (Table 1). The Bioinformatics Links Directory now contains more than 1200 unique web servers. This year’s NAR Web Server Issue introduces an additional 90 web servers and 12 server updates (Table 1). The complete listing of URLs cited in the 2012 Web Server Issue can be accessed online at the Nucleic Acids Research website, http://nar.oxfordjournals.org/, and at http://bioinformatics.ca/links_directory/narweb2012/.

Growth has also resulted in change. In 2011, the Bioinformatics Links Directory introduced all the databases published by Nucleic Acids Research over their 2010–2011 window and completed a major reorganization of its existing links (8). Links were reclassified into one of three types reflecting the growing variety of bioinformatics

Figure 5. All links within the Bioinformatics Links Directory now contain a per PMID and cumulative Google Scholar citation count for the citations listed for each link, as well as a Google+ and Twitter sharing count.
materials being generated to support research: (i) Resource—a static resource whose intention is to convey bioinformatic information; (ii) Tool—a bioinformatic web server or downloadable software tool that can query, analyze, extract or modify input data and (iii) Database—a biological data store that can be queried. Under this new schema and with this expanded content, the Bioinformatics Links Directory now contains 1794 links comprising 455 databases, 134 resources and 1205 web server tools (Table 1). The Bioinformatics Links Directory has also initiated active curation of its content, removing dead content and correcting content errors, which has resulted in more accurate although occasionally smaller counts for 2012 (Table 1). Over time, the Bioinformatics Links Directory is expected to continue its growth and evolution in keeping with the current pace of research.

FUTURE CHALLENGES

With such a vast number of bioinformatics tools available for biological research, now more than ever, there is a need for a well-annotated and user-friendly ‘resourceome’ (1). In 2011, the Bioinformatics Links Directory introduced features to facilitate peer review of its links and link ownership (8). The application of Google +1 and other social media features, such as Twitter to the Bioinformatics Links Directory, provides users with the opportunity to share useful links; registered users may also post comments and review a link’s usefulness or start a discussion forum. Along with enhanced search capabilities and MeSH term and keyword tagging, such a peer-review system allows users to rapidly narrow down the links most applicable to their research need. Other features added provide link owners with the capacity to edit, manage and communicate on their own link without requiring input or support from the Bioinformatics Links Directory (8).

Although necessary to remain current and to advance the utility of the Bioinformatics Links Directory, these improvements will only prove useful if driven by the community. As a community-driven repository, everyone in the research or bioinformatics community has the opportunity to help make the collection better and more meaningful. Anyone may (i) suggest a link through links@bioinformatics.ca or through the ‘Suggest a URL’ page on the Bioinformatics Links Directory; (ii) register for a free account and submit their own tool or database; (iii) rate a tool, resource or database; (iv) register for a free account and submit their review of a tool, resource or database; (v) enrich a tool or database with tutorials or user-oriented content; (vi) initiate a discussion forum on a problem related to the use of any given tool or database or (vii) share a useful tool, resource or database with their colleagues. After a decade of output from the NAR Web Server issues, it will be interesting to monitor the development, maturity and impact of these bioinformatics web servers and databases over the next decade.

Figure 6. Number of NAR Web Servers from selected subcategories of the Bioinformatics Links Directory (http://bioinformatics.ca/links_directory/). The number of NAR Web Servers published from 2003–2012 for five selected subcategories of the Bioinformatics Links Directory: Protein: 3D Structural Features; Expression: Gene Regulation; DNA: Tools for the Bench; Sequence Comparison: Analysis of Aligned Sequences and RNA: Functional RNAs is shown.
Table 1. Historical summary of the number of resources, databases and web server tools listed in each subcategory of the Bioinformatics Links Directory between 2005 and 2012

| Name                        | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|-----------------------------|------|------|------|------|------|------|------|------|
|                             | Total| Total| Total| Total| Total| Total| Total| Total |
|                             | Resources Databases Tools Total | Resources Databases Tools Total | Resources Databases Tools Total | Resources Databases Tools Total | Resources Databases Tools Total | Resources Databases Tools Total | Resources Databases Tools Total | Resources Databases Tools Total |
| Computer related            |      |      |      |      |      |      |      |      |
| Bio-* programming tools     | 20   | 20   | 20   | 20   | 18   | 12   | 18   | 6    |
| C/C++                       |      |      |      |      |      |      |      |      |
| Databases                   | 2    | 2    | 3    | 3    | 3    | 2    | 1    | 1    |
| Java                        | 4    | 4    | 4    | 4    | 4    | 3    | 1    | 1    |
| Linux/Unix                  | 12   | 12   | 11   | 11   | 10   | 5    | 1    | 1    |
| PERL                        | 5    | 5    | 5    | 5    | 5    | 4    | 1    | 1    |
| PHP                         | 3    | 3    | 1    | 1    | 1    | 1    | 1    | 1    |
| Statistics                  | 9    | 9    | 9    | 9    | 9    | 2    | 7    | 9    |
| Web development             | 6    | 6    | 2    | 2    | 2    | 2    | 2    | 2    |
| Web services                | 5    | 6    | 7    | 10   | 7    | 3    | 16   | 19   |
| Workflows                   | 3    | 1    | 5    | 6    | 1    | 7    | 8    |      |
| DNA                         |      |      |      |      |      |      |      |      |
| Annotations                 | 29   | 38   | 56   | 57   | 62   | 71   | 1    | 6    |
| Databases                   |      |      |      |      |      |      |      |      |
| DNA and genomic analysis    | 17   | 19   | 20   | 23   | 24   | 25   | 5    | 1    |
| Gene prediction             | 17   | 32   | 33   | 34   | 37   | 37   | 1    | 1    |
| Mapping and assembly        | 12   | 14   | 15   | 15   | 17   | 20   | 3    | 2    |
| Phylogeny reconstruction    | 24   | 37   | 43   | 46   | 49   | 53   | 4    | 4    |
| Sequence polymorphisms      | 22   | 32   | 39   | 41   | 42   | 44   | 1    | 8    |
| Sequence retrieval and submission | 23  | 26   | 30   | 32   | 32   | 30   | 1    | 6    |
| Structure and sequence feature detection | 71  | 118  | 142  | 145  | 150  | 161  | 3    | 10   |
| Tools for the bench         | 44   | 55   | 63   | 65   | 71   | 76   | 2    | 7    |
| Education                   |      |      |      |      |      |      |      |      |
| Bioinformatics-related news sources | 9    | 9    | 9    | 9    | 9    | 9    | 7    | 7    |
| Community                   | 19   | 19   | 24   | 23   | 23   | 18   | 17   | 2    |
| Courses, programs and workshops | 5    | 5    | 5    | 5    | 4    | 4    | 3    | 3    |
| Directories and portals     | 12   | 15   | 15   | 15   | 15   | 15   | 12   | 6    |
| General                     | 15   | 15   | 14   | 14   | 14   | 14   | 13   | 14   |
| Tutorials and directed learning resources | 9    | 9    | 9    | 9    | 9    | 10   | 10   | 9    |
| Expression                  |      |      |      |      |      |      |      |      |
| cDNA, EST, SAGE             | 24   | 29   | 36   | 44   | 48   | 49   | 1    | 10   |
| Databases                   |      |      |      |      |      |      |      |      |
| Gene regulation             | 59   | 96   | 119  | 120  | 123  | 138  | 3    | 16   |
| Gene set analysis           | 17   | 24   | 23   | 23   | 18   | 17   | 2    | 19   |
| Networks                    | 12   | 15   | 15   | 15   | 15   | 15   | 12   | 6    |
| Protein expression          | 7    | 8    | 9    | 17   | 22   | 23   | 3    | 18   |
| Splicing                    | 9    | 16   | 19   | 19   | 21   | 22   | 3    | 17   |
| Transcript expression and microarrays | 54  | 75   | 89   | 101  | 108  | 121  | 5    | 10   |
| Human genome                |      |      |      |      |      |      |      |      |
| Annotations                 | 22   | 31   | 37   | 38   | 39   | 46   | 2    | 5    |
| Databases                   |      |      |      |      |      |      |      |      |
| Ethics                      | 7    | 7    | 8    | 8    | 6    | 5    | 1    | 1    |
| Genomes                     | 4    | 4    | 3    | 3    | 10   | 19   | 3    | 2    |
| Health and disease          | 14   | 17   | 19   | 23   | 27   | 29   | 1    | 7    |
| Other resources             | 24   | 25   | 29   | 29   | 31   | 9    | 4    | 17   |
| Sequence polymorphisms      | 17   | 25   | 33   | 36   | 38   | 46   | 2    | 4    |
| Literature                  |      |      |      |      |      |      |      |      |
| Databases                   | 7    | 7    | 7    | 7    | 7    | 7    | 7    | 7    |
| Goldmines                   | 5    | 6    | 6    | 6    | 5    | 1    | 4    | 5    |
| Open access resources       | 2    | 2    | 2    | 3    | 3    | 4    | 1    | 3    |
| Search tools                | 6    | 10   | 12   | 13   | 14   | 14   | 6    | 8    |
| Text mining                 | 7    | 11   | 15   | 22   | 30   | 31   | 4    | 30   |
| Model organisms             |      |      |      |      |      |      |      |      |
| Databases                   | 95   | 95   | 94   | 94   | 94   | 94   | 4    | 5    |
| Fish                        | 11   | 11   | 11   | 11   | 11   | 11   | 5    | 6    |

(continued)
| Name | 2005 Total | 2006 Total | 2007 Total | 2008 Total | 2009 Total | 2010 Total | 2011 | 2012 Total |
|------|------------|------------|------------|------------|------------|------------|------|------------|
| Fly  | 12         | 16         | 17         | 17         | 21         | 2          | 8    | 13         |
| General resources | 20 | 23 | 27 | 28 | 29 | 32 | 3 | 12 |
| Microbes | 19 | 31 | 45 | 53 | 60 | 63 | 3 | 12 |
| Mouse and rat | 28 | 32 | 35 | 36 | 42 | 4 | 17 |
| Other organisms | 13 | 18 | 21 | 21 | 22 | 3 | 6 |
| Other vertebrates | 10 | 10 | 10 | 10 | 11 | 1 | 3 |
| Plants | 11 | 16 | 19 | 21 | 25 | 28 | 2 | 10 |
| Worm | 9 | 9 | 9 | 9 | 10 | 1 | 4 |
| Yeast | 11 | 15 | 18 | 18 | 21 | 2 | 9 |
| Other molecules | 5 | 6 | 6 | 6 | 7 | 7 | 7 |
| Carbohydrates | 5 | 6 | 6 | 6 | 7 | 7 | 7 |
| Compounds | 2 | 6 | 12 | 1 | 14 |
| Databases | 26 | 26 | 25 | 1 | 26 |
| Enzymes | 2 | 2 |
| Metabolites | 3 | 4 | 7 | 12 | 1 | 12 |
| Peptides | 1 | 3 | 6 | 6 | 9 | 13 |
| Small molecules | 116 | 116 | 115 | 2 | 117 |
| Proteins | 29 | 51 | 60 | 63 | 65 | 1 | 65 |
| 2D Structure prediction | 53 | 70 | 85 | 100 | 106 | 1 | 107 |
| 3D Structural features | 17 | 20 | 23 | 25 | 28 | 3 | 30 |
| 3D Structure comparison | 48 | 59 | 70 | 83 | 90 | 2 | 93 |
| 3D Structure retrieval, viewing | 28 | 45 | 51 | 56 | 58 | 3 | 60 |
| Annotation and function | 26 | 35 | 44 | 53 | 57 | 5 | 62 |
| Biochemical features | 25 | 37 | 40 | 46 | 46 | 2 | 47 |
| Databases | 116 | 116 | 115 | 2 | 117 |
| Do-it-all tools for protein | 6 | 8 | 9 | 14 | 15 | 3 | 12 |
| Domains and motifs | 50 | 86 | 112 | 121 | 124 | 14 | 115 |
| Identification, presentation and format | 12 | 13 | 14 | 14 | 14 | 1 | 13 |
| Networks & Interactions, pathways, enzymes | 19 | 30 | 38 | 39 | 41 | 3 | 39 |
| Localization and targeting | 28 | 36 | 44 | 53 | 54 | 5 | 47 |
| Molecular dynamics and docking | 7 | 8 | 8 | 10 | 10 | 1 | 7 |
| Phylogeny reconstruction | 21 | 25 | 27 | 34 | 40 | 3 | 44 |
| Protein expression | 5 | 7 | 8 | 10 | 10 | 1 | 7 |
| Proteomes | 21 | 25 | 27 | 34 | 39 | 3 | 43 |
| Sequence comparison | 12 | 15 | 17 | 19 | 21 | 2 | 18 |
| Sequence data | 14 | 15 | 17 | 19 | 21 | 2 | 18 |
| Sequence features | 21 | 27 | 29 | 31 | 32 | 1 | 28 |
| RNA Databases | 0 | 24 | 3 | 25 | 23 | 2 | 25 |
| Functional RNAs | 9 | 14 | 19 | 26 | 32 | 37 | 5 | 39 |
| General resources | 8 | 10 | 10 | 10 | 11 | 4 | 3 |
| Motifs | 9 | 19 | 21 | 22 | 23 | 25 | 1 | 26 |
| Sequence retrieval | 11 | 11 | 10 | 11 | 11 | 9 | 2 |
| Structure prediction, visualization and design | 24 | 38 | 47 | 54 | 62 | 0 | 65 |
| Sequence comparison | 14 | 20 | 21 | 23 | 25 | 0 | 25 |
| Analysis of aligned sequences | 27 | 43 | 59 | 60 | 64 | 4 | 65 |
| Comparative genomics | 19 | 26 | 33 | 35 | 37 | 4 | 47 |
| Multiple sequence alignments | 6 | 11 | 11 | 11 | 12 | 0 | 13 |
| Pairwise sequence alignments | 11 | 22 | 23 | 26 | 33 | 35 | 1 |
| Similarity searching | 13 | 31 | 47 | 49 | 50 | 1 | 49 |
| Bioinformatics Links Directory | 700+ | 1000+ | 1100+ | 1200+ | 1400+ | 1500 | 144 | 1874 |

* A complete listing of all URLs listed in the *Nucleic Acids Research* 2012 Web Server Issue can be accessed online at: http://bioinformatics.ca/links_directory/narweb2012.
ACKNOWLEDGEMENTS

The Bioinformatics Links Directory is a community resource built on a commitment to the spirit of open access under the Attribution-Share Alike 2.5 Canada (CC BY-SA 2.5) license. The authors acknowledge the efforts of Nucleic Acids Research and the researchers and developers worldwide who invest considerable effort into ensuring that their research is freely accessible to all. In particular, the authors acknowledge all the contributors to the Bioinformatics Links Directory for their valuable input and suggestions for improvements to the directory; these individuals are listed on the Acknowledgements page at http://bioinformatics.ca/links_directory/acknowledgements/. In particular, they recognize the continued support of the OICR Web Development group, especially Joseph T. Yamada.

FUNDING

Ontario Institute for Cancer Research and the government of Ontario, and from Genome Canada and the Ontario Genome Institute though a platform award to the The Center of Applied Genomics (TCAG) at the Hospital for Sick Children (SickKids). Funding for open access charge: Waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Cannata,N., Merelli,E. and Altman,R.B. (2005) Time to organize the bioinformatics resourceome. PLoS Comput. Biol., 1, e76.
2. Fox,J.A., Butland,S.L., McMillan,S., Campbell,G. and Ouellette,B.F. (2005) The Bioinformatics Links Directory: a compilation of molecular biology web servers. Nucleic Acids Res., 33, W3–W24.
3. Fox,J.A., McMillan,S. and Ouellette,B.F. (2006) A compilation of molecular biology web servers: 2006 update on the Bioinformatics Links Directory. Nucleic Acids Res., 34, W3–W5.
4. Fox,J.A., McMillan,S. and Ouellette,B.F. (2007) Conducting research on the web: 2007 update for the bioinformatics links directory. Nucleic Acids Res., 35, W3–W5.
5. Brazas,M.D., Fox,J.A., Brown,T., McMillan,S. and Ouellette,B.F. (2008) Keeping pace with the data: 2008 update on the Bioinformatics Links Directory. Nucleic Acids Res., 36, W2–W4.
6. Brazas,M.D., Yamada,J.T. and Ouellette,B.F. (2009) Evolution in bioinformatic resources: 2009 update on the Bioinformatics Links Directory. Nucleic Acids Res., 37, W3–W5.
7. Brazas,M.D., Yamada,J.T. and Ouellette,B.F. (2010) Providing web servers and training in Bioinformatics: 2010 update on the Bioinformatics Links Directory. Nucleic Acids Res., 38, W3–W6.
8. Brazas,M.D., Yim,D.S., Yamada,J.T. and Ouellette,B.F. (2011) The 2011 Bioinformatics Links Directory update: more resources, tools and databases and features to empower the bioinformatics community. Nucleic Acids Res., 39, W3–W7.
9. Google Scholar. http://scholar.google.ca/ (10 June 2012, date last accessed).
10. Google+. http://www.google.com/+1/button/ (10 June 2012, date last accessed).