

Evolution in bioinformatic resources: 2009 update on the Bioinformatics Links Directory

Michelle D. Brazas, Joseph Tadashi Yamada and B. F. Francis Ouellette*

Ontario Institute for Cancer Research, 101 College Street, Suite 800, Toronto, Ontario, Canada M5G 0A3

Received June 3, 2009; Revised June 4, 2009; Accepted June 5, 2009

ABSTRACT

All of the life science research web servers published in this and previous issues of *Nucleic Acids Research*, together with other useful tools, databases and resources for bioinformatics and molecular biology research are freely accessible online through the Bioinformatics Links Directory, http://bioinformatics.ca/links_directory/. Entirely dependent on user feedback and community input, the Bioinformatics Links Directory exemplifies an open access research tool and resource. With 112 websites featured in the July 2009 Web Server Issue of *Nucleic Acids Research*, the 2009 update brings the total number of servers listed in the Bioinformatics Links Directory close to an impressive 1400 links. A complete list of all links listed in this *Nucleic Acids Research* 2009 Web Server Issue can be accessed online at http://bioinformatics.ca/links_directory/narweb2009/. The 2009 update of the Bioinformatics Links Directory, which includes the Web Server list and summaries, is also available online at the *Nucleic Acids Research* website, <http://nar.oxfordjournals.org/>.

COMMENTARY

For 7 years now, *Nucleic Acids Research* has peer-reviewed and published a Web Server issue highlighting the latest web servers and open access bioinformatic tools available online to guide and aid research work. This year's Web Server Issue introduces an additional 112 web servers, 10 of which are server updates (Table 1). The complete listing of URLs cited in the 2009 Web Server Issue can be accessed online at the *Nucleic Acids Research* website, <http://nar.oxfordjournals.org/>, as well as at http://bioinformatics.ca/links_directory/narweb2009/.

In 2002, the Bioinformatics Links Directory was created as a repository for curated URLs to important bioinformatic tools, databases and general-purpose resources for bioinformatics and molecular biology research. Since 2005, the Bioinformatics Links Directory has partnered

with *Nucleic Acids Research* to organize all of the Web Server issue published links in its comprehensive directory (1–4). As a public repository, the Bioinformatics Links Directory, http://bioinformatics.ca/links_directory/, serves as a 'go-to' site for the research community seeking bioinformatic resource options. The 2009 *Nucleic Acids Research* Web Server update brings the total number of servers and tools listed in the Bioinformatics Links Directory close to 1400 unique links (Table 1). Each entry in the Directory contains a short description of the tool's function, as well as the accompanying PubMed citation and web server URL. The directory is organized by biological subject with subcategories of common tasks relevant to each subject listed. Subject categories and subcategories are also easily browsed and queried with a keyword search. The up-to-date complete listings accessible through the Bioinformatics Links Directory, including the *Nucleic Acids Research* 2009 web servers, can be accessed online at http://bioinformatics.ca/links_directory/.

Reflecting back on the history of the *NAR* Web Server issue and the Bioinformatics Links Directory, and on informatics over the past few years, it is interesting to note how research and the field of bioinformatics have changed in such a short timeframe. Table 1 displays data on the number of web server URLs for each category and subcategory of the Bioinformatics Links Directory for the years 2006–2009. Many interesting trends reflective of parallel changes in research areas and technologies are observable. In many instances there are sharp spikes in the number of informatics tools available in a given biological area or specific to a particular task. For example, the number of tools for 'DNA – Structure and Sequence Feature Detection', 'Gene Regulation' and 'Transcript Expression and Microarrays' increased sizably between 2006 and 2007, most likely in response to the pervasive use of expression platforms and transcript analyses at that time (5,6) (Table 1). Similarly, as the research focus on small non-coding RNAs, microRNAs and other RNA species has intensified, so too has the number of bioinformatics web servers for 'Functional RNA' analysis, particularly in 2008 and 2009 (Table 1). A steady increase in the number of useful informatics resources for 'Domain

*To whom correspondence should be addressed. Tel: +1 416 673 8511; Fax: +1 416 977 7446; Email: francis@oicr.on.ca

Table 1. Historical summary (2006–2009) of the number of web servers listed in each subcategory of the Bioinformatics Links Directory

Name	2006	2007	2008	2009 ^a
Computer-related				
Bio-* Programming Tools	20	20	20	20
C/C++	3	3	3	3
Databases	2	2	2	3
Java	4	4	4	4
Linux/Unix	12	11	11	11
PERL	5	5	5	5
PHP	3	1	1	1
Statistics	9	9	9	9
Web Development	6	2	2	2
Web Services	6	7	7	10
DNA				
Annotations	38	56	57	62
Gene Prediction	32	33	34	37
Mapping and Assembly	14	15	15	17
Phylogeny Reconstruction	37	43	46	49
Structure and Sequence Feature Detection	118	142	145	150
Sequence Polymorphisms	32	39	41	42
Sequence Retrieval and Submission	26	30	32	32
Tools For the Bench	55	63	65	71
Utilities	19	20	23	24
Education				
Bioinformatics Related News Sources	9	9	9	9
Community	19	24	23	23
Courses, Programs and Workshops	5	5	5	5
Directories and Portals	15	15	15	15
General	15	14	14	14
Tutorials and Directed Learning Resources	9	9	9	9
Expression				
cDNA, EST, SAGE	29	36	44	48
Gene Regulation	96	119	120	128
Transcript Expression and Microarrays	75	89	101	108
Protein Expression	8	9	17	22
Splicing	16	19	19	21
Networks			8	12
Gene Set Analysis				11
Human Genome				
Annotations	31	37	38	39
Ethics	7	8	8	8
Genomics	4	3	3	10
Health and Disease	14	19	23	27
Other Resources	25	29	29	29
Sequence Polymorphisms	25	33	36	38
Literature				
Goldmines	6	6	6	6
Open Access Resources	2	2	2	3
Search Tools	10	12	12	13
Text Mining	11	15	22	30
Model Organisms				
Fish	11	11	11	11
Fly	16	17	17	17
General Resources	23	27	28	29
Microbes	31	38	45	53
Mouse and Rat	32	35	35	36
Other Organisms	18	21	21	21
Other Vertebrates	10	10	10	10
Plants	16	19	21	25
Worm	9	9	9	9
Yeast	15	18	18	18
Other Molecules				
Carbohydrates	6	6	6	6
Metabolites		3	4	7
Small Molecules	3	6	6	9
Compounds			2	6

(continued)

Table 1. Continued

Name	2006	2007	2008	2009 ^a
Protein				
2-D Structure Prediction	51	58	60	63
3-D Structural Features	53	70	75	85
3-D Structure Comparison	35	45	50	59
3-D Structure Prediction	48	59	60	70
3-D Structure Retrieval, Viewing	45	51	52	56
Biochemical Features	37	40	41	46
Do-it-all Tools for Protein	8	8	13	14
Domains and Motifs	86	112	115	121
Function and Annotation	35	44	47	53
Interactions, Pathways, Enzymes	66	88	94	107
Localization and Targeting	30	38	38	39
Molecular Dynamics and Docking	19	21	27	34
Phylogeny Reconstruction	36	44	45	53
Presentation and Format	13	14	14	14
Protein Expression	8	8	8	10
Proteomics	25	27	33	37
Sequence Data	7	8	9	10
Sequence Comparison			7	14
Sequence Features	25	31	33	38
Sequence Retrieval	27	29	29	31
RNA				
Functional RNAs	14	19	26	32
General Resources	10	10	10	10
Motifs	19	21	22	23
Sequence Retrieval	11	10	11	11
Structure Prediction, Visualization, and Design	38	47	54	58
Sequence Comparison				
Alignment Editing and Visualization	20	21	21	23
Analysis of Aligned Sequences	43	59	60	62
Comparative Genomics	26	33	35	37
Multiple Sequence Alignments	38	50	56	57
Other Alignment Tools	11	11	11	11
Pairwise Sequence Alignments	22	23	26	33
Similarity Searching	31	47	47	49

^aA complete listing of all URLs listed in the *Nucleic Acids Research* 2009 Web Server Issue can be accessed online at: http://bioinformatics.ca/links_directory/narweb2009/

and Motif' analysis, 'Interactions, Pathways and Enzyme' analysis and 'Molecular Dynamics and Docking' since 2006 has also been observed, in part due to the increased need for tools to aid researchers evaluating biology from a pathway or whole organism perspective. Indeed in 2008, a new subcategory for 'Network' analysis was added to accommodate informatics tools and resources that take a systems approach to biological problems.

Prominent in research today has been the introduction and application of various next-generation sequencing platforms to a wide range of research problems from rapid de novo sequencing (7), to genomic evaluation of polymorphisms and other mutation events (8), and analyses of the transcriptome of various RNA macromolecules (9) and the epigenome (10). Given the quantity and variety of data being produced by these new high-throughput technologies, there is an increased need for tools in meta-data analysis, data integration and data mining that facilitate data exploration and provide opportunities for novel discoveries without the concomitant requirement for personal computational capacity. Accordingly, web servers introduced in the 2009 *NAR* Web Server Issue include several specialized web service interfaces such as

those from the European Bioinformatics Institute (<http://www.ebi.ac.uk/Tools/webservices/>) and the DNA Data Bank of Japan (<http://www.xml.nig.ac.jp/index.html>) which provide remote computational power, as well as several data integration packages such as BioBIKE (<http://biobike.csbc.vcu.edu/>), which 'enable biologists with little programming expertise to combine tools, data and knowledge'. Also introduced in this issue, were a host of online informatic resources for advanced text/data mining such as SENT (<http://sent.dacya.ucm.es/>) for detection of semantic features in text and LitInspector (<http://www.litinspector.org/>) for literature and signal transduction pathway mining in PubMed.

New to the Bioinformatics Links Directory in 2009 is the addition of 'Gene Set Analysis', a new subcategory under tools for 'Expression' analysis (Table 1), addressing the need and introduction of new resources that facilitate data integration and functional evaluation of gene sets identified through expression or genome wide association studies. Example web servers include ToppGene Suite (<http://toppgene.cchmc.org/>) for functional enrichment, candidate gene prioritization and identification of novel disease genes in an interactome, and GeneCodis (<http://genecodis.dacya.ucm.es/>) for functional analysis of gene lists that integrates different data sources (microRNA, transcription factors, KEGG, user annotations, etc.) and finds patterns of interrelated annotations.

These historical perspectives are not unique or isolated occurrences. As new technologies and applications continue to be introduced and as new data become available, we will continue to see the development and update of specialized web servers and bioinformatic resources to accompany emerging, data intensive and biologically complex research needs. The Bioinformatics Links Directory, as an excellent example of a community-driven resource, will also continue to expand its resources in support of the research community seeking bioinformatic resource options. Suggestions for new links or updates and corrections to existing links at the Bioinformatics Links Directory are always welcome, and may be submitted through email directly to links@bioinformatics.ca.

ACKNOWLEDGEMENTS

The authors wish to 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. The Bioinformatics Links Directory is a community resource

built on this commitment to the spirit of open access. In particular, the authors would like to acknowledge all of 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/.

FUNDING

This work was conducted with the support of the Ontario Institute for Cancer Research through funding provided by the Government of Ontario. The Open Access publication charge for this paper has been waived by the Oxford University Press, in recognition of the work on behalf of the journal.

Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. 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.
3. 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.
4. 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.
5. Aharoni, A. and Vorst, O. (2002) DNA microarrays for functional plant genomics. *Plant Mol. Biol.*, **48**, 99–118.
6. Majtan, T., Bukovska, G. and Timko, J. (2004) DNA microarrays – techniques and applications in microbial systems. *Folia Microbiol. (Praha)*, **49**, 635–664.
7. Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T. *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**, 872–876.
8. Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, J.I., Hickenbotham, M., Huang, W. *et al.* (2008) Whole-genome sequencing and variant discovery in *C. elegans*. *Nat. Methods*, **5**, 183–188.
9. Fullwood, M.J., Wei, C.L., Liu, E.T. and Ruan, Y. (2009) Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.*, **19**, 521–532.
10. Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M. and Jacobsen, S.E. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, **452**, 215–219.