ROBERT W. COTTINGHAM

117 Westview Lane

Oak Ridge, TN 37830

bob@cottingham.org

865-275-5125

<https://scholar.google.com/citations?hl=en&user=1PWI1ssAAAAJ>

<https://orcid.org/0000-0002-2775-3905>

**Education**: BS Mechanical Engineering, University of Washington, 1973

# Employment Experience:

2008 - Oak Ridge National Laboratory, Oak Ridge, TN.

**Group Leader**

Oversee the Computational Biology and Bioinformatics group in the Biosciences Division with a staff that includes expertise in biology, software engineering and computer science. The group conducts collaborative research developing and applying tools of data science, predictive modeling and high-performance computing to transform biological data into knowledge. Its work is funded primarily under Department of Energy Genomic Sciences programs, and also applies its expertise and capabilities toward biomedical and biosecurity efforts funded by other agencies such as NIH, CDC and NSF, and in collaborations with University researchers. Current major efforts are under The Center for Bioenergy, the Plant Microbe Interfaces project, and the DOE Systems Biology Knowledgebase with foci on various aspects of genome and proteome analysis and molecular systems biology. Elements of the effort include genome analysis and annotation; multiscale data management and integration; collaboratory systems; molecular modeling and simulation; mass spectrometry analysis; metabolic and regulatory modeling; high-performance biocomputing; computational biology graduate training.

2007 - 2008 Consultant, Redmond, WA.

**Consultant**

Provided business and management consulting primarily to one genomics startup company creating cell-based assays for drug discovery and target validation.

2000 - 2006 VizX Labs, Seattle, WA.

**Co-Founder and Chief Technology Officer**

Startup company (12 employees, 250 customers, subsequently sold to Geospiza in 2008, now owned by Perkin Elmer) providing bioinformatics solutions for microarray based research. Products based on the latest analysis and computational tools, and designed to be easy for lab bench researchers to use to reach biologically relevant conclusions with minimal support requirements. Oversaw software development, production deployment and infrastructure, while also assisting with the business strategy, planning and overall management. Received a Washington Technology Center grant for improvement of microarray design toward clinical diagnostic applications.

1997 - 2000 Celltech Chiroscience Inc., Bothell, WA.

**Vice President, Computing**

Oversaw all computational activities in the US and UK sites including bioinformatics, software development and systems administration. Directed development of an integrated computing infrastructure across the company. Guided transition toward improved software development technologies and methods. Provided scientific vision, technical strategy, and managerial direction. (Celltech subsequently merged with UBC SA in 2004).

1995 - 1997 John Hopkins University School of Medicine, Baltimore, MD.

**Assistant Professor**

**Operations Director** of the Genome Database that manages and curates data for the international Human Genome Project and related genomics research. Oversaw substantial expansion of the database content and a technical transition from a strict relational database to a more powerful object-relational form. This resulted in a more comprehensive and manageable database using automated tools for data modeling more complex biological concepts.

1991- 1995 Baylor College of Medicine, Houston, TX.

**Instructor**

**Co-Director**, Informatics Core in the BCM Human Genome Center. Developed computing infrastructure and staff organization. Manage projects and schedules for the Core activities. Oversaw the development of the databases, user interfaces and analysis tools for the Center. Coordinate interaction and collaboration with other BCM support and client groups, and external Centers and informatics groups working on the Human Genome Project.

Improved algorithms for genetic linkage analysis in humans. Developed the FASTLINK version of the LINKAGE package using new algorithmic approaches resulting in more than 1000 fold speed improvement and thereby enabling analysis of more complex genetics. Genotype error detection using maximum likelihood methods. Consulting and analysis of genetic linkage for construction of maps and localization of diseases.

Organize and teach quantitative genetics portion of the Human Genetics II course. Also teach linkage analysis and physical mapping portion of the Computational Methods in Molecular Biology course.

1989 – 1991 Centre d'Etude du Polymorphisme (CEPH), Paris, France.

###### Directeur Informatique

Managed computing activities at CEPH (now Fondation Jean Dausset-CEPH) as the first human genetic maps were developed. Designed and developed software to manage the CEPH Database, a database of human genetic markers. Implemented wide area networking to provide remote database access to more than 1000 CEPH collaborators worldwide. Provide consultation and assistance to users of the LINKAGE package of analysis programs used to create the genetic linkage maps. Developed software and algorithm improvements to the LINKAGE package. Reviewed autoradiogram image analysis products and initiated project for automation of genotype calling.

1987 – 1989 Control Applications (Teledyne), Dallas, TX.

##### Principal Systems Analyst

Project Management on projects for Texas Utilities Fuel Co., Shell Oil Co., Williams Pipe Line Co., and Seattle Metro. Managed teams of up to 70 software developers. Developed new Historical subsystem for collecting real-time data. Assisted and managed others in the design, coding, and implementation of various other subsystems including User Interface, Report Generation, Automatic Daylight Savings Time. Acted as client interface on each project and chaired monthly project status meetings with clients.

1984 – 1987 Cottingham Software Inc., Houston, TX.

##### Consultant

Consultant to Modular Data Systems in the design and implementation of a new subsystem to archive and retrieve historical data for MDSI’s ADACS system. While consulting to Scientific Software-Intercomp, developed and implemented the trend and profile packages for measured and fluid flow models on the Stand Alone Gas Applications system for Aramco’s natural gas pipeline network in Saudi Arabia.

1981 – 1984 Scientific Software-Intercomp, Houston, TX.

**System Analyst**

Project Manager on a feasibility study of model applications for the Colorado Interstate Gas transmission network. Set up initial integration for team working on the Trans-Siberian pipeline at client’s facilities in Paris. Implemented pipeline fluid flow models on ethylene pipeline for Alberta Gas Ethylene in Canada, and a batch products pipeline for New Zealand Refining Co. Taught one week modeling courses to each client.

1979 – 1981 Control Applications, Houston, TX.

**System Analyst**

Designed and implemented the man machine interface package for Peace Pipeline in Canada. This package was part of a complete Supervisory Control And Data Acquisition system running on a DEC PDP-11 and included device independent display drivers, interactive display building and menu driven command input.

1975 – 1979 University of Washington, Seattle, WA.

**Programmer**

Developed software for the collection, maintenance and analysis of online genetic databases. Worked under Jurg Ott, on LIPED, a genetic linkage analysis program used in the mapping of early human disease gene discoveries.

1973 – 1975 Boeing Computer Services, Seattle, WA.

**Programmer**

Real-time data collection and analysis at the wind tunnel facility.

**Research Interests**:

My career focus has been on the development of computational methods and systems to advance genomics research and its application to improve biological understanding. This has included development and analysis of reporter cell based assays used in genome wide screening applications, microarray-based gene expression analysis, gene discovery as a lead-in to drug development, development of GDB: the Human Genome Database, improved genetic linkage analysis algorithms and software, development of physical maps for the Human Genome Project, and development of informatics tools and methods as part of the international collaboration to create the first detailed genetic maps of the human genome. More recently I led the development of the DOE Systems Biology Knowledgebase Implementation Plan and am currently the ORNL co-PI for the KBase project.

The nature of computing in biology is changing as the volume and complexity of data continues to grow exponentially, and the number of new tools and algorithms similarly grow. Bench biologists and biochemists have become reliant on computational tools to aid their exploration and

prediction of behavior in systems and synthetic biology. DNA sequence alignment and assembly, and

metabolic modeling are a few examples of the many types of analyses that are performed in laboratories on a daily basis. Recent cyberinfrastructure projects have led in making such analysis more accessible and

shareable to the broader scientific community, via frameworks such as the predictive DOE Systems Biology Knowledgebase, KBase, a community informed online platform containing thousands of annotated models, tools, apps and user narratives used to answer fundamental questions about biological ecosystems. Yet, underlying many of the algorithms and tools is a complexity second only to that of the biological systems they explore. These tools are implementations of algorithms that approximate and optimize the biology being studied. They work as magic boxes, encapsulating the logic utilized in their decision making. And they are often highly customizable allowing the end user to modify tens to hundreds of options controlling the heuristics and resource constraints which drive their decision-making along the way. While this customizability produces flexible and general tools that work across a wide range of problems and types of data, it also adds a layer of obfuscation between the tools and the biologist, creating a ***trust gap*** across which results are to be believed and presented as facts. As shown in Cashman2, configuration and optimization of parameter choices can often produce incorrect results that are not obvious. The implications are broad --- ***the conclusions derived from these tools and the size of the space explored within a given time budget can vary greatly***. This leads to a lack of reliability and repeatability, and produce misleading conclusions, ultimately widening the trust gap. Furthermore, as these algorithms explore complex biological domains, often with a singular end objective (e.g. find a sequence, maximize growth), they learn valuable information about the space of all possible solutions, and this information is lost. What is needed are **explainable, observable bioinformatics algorithms and tools that can explore large and varied data sets exposing relevant facts and configuration options that they have used and learned during their exploration**.

Computational systems that guide research: Researchers frequently focus on limited results from experiments because of interest or bias. However, emerging technologies produce very large data sets that usually present many discovery opportunities. Unfortunately, typical computing analysis systems today are preprogrammed with specific analytical methods and little interaction or feedback from the research user and thus limit the discovery opportunities. This approach must be changed to one that involves the researcher not only in using the system but also in its design. Computational systems are in fact embodiments of biological or medical understanding. Therefore, these systems must be open to scrutiny and modification by the whole research community even as they are being used and as scientific understanding advances. My work continues to be developing more effective and productive approaches toward building computational systems that meet these goals. Model based simulation and inference would provide a basis for more powerful computation directed discovery.

Improve computational infrastructure with data models and standards: Data integration continues to be an important hurdle and opportunity for advancement of biomedical research. It is also a significant part of creating a community computational model of biological systems and provides a more specific and achievable focus. Advancement of systems biology research depends on common data models as a basis for exchange and integration of data. To achieve these standards depends as much on community awareness and initiative as it does on the technical solutions. My work will be to promote awareness, education and technical approaches to such standards and their broad acceptance as an important part of the advancement of research. The DOE Systems Biology Knowledgebase effort is a great opportunity to develop community collaboration and solutions encoded in an open shared platform that is open to scrutiny and modification by the whole research community.

Semantic search: Most biomedical knowledge represented in electronic form is stored as text. Humans can read the text and understand its meaning while to a computer it is just a string of letters. One way to store the meaning electronically is to have knowledgeable curators manually read the text and then enter the associated concepts into a database along with the article. Because of the exponentially growing literature and a lack of human resources there is no complete encoding of literature concepts. The only solution is the development of automatic semantic parsing.

MetaMap developed at the NLM is one such tool that parses concepts from natural language, and Collexis, a commercial tool, is another. There is currently no fully automatic tool. However, despite their shortcomings current methods while not perfect are adequate to provide automated discovery support that would help researchers. My interest is in applying these techniques as part of an integrated toolkit for discovery research to augment traditional analytic methods.

**Research Experience**:

Computational molecular biology: Focusing on methods for analysis and management of genomic information; mouse-human synteny, orthologous and paralogous relationships, information theoretic approaches to analyzing DNA sequence.

Computational methods for the analysis of genetics: Genetic linkage analysis, error detection, map construction and chromosomal localization of human genetic disease. Error modeling of genetic mapping experiments provided basis for predicting and correcting errors that made disease mapping possible. Improved computational performance of genetic linkage analysis software.

Bioinformatics: Application of computer hardware and software to genome informatics, including numerical methods, algorithmic optimization, data modeling, databases, user-interfaces and object-oriented techniques.

**Computing Experience**:

Software Engineering: CVS, UML, Rational Rose, Purify, ClearCase,

Web Services Programming and Administration: Apache, Perl modules, HTML, XML

Database Programming and Administration: Adabas, Sybase, Oracle, ObjectStore, mySQL

Programming Languages: BASIC, FORTRAN, Pascal, Simula, Smalltalk, C, C++, Perl, Java

Network Administration: Ethernet, BitNet, TCP/IP, DNS, DHCP, Cisco/Linksys routers, firewalls, security

System Administration: Unix, Linux, Windows

**Teaching Experience**:

Computational Methods in Molecular Biology (Linkage Analysis portion) 1992-1995

Human Genetics II, Quantitative Genetics, 1993-1995

**Advisory Boards:**

Board of Directors, VizX Labs, 2001-2008

Editorial Board, Briefings in Bioinformatics, 1999-2004

Member of Genetics Committee, Foundation for Fighting Blindness, 1998-2002

**Service**:

Organizer of the BER Plant Genomic Science workshop at the Plant and Animal Genome Conference 2019-2020

Organizer of the Annual KBase User Meeting 2015-2020

Organizing Committee member for the ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, 2016-2019

Invited participant in DOE Office of Biological and Environmental Research Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and Environment workshop on September 21-23, 2016

Invited participant on the CDC Blue Ribbon Panel to review the status of Bioinformatics at CDC, 2011

Invited presentation to the EPA Office of Research and Development Board of Scientific Counselors, 2010

Invited keynote speaker at the International Conference on Bioinformatics and Computational Biology (ACM-BCB 2010)

Invited participant in DOE Office of Biological and Environmental Research Long Term Vision Workshop, 2010

Invited keynote speaker for the 7th Annual Conference of the MidSouth Computational Biology and Bioinformatics Society (MCBIOS 2010)

Organizing Committee member for the DOE Knowledgebase System Development Workshop, June 1–3, 2010

Organizing Committee member for the DOE Workshop on Cloud Computing in Systems and Computational Biology at SuperComputing 2009.

Reviewer of NIDDK consortium grant, 2007

Site Visitor for the Complex Biological Systems Initiative Review at Oak Ridge National Laboratory, 2002

Site Visitor for Joint Genome Institute Production Sequencing Facility Review, 2000

Reviewer of Joint Genome Institute collaboration grant, 1999

Reviewer for DOE Human Genome Informatics grant applications, 1999

Site visitor of Lawrence Berkeley National Laboratory Computing Sciences Division, 1999

Reviewer for DOE Joint Genome Institute Functional Genomics grant applications, 1998

Invited participant in DOE/NIH Informatics Workshop to advise on HGP Informatics direction, 1998

Site visitor of Genome Annotation Consortium application to DOE, ORNL, 1998

Site visitor for NSF Macromolecular Structure Database center applications, 1998

Invited presentation for DOE Jasons review of Human Genome Project, 1997

Site visitor of the Human Genome Mapping Project Resource Center renewal application to the MRC, Hinxton, 1997

Panelist for DOE ELSI Study Section, 1997-1998

Organized Genome Database Editor Training, Heidelberg, 1996

Reviewer for NIH Research Resource grant applications, 1995-1996

Reviewer for DOE on GDB application, 1994

Reviewer for DOE on OMIM application, 1994

Reviewer for NSF Informatics and Database grant applications 1993-1995

Informatics support for Chromosome 6,8,17, X Workshops, 1992-1995

# Professional Associations:

SIGBio - Special Interest Group of ACM for Bioinformatics, Computational Biology, and Health Informatics., 2010-

International Society for Computational Biology, 1998-

American Society for Human Genetics, 1974-2000

Association for Computing Machinery, 1972-

American Association for the Advancement of Science, 1968-

**Publications:**

1. Samuel MD Seaver, Filipe Liu, Qizhi Zhang, James Jeffryes, Jose P Faria, Janaka Edirisinghe, Michael Mundy, Nicholas Chia, Elad Noor, Moritz E Beber, Aaron A Best, Matthew DeJongh, Jeffrey A Kimbrel, Patrik D'haeseleer, Erik Pearson, Shane Canon, Elisha M Wood-Charlson, Robert W Cottingham, Adam P Arkin, Christopher S Henry. The ModelSEED Database for the integration of metabolic annotations and the reconstruction, comparison, and analysis of metabolic models for plants, fungi, and microbes. Nucleic Acids Research (accepted, Jan 2021, IF 11.5).

2. Cashman M, Cohen MB, Ranjan P, Cottingham RW. Navigating the maze: the impact of configurability in bioinformatics software. Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering; 2018: ACM. **(Won ACM SIGSOFT Distinguished Paper Award)**

3. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: the United States department of energy systems biology knowledgebase. Nature Biotechnology. 2018;36(7), IF 33.8.

4. Yang X, Cushman JC, Borland AM, Edwards EJ, Wullschleger SD, Tuskan GA, et al. A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. New Phytologist. 2015;207(3):491-504.

5. Lee LL, Izquierdo JA, Blumer-Schuette SE, Zurawski JV, Conway JM, Cottingham RW, et al. Complete Genome Sequences of Caldicellulosiruptor sp. Strain Rt8. B8, Caldicellulosiruptor sp. Strain Wai35. B1, and “Thermoanaerobacter cellulolyticus”. Genome Announcements. 2015;3(3):e00440-15.

6. Brettin TS, Cottingham RW, Griffith SD, Quest DJ. Scenario driven data modelling: a method for integrating diverse sources of data and data streams. US Patent 9,129,039; 2015.

7. Elkins JG, Hamilton-Brehm SD, Lucas S, Han J, Lapidus A, Cheng J-F, et al. Complete genome sequence of the hyperthermophilic sulfate-reducing bacterium Thermodesulfobacterium geofontis OPF15T. Genome Announcements. 2013;1(2):e00162-13.

8. Mavromatis K, Land ML, Brettin TS, Quest DJ, Copeland A, Clum A, et al. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. PLoS One. 2012;7(12):e48837.

9. Brettin TS, Cottingham RW, Quest DJ. Screening tool for providers of synthetic double stranded dna. US Patent App. 14/355,767; 2012.

10. Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin Y, et al. Caldicellulosiruptor core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. Journal of Bacteriology. 2012;194(15):4015-28.

11. Griffith SD, Quest DJ, Brettin TS, Cottingham RW. Scenario driven data modelling: a method for integrating diverse sources of data and data streams. BMC Bioinformatics. 2011;12(10):1.

12. Quest DJ, Land ML, Brettin TS, Cottingham RW. Next generation models for storage and representation of microbial biological annotation. BMC Bioinformatics. 2010;11(6):1.

13. Elkins JG, Lochner A, Hamilton-Brehm SD, Davenport KW, Podar M, Brown SD, et al. Complete genome sequence of the cellulolytic thermophile Caldicellulosiruptor obsidiansis OB47T. Journal of Bacteriology. 2010;192(22):6099-100.

14. Letovsky SI, Cottingham RW, Porter CJ, Li PW. GDB: the human genome database. Nucleic Acids Research. 1998;26(1):94-9, IF 11.5.

15. Fasman KH, Letovsky SI, Li P, Cottingham RW, Kingsbury DT. The GDB™ Human Genome Database Anno 1997. Nucleic Acids Research. 1997;25(1):72-80, IF 11.5.

16. Cottingham Jr RW, Ehm MG, Kimmel M. Error Analysis of Genetic Linkage Data. Theoretical and Computational Methods in Genome Research: Springer US; 1997. p. 135-43.

17. Fasman KH, Letovsky SI, Cottingham RW, Kingsbury DT. Improvements to the GDB™ human genome data base. Nucleic Acids Research. 1996;24(1):57-63, IF 11.5.

18. Ehm MG, Kimmel M, Cottingham Jr RW. Error detection for genetic data, using likelihood methods. American Journal of Human Genetics. 1996;58(1):225.

19. Ehm MG, Kimmel M, Cottingham RW. Error detection in genetic linkage data for human pedigrees using likelihood ratio methods. Journal of Biological Systems. 1995;3(01):13-25.

20. Vora DK, Rosenbloom CL, Beaudet AL, Cottingham RW. Polymorphisms and linkage analysis for ICAM-1 and the selectin gene cluster. Genomics. 1994;21(3):473-7.

21. Volz A, Boyle JM, Cann HM, Cottingham RW, Orr HT, Ziegler A. Report of the second international workshop on human chromosome 6. Genomics. 1994;21(2):464-72.

22. Schaeffer AA, Gupta SK, Shriram K, Cottingham Jr RW. Avoiding recomputation in linkage analysis. Human heredity. 1994;44(4):225-37.

23. Dwarkadas S, Schäffer AA, Cottingham Jr RW, Cox AL, Keleher P, Zwaenepoel W. Parallelization of general-linkage analysis problems. Human Heredity. 1994;44(3):127-41.

24. Cottingham Jr RW, Idury RM, Schäffer AA. Faster sequential genetic linkage computations. American Journal of Human Genetics. 1993;53(1):252, Cited 1523.

25. Law A, Richard III CW, Cottingham Jr RW, Lathrop GM, Cox DR, Myers RM. Genetic linkage analysis of bipolar affective disorder in an Old Order Amish pedigree. Human Genetics. 1992;88(5):562-8.

26. Huang TH-M, Cottingham RW, Ledbetter DH, Zoghbi HY. Genetic mapping of four dinucleotide repeat loci, DXS453, DXS458, DXS454, and DXS424, on the X chromosome using multiplex polymerase chain reaction. Genomics. 1992;13(2):375-80.

27. Beckmann J, Richard I, Hillaire D, Broux O, Antignac C, Bois E, et al. A gene for limb-girdle muscular dystrophy maps to chromosome 15 by linkage. Comptes rendus de l'Académie des sciences Série III, Sciences de la vie. 1990;312(4):141-8.