

Variable selection and parameter estimation for spatially varying coefficient models

Wesley Brooks

Department of Statistics
University of Wisconsin–Madison

somesquares.org

Access in action

Interesting reference

- [8] Kang C. and Speller R. The effect of region of interest selection on dual energy x-ray absorptiometry emasurements of the calcaneus in 55 post-menopausal women. *The british Journal of Radiology*, 72:864–871, 1999.
- [9] The 1000 Genomes Project Consortium. A map of human genome variation from population scale sequencing. *Nature*, 467:1061–1073, 2010.
- [10] John C.V., Mark D.A., Eugene W.M., Peter W.L., Richard J.M., Granger G.S., and Hamilton O.S. The sequence of the human genome. *Science*, 291:1304–1351, 2001.
- [11] Schwartz D.C. and Waterman M.S. New generations: sequencing machines and their computational challenges. *Journal of Computer Science and Technology*, 25(1):3–9, 2010.
- [12] Church D.M., Goodstadt L., Hillier L.W., Zody M.C., Goldstein S., She X., Bult C.J., Agarwala R., Cherry J.L., DiCuccio M., Hlavina W., Kapustin Y., Meric P., Maglott D., Birtle Z., Marques A.C., Graves T., Zhou S., Teague B., Potamousis K., Churas C., Place M., Herschleb J., Runnheim R., Forrest D., Amos-Landgraf J., Schwartz D.C., Cheng Z., Lindblad-Toh K., Eichler E.E., and Ponting C.P. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biology*, 7.5:e1000112, 2009.
- [13] Tor D.W., Matthew C.K., Steven C.L., and John J. Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage*, 26:99–113, 2005.

Access in action

Google Scholar

The screenshot shows a Google Scholar search results page. The search query is "New generations: sequencing machines and their computational challenges". There are approximately 20,200 results found in 0.12 seconds. The results are listed in chronological order from most recent at the top.

1. New generations: Sequencing machines and their computational challenges (nih.gov [HTML])

DC Schwartz, MS Waterman - Journal of computer science and technology, 2010 - Springer

Schwartz DC, Waterman MS. **New generations: Sequencing machines and their computational challenges**. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 25(1): 3–9 Jan. 2010

... New Generations: Sequencing Machines and Their Computational

Cited by 13 Related articles All 14 versions Cite

2. [BOOK] Evolutionary computation: toward a new philosophy of machine intelligence

DB Fogel - 2006 - books.google.com

... If the process is understood, methods for its generation should converge functionally and become fundamentally identical, relying on ... of computing, we have envisioned **machines** that could go beyond our own ability to solve problems—intelligent **machines**. ... This is nothing **new**. ...

Cited by 2874 Related articles All 14 versions Cite More ▾

3. [HTML] How to map billions of short reads onto genomes (nih.gov [HTML])

C Trapnell, SL Salzberg - Nature biotechnology, 2009 - ncbi.nlm.nih.gov

... As a practical matter, the task of mapping billions of **sequences** to a mammalian-sized ... To reduce the computing cost of analysis for **sequencing**-based assays and to make them available to all investigators, we and others have created a **new generation** of alignment ...

Cited by 142 Related articles All 26 versions Cite

Access in action

Google Scholar

New generations: sequencing machines and their computational challenges

Scholar

New gene

DC Schwartz,
Schwartz DC,
challenges. J
... New Gener

Cited by 13

[BOOK] Evolu

DB Fogel - 20
... If the proce
fundamentally
beyond our ov

Cited by 2874

[HTML] How

C Trapnell, SL

... As a practi
the computing
investigators,

Cited by 142

Journal of Computer Science and Technology
January 2010, Volume 25, Issue 1, pp 3-9

New Generations: Sequencing Machines and Their Computational Challenges

David C. Schwartz, Michael S. Waterman

Purchase on Springer.com
\$39.95 / €34.95 / £29.95 *
Buy now

* Final gross prices may vary according to local VAT.

Get Access

Abstract

New generation sequencing systems are changing how molecular biology is practiced. The widely promoted \$1000 genome will be a reality with attendant changes for healthcare, including personalized medicine. More broadly the genomes of many new organisms with large samplings from populations will be commonplace. What is less appreciated is the explosive demands on computation, both for CPU cycles and storage as well as the need for new computational methods. In this article we will survey some of these developments and demands.



Look Inside

Share

f t in

Other actions

» Export citations

» Register for Journal Updates

» About This Journal

» Reprints and Permissions

What's the deal with the prices?

Broman K, Speed T, Tigges M (1996) Estimation of antigen-responsive T cell frequencies in PBMC from human subjects. J Immunol Meth 198:119–132	\$39.95
Broman KW, Weber JL (1999) Method for constructing confidently ordered linkage maps. Genet Epidemiol 16:337–343	\$35.00
Broman KW, Feingold E (2004) SNPs made routine. Nat Methods 1:104–105	\$18.00
Broman KW (2005) Mapping expression in randomized rodent genomes. Nat Genet 37:209–210	\$18.00

Access in action

journal.com.ezproxy.library.wisc.edu/blah

The screenshot shows the University of Wisconsin-Madison Libraries website. At the top, there's a red header bar with the university logo and the word "Libraries". Below the header, a main content area has a yellow background. It features a section titled "Remote Access to Library Resources" and a sub-section titled "Login With Your NetID". Under "Login With Your NetID", there are two bullet points: one about logging in giving access to restricted resources, and another about calling the Help Desk for assistance. Below this, there's a "Help" section with three bullet points linking to "NetID - Account Utilities", "NetID - Eligibility Policy", and "Help with Remote Access/Proxy Service for Electronic Resources". At the bottom of the page, there's a footer with links to "Accessibility", "Code of Conduct", and "Contact Us", along with the address "728 State Street, Madison, Wisconsin 53706" and the copyright notice "©2011 Board of Regents of the University of Wisconsin System".

UNIVERSITY OF WISCONSIN-MADISON

Campus Libraries | Mobile | My Accounts ▾

Search this site...

chat | email | phone | desk | ASK

Remote Access to Library Resources

Login With Your NetID

- Logging in will give you access to library resources that are restricted (due to software licenses, copyright restrictions and other contractual agreements) to UW-Madison students, faculty, and staff.
- If you need assistance with your NetID and password, please call the DoIT Help Desk at [608-264-HELP](#) (4357). For security reasons, the Help Desk will only handle NetID and password issues over the phone. The Help Desk is open from 6AM-1AM daily.

Help

- [NetID - Account Utilities](#)
Information on resetting passwords, activating accounts, and modifying accounts.
- [NetID - Eligibility Policy](#)
Please note that eligibility for a UW-Madison NetID and eligibility for remote access to library resources is not equivalent. **Some NetID accounts may not allow access to library resources from off campus.**
- [Help with Remote Access/Proxy Service for Electronic Resources](#)
More information on the UW-Madison Libraries remote access system.

Accessibility • Code of Conduct • Contact Us
728 State Street, Madison, Wisconsin 53706
©2011 Board of Regents of the University of Wisconsin System

Access in action

journal.com.ezproxy.library.wisc.edu/blah

The screenshot shows the University of Wisconsin-Madison Libraries homepage. The top navigation bar includes links for "Campus Libraries", "Mobile", and "My Accounts". A search bar and a "Search" button are also present. On the right, there are links for "chat", "email", "phone", "desk", and a yellow "ASK" button with a question mark icon.

The main content area features a large red banner with the UW System logo and the text "Select your organization". Below this, a message reads: "This service requires you to authenticate with your local organization, please select it from the list below." A list of radio buttons allows selection of a specific UW institution. The "University of Wisconsin - Madison" option is selected, indicated by a grey background. Other options listed include: University of Wisconsin - Colleges, University of Wisconsin - Eau Claire, University of Wisconsin - Extension, University of Wisconsin - Green Bay, University of Wisconsin - La Crosse, University of Wisconsin - Milwaukee, University of Wisconsin - Oshkosh, University of Wisconsin - Parkside, University of Wisconsin - Platteville, University of Wisconsin - River Falls, University of Wisconsin - Stevens Point, University of Wisconsin - Stout, University of Wisconsin - Superior, University of Wisconsin - System Administration, and University of Wisconsin - Whitewater.

At the bottom of the page, there is a "Select" button and a copyright notice: "© 2009 Board of Regents - University of Wisconsin System. All Rights Reserved. Accessibility Policy | Privacy Policy | Contact Webmaster".

Access in action

journal.com.ezproxy.library.wisc.edu/blah

The screenshot shows a web browser window with the following details:

- Header:** UNIVERSITY OF WISCONSIN-MADISON Libraries
- Top Right:** Campus Libraries | Mobile | My Accounts ▾, Search this site... Search, chat | email | phone | desk, ASK
- Left Sidebar:** Rem, Login, UW SYSTEM, University of Wisconsin-Madison, Campus NetID Login, Help, and several links.
- Center Content:** UW SYSTEM, UW System Home -- Authentication Redirector, Select your organization, University of Wisconsin-Madison, Campus NetID Login form (NetID and Password fields, LOG IN button), and a forgot password link.
- Right Sidebar:** Account section with links: Activate your NetID, Modify your account, Reset your password, Create a strong password, and Login help.
- Bottom:** A message: You are connecting to a website or application that requires authentication. Please enter your NetID and password to continue. To protect your personal information, remember to Exit all web browsers (Quit on a Mac.).
- Footer:** Select button, © 2009 Board of Regents - University of Wisconsin System, All Rights Reserved, Accessibility Policy | Privacy Policy | Contact Webmaster.

Access in action

Oh, crap.

The screenshot shows a layered web interface. At the top is the University of Wisconsin-Madison Libraries homepage with a red header containing the UW logo and the word "Libraries". Below it is a login overlay for the UW System, prompting the user to "Select your organization" with options for "University of Wisconsin-Madison" and "UW Health". The main content area displays an article from the "Journal of Computer Science and Technology" titled "New Generations: Sequencing Machines and Their Computational Challenges" by David C. Schwartz and Michael S. Waterman. The article summary includes a "Buy now" button for \$39.95 / €34.95 / £29.95. To the right, there's a thumbnail of the journal cover and a "Look Inside" button. At the bottom, there are "Share" and "Other actions" links, along with a "Get Access" button.

UNIVERSITY OF WISCONSIN-MADISON

Campus Libraries | Mobile | My Accounts ▾

Search this site... Search

chat | email | phone | desk | ASK

Rem

Login

- Lo
- re:
- If
- se
- frc

Help

- Ne
- In:
- Ne
- Pe
- He
- Mc

University of Wisconsin-Madison

UW SYSTEM

UW System Home -- Authentication Redirector

Select your organization

University of Wisconsin-Madison

Journal of Computer Science and Technology

January 2010, Volume 25, Issue 1, pp 3-9

New Generations: Sequencing Machines and Their Computational Challenges

David C. Schwartz, Michael S. Waterman

Purchase on Springer.com

\$39.95 / €34.95 / £29.95 *

* Final gross prices may vary according to local VAT.

Buy now

Get Access

Share

Look Inside

Journal of Computer Science & Technology

Abstract

Other actions

» Export citations

Access in action

Library catalog

The screenshot shows the UW-Madison Library Catalog homepage. At the top, there's a red header bar with the library logo and the text "Library Catalog". Below it is a light green navigation bar with the word "Catalog" in large letters, followed by "Explore books", "music", "movies", and "and more". To the right of these are icons for a search bar, a user profile, and other catalog features. The main content area has a white background. It starts with a section titled "Searching UW-Madison Items" containing a search bar with the query "Journal of Computer Science and Technology", a dropdown menu set to "Anywhere", and a red "Search" button. Below this is a link to "Advanced Search - Search All of UW System". To the right of this section is a sidebar titled "New Library Catalog" with a bulleted list: "Take the visual tour (coming soon)", "Give us feedback", and "Visit MadCat: the legacy catalog". At the bottom left is another section titled "Browse by Subject" with the sub-instruction "Explore the library collections via subject topics, places, languages, author or creator name."

Access in action

Library catalog

The screenshot shows a library catalog interface with a red header bar. The left sidebar has categories: Catalog (selected), Catalog Home, Advanced Search, Browse, Subjects, Publication Info, Formats, and OCLC. The main search bar contains "Search the Library Catalog" and "Anywhere". The search results page displays the title "Journal of computer science and technology". Below the title are sections for "Subjects" (Computer science -- Periodicals, Electronic data processing -- Periodicals), "Publication Info" (English language ed., Beijing, China : Science Press, c1986), "Physical details" (v. : ill. ; 28 cm.), "Dates of publication" (Vol. 1, no. 1 (Jan. 1986)-), "ISSNs" (1000-9000, 1860-4749), and "OCLC" (ocm20699620). There are also buttons for Select, Cite, Email, Print, Share, and navigation links for Back to Search Results, Result 2 of 5,345, and Help.

Access in action

Library catalog

The screenshot shows the University of Wisconsin-Madison Library Catalog interface. The top navigation bar has two tabs: "Catalog" and "Search". The "Catalog" tab is selected, showing a search bar with "Anywhere" and a "Search" button. Below the search bar, the university logo and the word "Libraries" are displayed. On the left, there is a sidebar with links for "Cat", "Sea", "Jou", "Adv", "Job", and "Bro". The main content area displays search results for "Journal of Computer Science and Technology [1000-9000]". It includes links for "Get Full Text Online" (ProQuest ABI/INFORM Complete), "Available from 1997", and "Most recent 1 year(s) not available". There is also a link to "Look for Print and Other Formats - Library Catalog". On the right, there are several call-to-action buttons: "More Options", "Request A Copy", "Report A Problem", "See Table of Contents and More", "Start A New Search", "Permalink", and a prominent yellow "Chat Now" button. At the bottom, there is a "Refine This Search" section with fields for "Journal Title" (set to "Journal of Comput"), "Year", "Volume", "Issue", "Start Page", and "ISSN" (set to "1000-9000"). A "Go" button is located at the bottom right of this section.

Access in action

Library catalog

The screenshot shows the University of Wisconsin-Madison Library Catalog interface. The top navigation bar is red with the text "Library Catalog" and "Anywhere" search fields. Below the bar, the university logo and the word "Libraries" are displayed. A sidebar on the left contains links for "Cat", "Sea", "Jou", "Advi", "Bro", and "Explore". The main content area is titled "Find It Publication Information" for "Journal of Computer Science and Technology". It shows publication details like title, coverage, ISSN, language, and subjects. A search bar at the bottom allows users to search within the publication. At the very bottom, there's a section for browsing specific issues from 2010 to 2013.

Library Catalog

Anywhere

Search

UNIVERSITY OF WISCONSIN-MADISON

Libraries

Set up alert Create RSS feed

Source: Journal of Computer Science and Technology

Get Full Text

Publication title: Journal of Computer Science and Technology

Coverage (any format): Jan 1997 (Vol. 12, no. 1) - present
- Mc

Look for

ISSN: 1000-9000

Language: English

Subjects: Computers--Computer Architecture ; Computers

Show all ▾

Search within this publication:

Search Advanced Search

Browse specific issues

The list is currently ordered by most recent issue first. View oldest issue first

2010 - 2013

2013

2012

2011

2010

Nov 2010; Vol. 25 (6) | Sep 2010; Vol. 25 (5) | Jul 2010; Vol. 25 (4) | May 2010; Vol. 25 (3) | Mar 2010; Vol. 25 (2)

| Jan 2010; Vol. 25 (1)

Access in action

Finally.

Schwartz DC, Waterman MS. New generations: Sequencing machines and their computational challenges. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 25(1): 3-9 Jan. 2010

New Generations: Sequencing Machines and Their Computational Challenges

David C. Schwartz¹ and Michael S. Waterman^{2,3}

¹Laboratory for Molecular and Computational Genomics, Department of Chemistry and Laboratory of Genetics
University of Wisconsin-Madison, WI 53706, U.S.A.

²Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089-2916, U.S.A.

³Department of Automation, Tsinghua University, Beijing 100084, China

E-mail: dcschwartz@scitaff.wisc.edu, msw@usc.edu

Received September 5, 2009; revised November 24, 2009.

Abstract New generation sequencing systems are changing how molecular biology is practiced. The widely promoted \$1000 genome will be a reality with attendant changes for healthcare, including personalized medicine. More broadly the genomes of many new organisms with large samples from populations will be commonplace. What is less appreciated is the explosive demands on computation, both for CPU cycles and storage as well as the need for new computational methods. In this article we will survey some of these developments and demands.

Keywords genome sequencing, new generation sequencing, read mapping, optical mapping, sequence assembly, Eulerian graphs

1 Introduction

It may be somewhat futile to attempt to track perfectly an explosion. But here we hope to give some hints about the technological and computational challenges that will surely be addressed along the path to the commoditization of sequence information. As the cost of sequence information drops, its utility will grow as sequencing directly alters medical care, the type and safety of our food supply, and, of course, now unfortunately national security, which had predicted 50 years ago that lasers would find broad application as "pointers."¹ Accordingly, we expect that the experimental and computational challenges will become progressively intertwined in ways that foster development of completely new disciplines for tackling the even greater challenges that are now unthinkable. In this regard, we present here a brief overview of the current state of DNA sequencing, and our best guesses for how technology and computation may interact for creating this future.

2 Current Technology

Although commercial next generation platforms differ from each other in how sequence is actually obtained, they share the common advantage of not

requiring bacterial clone libraries. In many ways, the obviation of clone library construction and handling is a major reason why genome sequencing costs have plummeted, while platform throughput is dramatically increasing. Templates for large scale DNA sequencing are made from a library spread across massive culture plates and individual clones are isolated by "picking robots" for downstream sequencing reactions. Such operations, for large genomes such as human, require factory floor settings bristling with robots and technicians before any sequencing data is acquired. In contrast, next generation platforms construct "close" libraries directly from individual genomic DNA molecules, which are amplified by emulsion or bridge PCR (polymerase chain reaction). Existing genome libraries consist of small vesicles, or surfaces laden with amplicons, but there is one company^[1] whose libraries comprise unamplified genomic templates that are bound to surfaces.

2.1 Next-Generation Sequencing

Today an investigator can choose between four commercially available systems, each offering a panoply of technical strengths and weaknesses that need to be considered against overall cost and application: 1) Illumina's Genome Analyzer, 2) Life Technologies' SOLID

Access in action

There's also PubMed

The screenshot shows the PubMed homepage with a search bar containing the query "new generations: sequencing machines and their computational challenges". Below the search bar, a message states that PubMed is open but being maintained with minimal staffing due to government funding issues. The main content area displays a large image of a book and a brief description of what PubMed is. To the right, there's a "PubReader" section with an image of a tablet displaying a document. Below these sections are three columns of links: "Using PubMed", "PubMed Tools", and "More Resources".

NCBI Resources How To

PubMed new generations: sequencing machines and their computational challenges Advanced

PubMed is open, however it is being maintained with minimal staffing due to the lapse in government funding. Information will be updated to the extent possible, and the agency will respond to urgent operational inquiries. For updates regarding government operating status see USA.gov.

PubMed

PubMed comprises more than 23 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

PubReader

A whole new way to read scientific literature at PubMed Central

Using PubMed

- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools

- PubMed Mobile
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MeSH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Utilities
- LinkOut

Access in action

There's also PubMed

The screenshot shows the PubMed.gov homepage with a search bar at the top containing the query "new generations: sequencing machines and their computational challenges". Below the search bar, there is a message from the US National Library of Medicine stating that PubMed is open but being maintained with minimal staffing due to government funding issues. The main content area displays a single search result for an article titled "New Generations: Sequencing Machines and Their Computational Challenges". The article is by Somervell DC and Waterman MS, from the University of Wisconsin-Madison. The abstract discusses how new sequencing technologies are changing molecular biology and the resulting challenges for computation. The page includes links for "Full Text At" (PubMed Central), "Grant Support", "LinkOut - more resources", and "New and Notable". On the right side, there are sections for "Save items", "PubReader" (with a thumbnail image of a handheld device displaying the article), and "Related citations in PubMed".

new generations: sequencing machines and their computational challenges

Display Settings: Abstract

J Comput Biol. 2010 Jan; 12(1):3-8.

New Generations: Sequencing Machines and Their Computational Challenges.

Somervell DC, Waterman MS

Laboratory for Molecular and Computational Genomics, Department of Chemistry and Laboratory of Genetics, University of Wisconsin-Madison, WI 53706 USA.

Abstract

New generation sequencing systems are changing how molecular biology is practiced. The widely promoted \$1000 genome will be a reality with attendant changes for healthcare, including personalized medicine. More broadly the genomes of many new organisms with large sample sizes from populations will be commonplace. What is less appreciated is the explosive demands on computation, both for CPU cycles and storage as well as the need for new computational methods. In this article we will survey some of these developments and demands.

PubMed ID: 22121326 [PubMed] PMC ID: PMC3222932 Free PMC Article

Grant Support

LinkOut - more resources

New and Notable

Save items

Add to Favorites

PubReader

A whole new way to read scientific literature at PubMed Central

Click here to read article using PubReader

Related citations in PubMed

Carbon sequestration in *Byneococcus* Sp.: from molecular machines to Hera [OMICS. 2002]

Computational challenges of personal genomics [Cur Genomics. 2008]

Review Green systems biology - From single genomes, proteomes and n [J Proteomics. 2011]

Review Comprehensive next-generation cancer genome sequencing in the v [Biomark Med. 2011]

Review Where are we in genomics? [J Physiol Pharmacol. 2009]

Access in action

There's also PubMed

PubMed.gov Resources How To

PubMed Resources How To

Display Settings: Abstract

J Comput Biol Technol. 2010 Jan 12(1):1-8.

New Generations: Sequencing Machines and Their Computational Challenges

Sommer DC, Waterman MS

Laboratory for Molecular and Computational Genomics, Department of Chemistry and Laboratory of Genetics, University of California, Los Angeles, CA, USA

Abstract

New generation sequencing systems are changing how molecular biology is practiced. The widely promoted \$1000 genome will be a reality with attendant changes for healthcare, including personalized medicine. More broadly the genomes of many new organisms with large samples from populations will be commonplace. What is less appreciated is the explosive demands on computation, both for CPU cycles and storage as well as the need for new computational methods. In this article we will survey some of these developments and demands.

Using PubMed

PubMed Quick Search Full Text At PubMed PMID: 22121326 [PubMed] PMID: PMC3222502 Free PMC Article

Grant Support

LinkOut - more resources

New and Noteworthy

Related citations in PubMed

Carbon sequestration in *Byneococcus* Sp.: from molecular machines to Hera [OMICS. 2002]

Computational challenges of personal genomics [Cur Genomics. 2008]

Review Green systems biology - From single genomes, proteomes and n [J Proteomics. 2011]

Review Comprehensive next-generation cancer genome sequencing in the v [Biomark Med. 2011]

Review Where are we in genomics? [J Physiol Pharmacol. 2009]

free full-text archive

PubMed Central

PubReader

A whole new way to read scientific literature at PubMed Central

Click here to read article using PubReader

Access in action

There's also PubMed

The screenshot shows the PubMed.gov homepage with a search result for the article "New Generations: Sequencing Machines". The page includes the NIH logo, a search bar, and links for "Using PubMed" and "NIH/PA Manuscripts". The main content area displays the article title, authors (David C. Schwartz and Michael S. Waterman), and abstract.

PubMed.gov
US National Library of Medicine
NIH

PubMed respond

Display Settings: Abstract

J Comput Sci Technol. 2010 Jan; 25(1):3-8.

New Generations: Sequencing Machines

Schwartz DC, Waterman MS

Laboratory for Molecular and Computational Genomics, De

Abstract

New generation sequencing systems are changing attendant changes for healthcare, including personal populations will be commonplace. What is less appreciated is the need for new computational methods. In this ap

PubMed Qs

Full Text At

PubMed FE

PubMed TU

New and Ni

PMID: 22121326 [Published] PMCID: PMC3222932 Free

Grant Support

LinkOut - more resources

Journal List > NIH/PA Author Manuscripts > PMC3222932

 NIH Public Access
Author Manuscript
Accepted for publication in a peer reviewed journal
About Author manuscripts Submit a manuscript

PMCID: PMC3222932
NIHMSID: NIHMS335688

J Comput Sci Technol. Author manuscript; available in PMC 2011 November 23.
Published in final edited form as:
J Comput Sci Technol. 2010 January 1; 25(1): 3–8.
doi: 10.1007/s11390-010-9300-x

New Generations: Sequencing Machines and Their Computational Challenges

David C. Schwartz¹ and Michael S. Waterman²

Author information ► Copyright and License Information ►

See other articles in PMC that cite the published article.

Abstract Go to: ▾

New generation sequencing systems are changing how molecular biology is practiced. The widely promoted \$1000 genome will be a reality with attendant changes for healthcare, including personalized medicine. More broadly the genomes of many new organisms with large samplings from populations will be commonplace. What is less appreciated is the explosive demands on computation, both for CPU cycles and storage as well as the need for new computational methods. In this article we will survey some of these developments and demands.

Keywords: genome sequencing, new generation sequencing, optical mapping, sequence assembly, Eulerian graphs

1 Introduction Go to: ▾

It may be somewhat futile to attempt to track perfectly an explosion. But here we hope to give some hints about the technological and computational challenges that will be surely be addressed along the path to the commercialization of sequence information. As the cost of sequence information drops, its utility will grow as sequencing directly alters medical care, the type and safety of our food supply, and of course, now unfathomable applications: who would have predicted 50 years ago that lasers would find broad application as "pointers?" Accordingly, we expect that the experimental and computational challenges will become progressively intermingled in ways that may foster development of completely new disciplines for tackling the even greater challenges that are now unthinkable. In this regard, we present here a brief overview of the current state of DNA sequencing, and our best guesses for how technology and computation may interact for creating this future.

It's all about money

(Costs in scientific publishing)

- ▶ Research
- ▶ Writing
- ▶ Peer review, editorial oversight
- ▶ Journal administration
- ▶ Copy editing, typesetting
- ▶ Distribution

It's all about money

(Costs in scientific publishing)

- ▶ Research
- ▶ Writing
- ▶ Peer review, editorial oversight
- ▶ Journal administration
- ▶ Copy editing, typesetting
- ▶ Distribution
- ▶ Profit

It's all about money

(Costs in scientific publishing)

- ▶ Research
- ▶ Writing
- ▶ Peer review, editorial oversight
- ▶ Journal administration
- ▶ Copy editing, typesetting
- ▶ Distribution
- ▶ Profit

It's all about money

(Costs in scientific publishing)

- ▶ Research
- ▶ Writing
- ▶ Peer review, editorial oversight
- ▶ Journal administration
- ▶ Copy editing, typesetting
- ▶ Distribution
- ▶ Profit

It's not about

- ▶ Peer review
- ▶ Predatory publishing
- ▶ Impact factors
- ▶ Evaluating researchers
(for grants & promotions)

It's not about

- ▶ Peer review
- ▶ Predatory publishing
- ▶ Impact factors
- ▶ Evaluating researchers
(for grants & promotions)

Well, it sort of is...

It's not about

- ▶ Peer review
- ▶ Predatory publishing
- ▶ Impact factors
- ▶ Evaluating researchers
(for grants & promotions)

Well, it sort of is...

Paying for it

- ▶ Traditional approach
 - subscriptions
 - page charges
- ▶ Open access
 - bigger page charges
 - charge submissions?
- ▶ Endowments
- ▶ Direct grants to journals

\$7000 page charges

Broman KW (2012) Genotype probabilities at intermediate generations in the construction of recombinant inbred lines. <i>Genetics</i> 190:403–412	\$2,548
Broman KW (2012) Haplotype probabilities in advanced inter-cross populations. <i>G3</i> 2:199–202	\$1,650
Broman KW, Kim S, Sen Š, Ané C, Payseur BA (2012) Mapping quantitative trait loci onto a phylogenetic tree. <i>Genetics</i> 192:267–279	\$2,891

Invoice

GENETICS

Review Invoice

Article Information

Publisher:	Genetics Society Of America
Title:	Genetics
Issue:	Volume 192, Number 1
Manuscript Title:	Mapping Quantitative Trait Loci onto a Phylogenetic Tree
Manuscript Number:	142448
Article Type:	Regular Research Papers
Corr. Author Name (e-mail addr.):	Karl W Broman (kbroman@biostat.wisc.edu)
Membership Status:	Member

Charge Information

Review Estimated Publication Charges

Items	Unit Price	Quantity	Amount
Page Charges	\$70.00	13	\$910.00
Figure Charges	\$40.00	6	\$240.00
Supplemental Files (six pages or greater)	\$500.00	1	\$500.00
Open Access Option	\$1,200.00	1	\$1,200.00
Author Alterations	\$2.55	16	\$40.80
Subtotal:			\$2,890.80

Total Charges

Total:	\$2,890.80
Balance Due:	\$2,890.80

Invoice

GENETICS

Review Invoice

Article Information

Publisher:	Genetics Society Of America
Title:	Genetics
Issue:	Volume 192, Number 1
Manuscript Title:	Mapping Quantitative Trait Loci onto a Phylogenetic Tree
Manuscript Number:	142448
Article Type:	Regular Research Papers
Corr. Author Name (e-mail addr.):	Karl W Broman (kbroman@biostat.wisc.edu)
Membership Status:	Member

Review Estimated Publication Charges

Items	Unit Price	Quantity	Amount
Page Charges	\$70.00	13	\$910.00
Figure Charges	\$40.00	6	\$240.00
Supplemental Files (six pages or greater)	\$500.00	1	\$500.00
Open Access Option	\$1,200.00	1	\$1,200.00
Author Alterations	\$2.55	16	\$40.80
Subtotal:			\$2,890.80

Choices for young investigators

- ▶ Pay for open access
- ▶ Support young open access journals

or

- ▶ Let subscribers pay & do more experiments
- ▶ Continue to go after Science, Nature, & Cell

What can we do?

- ▶ Send our best work to open access journals
- ▶ Support junior faculty to keep their papers open
- ▶ Pay attention to the quality of the work
(not the impact factor of the journal)
- ▶ Raise endowments for trusted journals
- ▶ Reform copyright law

What can we do?

- ▶ Send our best work to open access journals
- ▶ Support junior faculty to keep their papers open
- ▶ Pay attention to the quality of the work
(not the impact factor of the journal)
- ▶ Raise endowments for trusted journals
- ▶ Reform copyright law