

Open access publishing

A researcher's perspective

Karl Broman

Biostatistics & Medical Informatics
University of Wisconsin–Madison

`biostat.wisc.edu/~kbroman`
`github.com/kbroman`

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., Potamouis 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 interface. At the top, the Google logo is on the left, and a search bar contains the text "New generations: sequencing machines and their computational challenges". To the right of the search bar is a blue search button with a magnifying glass icon. Below the search bar, the word "Scholar" is displayed in red. To its right, it says "About 20,200 results (0.12 sec)". Further right are three buttons: "Any time" with a dropdown arrow, a notification bell icon with the number "0", and another dropdown arrow.



The first search result is titled "New generations: Sequencing machines and their computational challenges" in purple text, with a "[HTML]" link to the right. The author information is "DC Schwartz, MS Waterman - Journal of computer science and technology, 2010 - Springer". The abstract text reads: "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**". Below the abstract are links for "Cited by 13", "Related articles", "All 14 versions", and "Cite".

The second search result is titled "Evolutionary computation: toward a new philosophy of machine intelligence" in blue text, with a "[BOOK]" icon to the left and a "[HTML]" link to the right. The author information is "DB Fogel - 2006 - books.google.com". The abstract text reads: "... 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**. ...". Below the abstract are links for "Cited by 2874", "Related articles", "All 14 versions", "Cite", and "More" with a dropdown arrow.

The third search result is titled "How to map billions of short reads onto genomes" in blue text, with a "[HTML]" icon to the left and a "[HTML]" link to the right. The author information is "C Trapnell, SL Salzberg - Nature biotechnology, 2009 - ncbi.nlm.nih.gov". The abstract text reads: "... 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 ...". Below the abstract are links for "Cited by 142", "Related articles", "All 26 versions", and "Cite".

Access in action

Google Scholar



Scholar

New gene
DC Schwartz,
Schwartz DC,
challenges. J
... **New Gene**
Cited by 13

[BOOK] Evol
DB Fogel - 20
... If the proce
fundamentally
beyond our ov
Cited by 2874

[HTML] How
C Trapnell, Sl
... As a practic
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**



Share

Other actions

- » [Export citations](#)
- » [Register for Journal Updates](#)
- » [About This Journal](#)
- » [Reprints and Permissions](#)

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.

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**

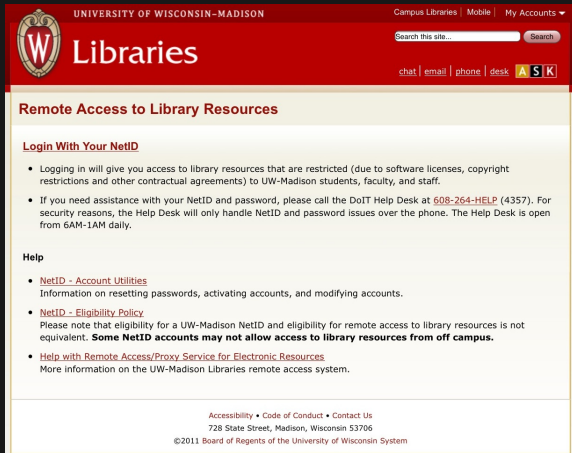
Broman KW, Weber JL (1999) Method for constructing confidently ordered linkage maps. *Genet Epidemiol* 16:337–343 **\$35.00**

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 top navigation bar of the University of Wisconsin-Madison Libraries website. The header is red with the university logo on the left, the text 'UNIVERSITY OF WISCONSIN-MADISON' and 'Libraries' in the center, and navigation links for 'Campus Libraries', 'Mobile', and 'My Accounts' on the right. Below the header is a search bar and a row of links for 'chat', 'email', 'phone', 'desk', and 'ASK'. The main content area has a yellow background and is titled 'Remote Access to Library Resources'. It includes a section for 'Login With Your NetID' with a bulleted list of instructions and a 'Help' section with three links: 'NetID - Account Utilities', 'NetID - Eligibility Policy', and 'Help with Remote Access/Proxy Service for Electronic Resources'. The footer contains accessibility and contact information, and a copyright notice for 2011.

UNIVERSITY OF WISCONSIN-MADISON

Campus Libraries | Mobile | My Accounts ▾

Search this site... Search

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 website. At the top, there is a red header with the University of Wisconsin-Madison logo and the text "UNIVERSITY OF WISCONSIN-MADISON" and "Libraries". To the right of the header, there are links for "Campus Libraries", "Mobile", and "My Accounts", along with a search bar. Below the header, there is a navigation bar with "chat", "email", "phone", "desk", and "ASK" buttons. The main content area is divided into several sections. On the left, there is a sidebar with "Rem" and "Logi" links, and a list of links including "Lo re:", "If se frc", "Help", "Me In:", "Me Plk eq", and "Me Mc". The main content area features a "UW SYSTEM" logo and a "UW System Home -- Authentication Redirector" heading. Below this, there is a "Select your organization" section with a list of radio buttons for various University of Wisconsin locations. The "University of Wisconsin - Madison" option is selected. At the bottom, 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".

UNIVERSITY OF WISCONSIN-MADISON

Campus Libraries | Mobile | My Accounts

Search this site... Search

chat | email | phone | desk | ASK

Rem

Logi

- Lo re:
- If se frc

Help

- Me In:
- Me Plk eq
- Me Mc

UNIVERSITY OF WISCONSIN SYSTEM

UW SYSTEM

UW System Home -- Authentication Redirector

Select your organization

This service requires you to authenticate with your local organization, please select it from the list below.

- 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 - Madison
- 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
- University of Wisconsin - Whitewater

Select

© 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

UNIVERSITY OF WISCONSIN-MADISON

Campus Libraries | Mobile | My Accounts ▾

Search this site... Search

chat | email | phone | desk **A S K**

Rem

UNIVERSITY OF WISCONSIN SYSTEM **UW SYSTEM**

UW System Home -- Authentication Redirector

Select your organization

University of Wisconsin-Madison

Campus NetID Login

NetID

Example: bbadger

Password

[Forgot your password?](#)

LOG IN

Account

- [Activate your NetID](#)
- [Modify your account](#)
- [Reset your password](#)
- [Create a strong password](#)
- [Login help](#)

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.)

Select

© 2009 Board of Regents - University of Wisconsin System. All Rights Reserved
[Accessibility Policy](#) | [Privacy Policy](#) | [Contact Webmaster](#)

Access in action

Oh, crap.

UNIVERSITY OF WISCONSIN-MADISON

Campus Libraries | Mobile | My Accounts

Libraries

Search this site... Search

chat | email | phone | desk A S K

Rem

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 *

Buy now

* Final gross prices may vary according to local VAT.

Get Access

Journal of Computer Science and Technology

Look Inside

Share

f t in

Other actions

Export citations

Abstract

Access in action

Library catalog

The screenshot shows the UW-Madison Library Catalog homepage. At the top, there is a red header with the UW-Madison logo and the text "Library Catalog". Below this is a navigation bar with the word "Catalog" and links for "Explore books", "music", "movies", and "and more". There are also icons for a bar chart, a globe, a magnifying glass, a person, and a gear. Below the navigation bar is a secondary navigation bar with links for "Catalog Home", "Advanced Search", "Browse", "Sign In", and "Help".

The main content area is titled "Searching UW-Madison Items". It features a search box containing the text "Journal of Computer Science and Technology". To the right of the search box is a dropdown menu set to "Anywhere" and a red "Search" button. Below the search box is a link for "Advanced Search - Search All of UW System".


To the right of the search area is a section titled "New Library Catalog" with a list of items:

- Take the visual tour (coming soon)
- Give us feedback
- Visit MadCat: the legacy catalog

Below the search area is a section titled "Browse by Subject" with the text "Explore the library collections via subject topics, places, languages, author or creator name."

Access in action

Library catalog



The screenshot displays a library catalog interface. At the top, there is a red header with the text "Library Catalog" and a search bar. Below the header, the main content area is white. The search bar contains the text "Search the Library Catalog" and a search button. The search results show the title "Journal of computer science and technology". Below the title, there are sections for "Subjects", "Publication Info", and "Formats". The "Subjects" section lists "Computer science -- Periodicals." and "Electronic data processing -- Periodicals.". The "Publication Info" section includes "Publication info: English language ed. ; Beijing, China : Science Press, c1986-" and "Physical details: v. : ill. ; 28 cm.". The "Formats" section lists "Electronic Resources" and "Journals, Magazines, Newspapers".

Library Catalog

Library Catalog

Cat

Catalog Search the Library Catalog Anywhere Search

Home Catalog Home Advanced Search Browse Sign In Help

Select Cite Email Print Share Back to Search Results < Result 2 of 5,345 >

Journal of computer science and technology

Subjects

- Computer science -- Periodicals.
- Electronic data processing -- Periodicals.

Publication Info

Formats:

- Electronic Resources
- Journals, Magazines, Newspapers

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

OCLC: ocm20699620

Access in action

Library catalog

The screenshot displays the University of Wisconsin-Madison Library Catalog interface. At the top, a red header contains the University of Wisconsin-Madison logo and the text 'UNIVERSITY OF WISCONSIN-MADISON Libraries'. Below the header, a search bar is visible with the text 'Search the Library Catalog' and a search button. The main content area shows search results for 'Journal of Computer Science and Technology [1000-9000]'. The results include a source link, a 'Get Full Text Online' link, and a 'Look for Print and Other Formats' link. On the right side, there are sections for 'More Options' and 'Refine This Search'. The 'More Options' section includes links for 'Request A Copy', 'Report A Problem', 'See Table of Contents and More', 'Start A New Search', and 'Permalink'. The 'Refine This Search' section includes a 'Note' and input fields for 'Journal Title', 'Year', 'Volume', 'Issue', 'Start Page', and 'ISSN', along with a 'Go' button.

Library Catalog

Catalog

Search the Library Catalog

Anywhere

Search

UNIVERSITY OF WISCONSIN-MADISON

Libraries

Search the library

chat | email | phone | desk

Find It: Results

Source: Journal of Computer Science and Technology [1000-9000]

[Get Full Text Online](#) - ProQuest ABI/INFORM Complete
Available from 1997
- Most recent 1 year(s) not available

[Look for Print and Other Formats](#) - Library Catalog

More Options

[Request A Copy](#) -

[Report A Problem](#) -

[See Table of Contents and More](#) - Ulrich's Periodical Directory

[Start A New Search](#)

Permalink

[Chat Now](#)

Refine This Search

Note: You must provide a Journal Title.

Journal Title:

Year:

Volume:

Issue:

Start Page:

ISSN:

[Go](#)

Access in action

Library catalog

The screenshot displays the University of Wisconsin-Madison Library Catalog interface. At the top, there is a red navigation bar with the University of Wisconsin-Madison logo and the text "Library Catalog". Below this, a search bar is visible with the text "Search the Library Catalog" and a dropdown menu set to "Anywhere". A red search button with a magnifying glass icon is also present.

The main content area features a red header with the University of Wisconsin-Madison logo and the text "Libraries". To the right of the logo, there are links for "Campus Libraries", "Mobile", and "My Accounts". Below the header, there is a search bar with the text "Search the library catalog" and a red search button.

The main content area is titled "Find It Publication Information" and displays details for the "Journal of Computer Science and Technology". The publication title is "Journal of Computer Science and Technology" with a link to "View most recent issue >". The publication title is "Journal of Computer Science and Technology". The coverage is "Jan 1997 (Vol. 12, no. 1) - present (delayed 1 year)". The ISSN is "1000-9000". The language is "English". The subjects are "Computers--Computer Architecture ; Computers".

Below the publication information, there is a section titled "Search within this publication:" with a search bar and a "Search" button. A link for "Advanced Search" is also visible.

At the bottom, there is a section titled "Browse specific issues" with a list of years from 2010 to 2013. Below the list, there is a link for "View oldest issue first".

At the very bottom, there is a list of issues: "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) | Dec 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 90089-2910, U.S.A.

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

E-mail: dschwartz@facstaff.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 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, 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, new unfathomable applications, who would have predicted 20 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.

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 "clone" libraries directly from individual genomic DNA molecules, which are amplified by emulsion or bridge PCR (polymerase chain reaction). Entire genome libraries consist of small vesicles, or surfaces laden with amplicons, but there is one company^[1] whose libraries comprise manipulated 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 displays the PubMed website interface. At the top, there is a navigation bar with "NCBI Resources" and "How To" links. Below this is the PubMed logo and the text "US National Library of Medicine National Institutes of Health". A search bar contains the text "new generations: sequencing machines and their computational challenges" and a "Search" button. Below the search bar, there is a blue banner with a warning icon and text: "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 app respond to urgent operational inquiries. For updates regarding government operating status see [USA.gov](#)." Below the banner, there are three main sections: "PubMed" with a stack of books image and text describing the database's size and content; "PubReader" with a tablet image and text describing a new way to read scientific literature; and three columns of navigation links: "Using PubMed" (Quick Start Guide, Full Text Articles, FAQs, Tutorials, New and Noteworthy), "PubMed Tools" (Mobile, Citation Matchers, Clinical Queries, Topic-Specific Queries), and "More Resources" (MeSH Database, Journals in NCBI Databases, Clinical Trials, E-Utilities, LinkOut).

NCBI Resources How To

PubMed
US National Library of Medicine
National Institutes of Health

PubMed new generations: sequencing machines and their computational challenges Search

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 app 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 displays the PubMed website interface. At the top, there's a navigation bar with 'NCBI Resources' and 'How To' links. The main search bar contains the query 'new generations: sequencing machines and their computational challenges'. Below the search bar, a notification states: '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 attempt to respond to urgent operational inquiries. For updates regarding government operating status see USA.gov.'

The search results section shows a single entry with the following details:

- Display Settings:** Abstract
- Journal:** J Genet Sci Technol. 2010 Jan 1;28(1):3-9.
- Title:** New Generations: Sequencing Machines and Their Computational Challenges.
- Author:** SCHWARTZ DC, WAINSTEIN MS
- Department:** Laboratory for Molecular and Computational Genomics, Department of Chemistry and Laboratory of Genetics, University of Wisconsin-Madison, WI 53706 USA.

The **Abstract** section reads: '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.'

On the right side of the page, there are several utility sections:

- Save Items:** Includes an 'Add to Favorites' button.
- PubReader:** A section titled 'A whole new way to read scientific literature at PubMed Central' with a 'Click here to read article using PubReader' link.
- Related citations in PubMed:** Lists several related articles with brief descriptions and dates, such as 'Carbon sequestration in *Synechococcus* Sp. from molecular machines to here [OMICS]. 2008' and 'Comprehensive next-generation cancer genome sequencing in the e [Biomark Med]. 2011'.

On the left side, there is a vertical sidebar with the following links:

- Using Pt
- PubMed Cl
- Full Text A
- PubMed F#
- PubMed Tu
- New and N

Access in action

There's also PubMed

The screenshot displays the PubMed interface with the following elements:

- Search Bar:** Contains the query "new generations: sequencing machines and their computational challenges".
- Article Title:** "New Generations: Sequencing Machines and Their Computational Challenges".
- Author:** Schwartz, D.C.; Weinert, M.S.
- 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."
- PMID:** 22121328 [PubMed] PMID: PMC322952 Free PMC Article
- Navigation Links:** Display Settings, Abstract, Grant Support, LinkOut - more resources, New and Notable.
- PubReader:** A section titled "PubReader" with the text "A whole new way to read scientific literature at PubMed Central" and a link "Click here to read article using PubReader".
- Related citations in PubMed:** A list of related articles including "Carbon sequestration in *Synechococcus* Sp. from molecular machines to hiems [OMICS, 2002]", "Computational challenges of personal genomics. [Curr Genomics, 2008]", "Review Green systems biology - From single genomes, proteomes and r [J Proteomics, 2011]", "Review Comprehensive next-generation cancer genome sequencing in the x [Biomark Med, 2011]", and "Review Where are we in genomics? [J Physiol Pharmacol, 2009]".
- Free in PMC:** A large yellow banner with the text "Free in PMC full-text archive" in red and black.

Access in action

There's also PubMed

The screenshot shows the PubMed search results page. At the top, there are navigation links for 'NCBI Resources' and 'How To'. The main search bar contains 'PubMed' and 'PubMed'. Below the search bar, there are several links: 'PubMed respond', 'PubMed is open, however it is being maintained respond to urgent operational inquiries. For up', 'Display Settings: Abstract', 'J Comput Sci Technol. 2010 Jan 1;25(1):3-9.', 'New Generations: Sequencing Machines and Their Computational Challenges', 'Schwartz DC, Waterman MS', 'Laboratory for Molecular and Computational Genomics, Dep', 'Abstract', 'New generation sequencing systems are changing attendant changes for healthcare, including senior populations will be commonplace. What is less appreciated is the need for new computational methods. In this a', 'Using Pt', 'PubMed Cl', 'Full Text At', 'PubMed E', 'PubMed Tu', 'New and N'. On the right side of the page, there are vertical labels 'NIH-PA Manuscripts' and 'NIH-PA Manuscripts'.

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

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): 3-9.

doi: 10.1007/s11390-010-9300-x

PMCID: PMC3222932

NIHMSID: NIHMS335688

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 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 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.

Another example

NCBI Resources How To Sign in to NCBI

PubMed Clustering Tooth Surfaces into Biologically Informative Caries Outcomes Search

US National Library of Medicine
National Institutes of Health

RSS Save search Advanced Help

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 attempt to respond to urgent operational inquiries. For updates regarding government operating status see [USA.gov](#).

Display Settings: Abstract Send to: View Full-Text Article at SAGE Publications

J Dent Res. 2013 Jan;92(1):32-7. doi: 10.1177/0022034512463241. Epub 2012 Oct 11.

Clustering tooth surfaces into biologically informative caries outcomes.

Shaffer JR, Feingold E, Wang X, Weeks DE, Wevart RJ, Crout B, McNeil DW, Marazita ML.
Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA. jr8@pitt.edu

Abstract

Dental caries affects most adults worldwide; however, the risk factors for dental caries do not necessarily exert their effects uniformly across all tooth surfaces. Instead, the actions of some risk factors may be limited to a subset of teeth/surfaces. Therefore, we used hierarchical clustering on tooth surface-level caries data for 1,068 Appalachian adults (ages 18-75 yrs) to group surfaces based on co-occurrence of caries. Our cluster analysis yielded evidence of 5 distinct groups of tooth surfaces that differ with respect to caries: (C1) pit and fissure molar surfaces, (C2) mandibular anterior surfaces, (C3) posterior non-pit and fissure surfaces, (C4) maxillary anterior surfaces, and (C5) mid-dentition surfaces. These clusters were replicated in a national dataset (NHANES 1999-2000, N = 3,123). We created new caries outcomes defined as the number of carious tooth surfaces within each cluster. We show that some cluster-based caries outcomes are heritable (i.e., under genetic regulation; $p < 0.05$), whereas others are not. Likewise, we demonstrate the association between some cluster-based caries outcomes and potential risk factors such as age, sex, educational attainment, and toothbrushing habits. Together, these results suggest that the permanent dentition can be subdivided into groups of tooth surfaces that are useful for understanding the factors influencing cariogenesis. Abbreviations: COHRA, Center for Oral Health in Appalachia, the principal study sample; C1-5, clusters 1-5, groups of similarly behaving tooth surfaces identified through hierarchical clustering; DMFS index, decayed, missing, or filled surfaces, a traditional caries measure representing the number of affected surfaces across the entire dentition; DMFS1-5, partial DMFS indices representing the number of affected surfaces within a hierarchical cluster; and NHANES, National Health and Nutrition Examination Survey, the secondary study sample.

PMID: 23064960 [PubMed - indexed for MEDLINE] PMCID: PMC3521447 [Available on 2014/1/1]

Save items

Add to Favorites

Related citations in PubMed

Genetic susceptibility to dental caries on pit and fissure and sr [Caries Res. 2012]
GWAS of dental caries patterns in the permanent dentition. [J Dent Res. 2013]
Incremental susceptibility of individual [Community Dent Oral Epidemiol. 1996]
Comparison of tooth surface-specific dental caries attack p [J Dent Res. 1993]
Review Factors that modify de- and remineraliz [Ann Acad Med Stetin. 1999]

See reviews...
See all...

Related information

Related Citations

Another example

NCBI Resources How To Sign in to NCBI

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed is open possible, and the

Display Settings

J Dent Res, 2013 Jan;f

Clustering tooth surfaces into biologically informative caries outcomes

Shaffer JR, Feingold E
Department of Human


Abstract
Dental caries affect uniformly across all Therefore, we used yrs) to group surface surfaces that differ v posterior non-pit an were replicated in a number of carious t (i.e., under genetic r some cluster-based toothbrushing habit surfaces that are us Health in Appalachia through hierarchica representing the nu number of affected t Survey, the second

PMID: 23064960 [Publ

Institution: University of Wisconsin-Madison
SAGE journals
Search all journals Advanced Search Search History Browse Journals

Impact Factor: 3.826 | Ranking: 2/82 in Dentistry, Oral Surgery & Medicine | Eigenfactor & Score: 0.02325 | 5-Year Impact Factor: 4.286

This item requires a subscription to Journal of Dental Research.

 **Full Text (PDF)**
Clustering Tooth Surfaces into Biologically Informative Caries Outcomes
J DENT RES January 2013 92: 32-37, first published on October 11, 2012

To view this item, select one of the options below:

- ▶ **IADR Member Sign In**
 - ▶ IADR Members, please sign in at IADR to access the journal online.
- ▶ **Sign In**

Already an individual subscriber?
If so, please sign in to SAGE Journals with your User Name and Password.

User Name
 - ▶ Can't get past this page?
 - ▶ Help with Cookies.
 - ▶ Need to Activate?

Password

Remember my user name & password.
[Forgot your user name or password?](#)
- ▶ **Purchase Short-Term Access**
 - ▶ **Pay per Article** - You may purchase this article for US\$32.00. You must download your purchase, which is yours to keep, within 24 hours.
 - ▶ **Regain Access** - You can regain access to a recent Pay per Article purchase if your access period has not yet expired.

Search this journal
Advanced Journal Search Return to
Search Results |
Edit My Last Search

This Article
Published online before print October 11, 2012, doi: 10.1177/0022034512463241
JDR January 2013 vol. 92 no. 1 32-37
Show PDF in full window
Abstract
Figures Only
Full Text
▶ **Full Text (PDF)**
Appendix

All Versions of this Article:
▶ Version of Record - Dec 10, 2012
▶ OnlineFirst Version of Record - Oct 11, 2012
What's this?

Services

- ▶ Email this article to a colleague
- ▶ Alert me when this article is cited
- ▶ Alert me if a correction is posted
- ▶ Article Metrics (beta)
- ▶ Similar articles in this journal
- ▶ Similar articles in PubMed
- ▶ Download to citation manager
- ▶ Request Permissions
- ▶ Request Reprints
- ▶ Load patient INFORMATION

led
al caries on
s Res. 2012)
ms in the
t Res. 2013)
individual
smiol. 1996]
-specific
t Res. 1993]
de- and
Stein. 1999)
See reviews...
See all...

Another example

NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed is open
possible, and the

Display Settings

J Dent Res, 2013 Jan;91

Clustering tooth

Shaffer JR, Feingold E
Department of Human

Abstract

Dental caries affects uniformly across all Therefore, we used yrs) to group surface surfaces that differ v posterior non-pit an were replicated in a number of carious t (i.e., under genetic s some cluster-based toothbrushing habit surfaces that are us Health in Appalachia through hierarchica representing the nu number of affected t Survey, the second

PMID: 23064960 (Publ

Institution: University of Wisconsin-Madison
SAGE journals
Search

Impact Factor: 3.826 | Ranking: 2/82 in Dentistry, C
Factor: 4.286

This item requires a subscription to Journal of



Full Text (PDF)

Clustering Tooth Surface:
Biologically Informative C
J DENT RES January 2013
published on October 11, 2

To view this item, select one of the options

IADR Member Sign In

IADR Members, please sign in IADR to access th

Sign In

Already an individual subscriber?

If so, please sign in to SAGE Journals with your Use

User Name Sign In
Password
 Remember my user name & password.
[Forgot your user name or password?](#)

Purchase Short-Term Access

- Pay per Article - You may purchase this article for l your purchase, which is yours to keep, within 24 ho
- Regain Access - You can regain access to a recent access period has not yet expired.

RESEARCH REPORTS

Clinical

J.R. Shaffer^{1,2*}, E. Feingold^{1,3},
X. Wang^{2,4,5}, D.E. Weeks^{1,3},
R.J. Weyant^{6,8}, R. Crou^{7,9},
D.W. McNeill⁸, and M.L. Maraziti^{1,2,4,5,9}

¹Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; ²Center for Oral Health Research in Appalachia, University of Pittsburgh, Pittsburgh, PA, USA; ³USA; and West Virginia University, Morgantown, WV 26506, USA; ⁴Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; ⁵Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; ⁶Department of Oral Medicine, University of Pittsburgh, Pittsburgh, PA, USA; ⁷Department of Dental Public Health and Information Management, University of Pittsburgh, School of Dental Medicine, Pittsburgh, PA, USA; ⁸Department of Periodontology, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; ⁹Oral Practice and Rural Health, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; and ¹⁰Clinical and Translational Science Institute, and Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; *Corresponding author; jrs@pitt.edu

J Dent Res 92(1):32-37, 2013

ABSTRACT

Dental caries affects most adults worldwide; however, the risk factors for dental caries do not necessarily meet their effects uniformly across all tooth surfaces. Instead, the extent of some risk factors may be a subset of tooth/human. Therefore, we used hierarchical clustering on tooth surface categories from 1,068 Appalachian adults (ages 18-75 yrs) to group surfaces based on co-occurrence of caries. Our cluster analysis yielded evidence of 5 distinct groups of tooth surfaces that differ with respect to caries: (C1) pit and fissure molar surfaces, (C2) mesial-labial anterior surfaces, (C3) premolar non-pit and fissure anterior, (C4) maxillary anterior surfaces, and (C5) mid-distal anterior. These clusters were replicated in a national dataset (NHANES 1999-2004, N = 3,123), the overall caries outcome defined as the number of carious tooth surfaces within each cluster. We show that some cluster-based caries outcomes are heritable (i.e., under genetic regulation; $p < 0.05$), whereas others are not. Likewise, we demonstrate the association between some cluster-based caries outcomes and potential risk factors such as age, sex, educational attainment, and breastfeeding habits. Together, these results suggest that the permanent dentition can be subdivided into groups of tooth surfaces that are useful for understanding the factors influencing caries etiology. **Abbreviations:** COHRA, Center for Oral Health in Appalachia; the principal study samples (C1-5, cluster 1-5, groups of similarly behaving tooth surfaces identified through hierarchical clustering; DMFS index, decayed, missing, or filled surface, a traditional caries measure representing the number of affected surfaces across the entire dentition; DMFTS-5, partial DMFS index representing the number of affected surfaces within a hierarchical cluster; and NHANES, National Health and Nutrition Examination Survey, the secondary study sample.

KEY WORDS: dental caries, permanent dentition, white spots, hierarchical clustering, cluster analysis, heritability.

DOI: 10.1177/0022034512463431

Received June 29, 2012; Last revision August 30, 2012; Accepted September 8, 2012

A supplemental appendix to this article is published electronically only at <http://jdr.sagepub.com/supplemental>.

© International & American Associations for Dental Research

Clustering Tooth Surfaces into Biologically Informative Caries Outcomes

INTRODUCTION

Dental caries, which affects the great majority of adolescents and adults throughout the world, is a multi-factorial disease caused by the effects of numerous environmental, behavioral, and genetic factors. Many risk factors have been identified, such as host genetics (Horowitz et al., 1958); environmental exposures, including fluoride, cariogenic bacteria, and pH-altering agents; behavioral factors, including diet and oral hygiene; characteristics of the dentition, including enamel composition and position and morphology of teeth; characteristics of the oral environment, including saliva composition, flow rate, and pH buffering capacity; and demographic factors, including age, sex, race, ethnicity, socio-economic status, and access to oral health care (Hunter, 1988). This complexity is further compounded by the disease phenotype, which may manifest as innumerable combinations of caries lesions across different tooth surfaces of the full dentition. Because caries risk factors may exert differential effects across tooth surfaces of the permanent dentition, measurable caries experience may be modeled as the cumulative result of multiple superimposed patterns of decay due to the various risk factors (Batchelor and Sheiham, 2004; Shaffer et al., 2012a).

In epidemiological studies, however, caries experience is typically reduced to a single measure of decay, such as the widely used DMFT/S index (calculated as the sum of decayed, missing due to decay, or filled/restored teeth/surfaces). Such global measures of caries experience ignore the fact that categories of tooth surfaces exhibit differences in susceptibility to decay and are differentially affected by risk factors. Because they ignore the patterns of dental caries across the dentition, global measures of decay such as DMFT/S index may be limited in their ability to identify caries risk factors that exert their effects on specific categories of tooth surfaces. Indeed, previous studies have demonstrated that modeling the patterns of tooth decay is beneficial for epidemiological (Pester et al., 2003, 2009) and genetic studies (Shaffer et al., 2012a, 2012b). We used hierarchical clustering analysis to group tooth surfaces into categories based on co-occurrence of caries lesions. We then generated novel caries outcomes reflecting these tooth-surface categories, and explored their utility for studying caries etiology.

Another example

NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed is open possible, and the

Display Settings

J Dent Res, 2013 Jan;1

Clustering tooth surfaces
Shaffer JR, Feingold E
Department of Human

Abstract

Dental caries affects uniformly across all Therefore, we used yrs) to group surface surfaces that differ v posterior non-pit an were replicated in a number of carious t (i.e., under genetic r some cluster-based toothbrushing habit surfaces that are us Health in Appalachia through hierarchica representing the nu number of affected r Survey, the second

PMID: 23064960 [PubM

Institution: University of Wisconsin-Madison
SAGE journals

Impact Factor: 3.826 | Ranking: 2/82 in Dentistry, C
Factor: 4.286

This item requires a subscription to Journal of



Full Text (PDF)
Clustering Tooth Surface:
Biologically Informative C
J DENT RES January 2013
published on October 11, 2

To view this item, select one of the options

IADR Member Sign In
IADR Members, please sign in at IADR to acc

Sign In
Already an individual subscriber?
If so, please sign in to SAGE Journals with you
User Name Sign In
Password
 Remember my user name & passw
Forgot your user name or password?

Purchase Short-Term Access
Pay per Article - You may purchase this articl
your purchase, which is yours to keep, with in ;
Regain Access - You can regain access to a n
access period has not yet expired.

RESEARCH REPORTS

Clinical

J.R. Shaffer^{1,2,*}, E. Feingold^{1,3},
X. Wang^{2,4,5}, D.E. Weeks^{1,3},
R.J. Weyant^{6,8}, R. Crou^{7,9},
D.W. McNeil⁸, and M.L. Maraziti^{1,2,4,5,9}

¹Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; ²Center for Oral Health Research in Appalachia, University of Pittsburgh, Pittsburgh, PA 15261, USA; ³West Virginia University, Morgantown, WV 26506, USA; ⁴Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; ⁵Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA; ⁶Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA; ⁷Department of Dental Public Health and Information Management, University of Pittsburgh, School of Dental Medicine, Pittsburgh, PA 15261, USA; ⁸Department of Periodontics, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; ⁹Oral Practice and Rural Health, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; and ¹⁰Clinical and Translational Science Institute, and Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. *Corresponding author: jrs7@pitt.edu

J Dent Res 92(1):32-37, 2013

ABSTRACT

Dental caries affects most adults worldwide, however, the risk factors for dental caries do not necessarily meet their effects uniformly across all tooth surfaces. Instead, the actions of some risk factors may be limited to a subset of tooth/surfaces. Therefore, we used hierarchical clustering on tooth surface-level caries data for 1,068 Appalachian adults (ages 18-75 yrs) to group surfaces based on co-occurrence of caries. Our cluster analysis yielded evidence of 5 distinct groups of tooth surfaces that differ with respect to caries: (C1) pit and fissure occlusal surfaces, (C2) mesial-labial anterior surfaces, (C3) posterior non-pit and fissure surfaces, (C4) maxillary anterior surfaces, and (C5) mid-distal anterior surfaces. These clusters were replicated in a national dataset (NHANES 1999-2004, n = 3,123). We created new caries outcomes defined as the number of carious tooth surfaces within each cluster. We show that some cluster-based caries outcomes are heritable (i.e., under genetic regulation; $p < 0.05$), whereas others are not. Likewise, we demonstrate the association between some cluster-based

Clustering Tooth Surfaces into Biologically Informative Caries Outcomes

INTRODUCTION

Dental caries, which affects the great majority of adolescents and adults throughout the world, is a multi-factorial disease caused by the effects of numerous environmental, behavioral, and genetic factors. Many risk factors have been identified, such as: host genetics (Horowitz et al., 1958); environmental exposures, including fluoride, cariogenic bacteria, and pH-altering agents; behavioral factors, including diet and oral hygiene; characteristics of the dentition, including enamel composition and positions and morphology of teeth; characteristics of the oral environment, including saliva composition, flow rate, and pH buffering capacity; and demographic factors, including age, sex, race, and ethnicity, socio-economic status, and access to oral health care (Hunter, 1988). This complexity is further compounded by the disease phe-

... caries lesions factors may tion, mea- of multiple tchelor and

... ily reduced dex (calculus/ dented teeth/ act that cat- way and are patterns of as DMFT/S s this exort ions studies merical for uffer et al., p tooth sur- then gen- eries, and

DOI: 10.1177/0022034512463241

Received June 29, 2012; Last revision August 30, 2012;
Accepted September 8, 2012

A supplemental appendix to this article is published electronically only at <http://jdr.sagepub.com/supplemental>.

© International & American Associations for Dental Research

Another example

NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed is open possible, and the

Display Settings

J Dent Res, 2013 Jan; 92(1):1-6

Clustering tooth surfaces
Shaffer JR, Feinold E, Department of Human

Abstract
Dental caries affects uniformly across all Therefore, we used yrs) to group surface surfaces that differ v posterior non-pit an were replicated in a number of carious t (i.e., under genetic some cluster-based toothbrushing habit surfaces that are us Health in Appalachia through hierarchica representing the nu number of affected Survey, the second

PMID: 23064960 | PubM

SAGE journals

Institution: University of Wisconsin-Madison

Impact Factor: 3.826 | Ranking: 2/82 in Dentistry, Oral Surgery & Medicine | Eigenfactor Score: 0.02325 | 5-Year Impact Factor: 4.286

This item requires a subscription to Journal of Dental Research.

 **Full Text (PDF)**
Clustering Tooth Surface: Biologically Informative C
J DENT RES January 2013
published on October 11, 2012

To view this item, select one of the options

- IADR Member Sign In**
IADR Members, please sign in at IADR to access the journal online.
- Sign In**
Already an individual subscriber?
If so, please sign in to SAGE Journals with your
User Name
Password
 Remember my user name & password
[Forgot your user name or password?](#)

- Purchase Short-Term Access**
- Pay per Article** - You may purchase this article purchase, which is yours to keep, within 30 days.
- Regain Access** - You can regain access to a limited access period has not yet expired.

RESEARCH REPORTS

Clinical

Institution: University of Wisconsin-Madison

SAGE journals

Search all journals Advanced Search Search History Browse

JDR JOURNAL OF DENTAL RESEARCH

Home OnlineFirst All Issues Subscribe RSS Email Alerts

Return to Search Results | Edit My Last Search

Search this journal

Impact Factor: 3.826 | Ranking: 2/82 in Dentistry, Oral Surgery & Medicine | Eigenfactor Score: 0.02325 | 5-Year Impact Factor: 4.286

This item requires a subscription to Journal of Dental Research.

Abstract
Dental caries affects uniformly across all tooth surfaces, but is limited to a subset of clustering on tooth in adults (ages 18-75 years). Our cluster analysis of tooth surfaces that differ in caries risk (i.e., under genetic control) in a population-based toothbrushing habit survey, the second National Health and Nutrition Examination Survey, the second

DOI: 10.1177/0000725612471111

Received
Accepted
A supplement
Internati

- IADR Member Sign In**
IADR Members, please sign in at IADR to access the journal online.
- Sign In**
Already an individual subscriber?
If so, please sign in to SAGE Journals with your User Name and Password.
User Name
Password
 Remember my user name & password.
[Forgot your user name or password?](#)
[Can't get past this page?](#)
[Help with Cookies.](#)
[Need to Activate?](#)
- OpenAthens Users**
Sign in via OpenAthens: If your organization uses OpenAthens, you can log in using your OpenAthens username and password. Contact your library for more details.
List of OpenAthens registered sites, including contact details.
- Login via Your Institution**
Login via your institution: You may be able to gain access using your login credentials for your institution. Contact your library if you do not have a username and password.
- Subscribe/Recommend**
Become a Member - Join the International and American Associations for Dental Research.
Click here to subscribe to the print and/or online journal.
Click here to recommend to your library.

Current Issue
October 2013, 92 (10)



Alert me to new issues of Journal of Dental Research

Submit a Manuscript

Free Sample Copy

Email Alerts

RSS feed

More about this journal

Other IADR Journals

Advances in Dental Research

Critical Reviews in Oral Biology & Medicine

International and American Associations for Dental Research

Interested in article-level

Twitter is useful

(for venting)



Twitter is useful (for venting and more)



Karl B
@kwbr

You SAGE b
from J Dent
supplement,
jdr.sagepub.
#OA

10/14/13, 11:23 A

RESEARCH REPORTS

Clinical

J.R. Schaffer^{1,2*}, E. Feingold^{1,3}, X. Wang^{2,4,5},
D.E. Weeks^{1,3}, R.J. Weyant^{2,6}, R. Croft^{2,7},
D.W. McNeill⁸, and M.L. Marazita^{1,2,4,5,9}

¹Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; ²Center for Oral Health Research in Appalachia, University of Pittsburgh, Pittsburgh, PA 15261, USA, and West Virginia University, Morgantown, WV 26506, USA; ³Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; ⁴Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15219, USA; ⁵Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA; ⁶Department of Dental Public Health and Information Management, University of Pittsburgh, School of Dental Medicine, Pittsburgh, PA 15261, USA; ⁷Department of Periodontics, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; ⁸Dental Practice and Rural Health, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; and ⁹Clinical and Translational Science Institute, and Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; *Corresponding author, jrs51@pitt.edu

J Dent Res DOI: 10.1177/0022034512463241

APPENDIX

COHRA Population Characteristics

Descriptive characteristics of the COHRA sample are shown in the Appendix Table. Compared with the general US population, this Appalachian sample from rural West Virginia and Pennsylvania is poorer, less educated, and underserved with respect to oral health care.

Cross-validation and Defining Separate Clusters

Determining the number of separate clusters is the principal challenge in the field of cluster analysis, and there is currently very little statistical theory in this area. Defining *a priori* minimum distances between clusters is arbitrary and may lead to over-fitting. Therefore, we instead performed two-fold cross-validation (e.g., Salvador and Chan, 2004), a commonly used approach for determining the number of clusters. Cross-validation allows us to determine the sensitivity of our clusters to perturbations of the input data and to identify what level of within-cluster similarity and between-cluster dissimilarity defines stable clusters. We randomly divided our sample into two halves and performed hierarchical clustering on each half. This process was repeated for 10 random halves. By comparing dendrograms, we determined the maximum number of separate clusters that were consistently observed across all random halves. Overall, tooth surfaces were distributed across 5 very stable clusters. Example cluster results for two complementary halves (which total to the full sample) are shown in Appendix

Clustering Tooth Surfaces into Biologically Informative Caries Outcomes

Fig. 1. Similarly, to assess any effects of the inclusion of biological relatives in our sample, we repeated hierarchical clustering in the maximal subset of unrelated individuals (see Appendix Fig. 2), which were newly identical to the total COHRA and NHANES 1999-2000 samples.

In both the total COHRA sample and in the NHANES 1999-2000 sample, the fifth cluster was subdivided into maxillary and mandibular components (indicating possibly 6 instead of 5 clusters). However, these subdivisions were not consistently observed *via* cross-validation of the COHRA sample. That is, in some randomly chosen halves of the COHRA sample, the sixth cluster did not distinguish maxillary and mandibular components of the fifth cluster. Therefore, we have conservatively shown results for the 5 stable clusters, as well as the maxillary and mandibular subdivisions of the fifth cluster.

Overwhelmingly, the cluster results were stable within the COHRA sample, and consistent between the COHRA sample and NHANES 1999-2000 sample. That being stated, there were subtle differences observed among cluster results. For example, left-right asymmetry was observed, albeit rarely, in the random halves (e.g., Appendix Fig. 1A, tooth #20 and tooth #29). Likewise, some tooth surfaces physically positioned on the border between two adjacent clusters shifted membership (e.g., Appendix Figs. 1A and 2, tooth #21 and tooth #28). In many cases, the result of such shifts in cluster membership echoed the subtle differences between COHRA and NHANES 1999-2000. We speculate that the all-or-nothing nature of our clustering approach represents an oversimplification of the relationships among tooth surfaces with respect to dental caries risk factors

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
 - submission charges?
- ▶ Endowments
- ▶ Direct grants to journals

\$7000 page charges

Broman KW (2012) Genotype probabilities at intermediate generations in the construction of recombinant inbred lines. *Genetics* 190:403–412

\$2,548

Broman KW (2012) Haplotype probabilities in advanced inter-cross populations. *G3* 2:199–202

\$1,650

Broman KW, Kim S, Sen S, Ané C, Payseur BA (2012) Mapping quantitative trait loci onto a phylogenetic tree. *Genetics* 192:267–279

\$2,891

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

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