Open access publishing
A researcher's perspective

Karl Broman
Biostatistics & Medical Informatics
University of Wisconsin–Madison

kbroman.org
github.com/kbroman
@kwbroman


Access in action

Google Scholar

New generations: sequencing machines and their computational challenges
DC Schwartz, MS Waterman - Journal of computer science and technology, 2010 - Springer
... New Generations: Sequencing Machines and Their Computational
Cited by 13  Related articles  All 14 versions  Cite

[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

[HTML] How to map billions of short reads onto genomes
... 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
New Generations: Sequencing Machines and Their Computational Challenges

David C. Schwartz, Michael S. Waterman

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?


Access in action

There's also PubMed
Access in action

There's also PubMed
Another example

*Clustering tooth surfaces into biologically informative caries outcomes.*

Shaffer JR, Feingold E, Wang X, Weeks DE, Weyant RJ, Crout R, McNeil DW, Marazita ML. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA. jhj@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.

Another example
Another example
Another example
Another example
Twitter is useful

(for venting)

You SAGE bailiffs. ILL’d an article from J Dent Res, but it didn’t include supplement, also behind pay wall. jdr.sagepub.com/content/92/1/3...
#OA

10/14/13, 11:23 AM
Twitter is useful
(for venting and more)

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 a 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 overfitting. 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.
Twitter is useful

#icanhazpdf
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...
Paying for it

- **Traditional approach**
  - subscriptions
  - page charges

- **Open access**
  - bigger page charges
  - submission charges?
Paying for it

- Traditional approach
  - subscriptions
  - page charges

- Open access
  - bigger page charges
  - submission charges?

- Endowments

- Direct grants to journals
# Genetics

**Review Invoice**

<table>
<thead>
<tr>
<th>Article Information</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Publisher:</td>
<td>Genetics Society Of America</td>
</tr>
<tr>
<td>Title:</td>
<td>Genetics</td>
</tr>
<tr>
<td>Issue:</td>
<td>Volume 192, Number 1</td>
</tr>
<tr>
<td>Manuscript Title:</td>
<td>Mapping Quantitative Trait Loci onto a Phylogenetic Tree</td>
</tr>
<tr>
<td>Manuscript Number:</td>
<td>142448</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Regular Research Papers</td>
</tr>
<tr>
<td>Corr. Author Name (e-mail addr.):</td>
<td>Karl W Broman (<a href="mailto:kbroman@biostat.wisc.edu">kbroman@biostat.wisc.edu</a>)</td>
</tr>
<tr>
<td>Membership Status:</td>
<td>Member</td>
</tr>
</tbody>
</table>

**Charge Information**

<table>
<thead>
<tr>
<th>Review Estimated Publication Charges</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>Unit Price</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Page Charges</td>
<td>$70.00</td>
</tr>
<tr>
<td>Figure Charges</td>
<td>$40.00</td>
</tr>
<tr>
<td>Supplemental Files (six pages or greater)</td>
<td>$500.00</td>
</tr>
<tr>
<td>Open Access Option</td>
<td>$1,200.00</td>
</tr>
<tr>
<td>Author Alterations</td>
<td>$2.55</td>
</tr>
<tr>
<td><strong>Subtotal:</strong></td>
<td><strong>$2,890.80</strong></td>
</tr>
</tbody>
</table>

**Total Charges**

| Total:       | $2,890.80 |
| Balance Due: | $2,890.80 |
**GENETICS**

**Review Invoice**

<table>
<thead>
<tr>
<th>Article Information</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Publisher:</td>
<td>Genetics Society Of America</td>
</tr>
<tr>
<td>Title:</td>
<td>Genetics</td>
</tr>
<tr>
<td>Issue:</td>
<td>Volume 192, Number 1</td>
</tr>
<tr>
<td>Manuscript Title:</td>
<td>Mapping Quantitative Trait Loci onto a Phylogenetic Tree</td>
</tr>
<tr>
<td>Manuscript Number:</td>
<td>142448</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Regular Research Papers</td>
</tr>
<tr>
<td>Corr. Author Name (e-mail addr.):</td>
<td>Karl W Broman (<a href="mailto:kbroman@biostat.wisc.edu">kbroman@biostat.wisc.edu</a>)</td>
</tr>
<tr>
<td>Membership Status:</td>
<td>Member</td>
</tr>
</tbody>
</table>

**Review Estimated Publication Charges**

<table>
<thead>
<tr>
<th>Items</th>
<th>Unit Price</th>
<th>Quantity</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page Charges</td>
<td>$70.00</td>
<td>13</td>
<td>$910.00</td>
</tr>
<tr>
<td>Figure Charges</td>
<td>$40.00</td>
<td>6</td>
<td>$240.00</td>
</tr>
<tr>
<td>Supplemental Files (six pages or greater)</td>
<td>$500.00</td>
<td>1</td>
<td>$500.00</td>
</tr>
<tr>
<td>Open Access Option</td>
<td>$1,200.00</td>
<td>1</td>
<td>$1,200.00</td>
</tr>
<tr>
<td>Author Alterations</td>
<td>$2.55</td>
<td>16</td>
<td>$40.80</td>
</tr>
</tbody>
</table>

**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