


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.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Title</th>
<th>Journal</th>
<th>Price</th>
</tr>
</thead>
</table>
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](https://www.library.wisc.edu/help/netid-account-utilities)
  - Information on resetting passwords, activating accounts, and modifying accounts.
- [NetID - Eligibility Policy](https://www.library.wisc.edu/help/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](https://www.library.wisc.edu/help/remote-access-proxy-service)
  - More information on the UW-Madison Libraries remote access system.
Access in action

journal.com.ezproxy.library.wisc.edu/blah
Access in action

journal.com.ezproxy.library.wisc.edu/blah
Access in action

Oh, crap.
Access in action

Library catalog
Access in action

Library catalog

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
Access in action

Library catalog
Finally.

New Generations: Sequencing Machines and Their Computational Challenges

David C. Schwartz1 and Michael S. Waterman2,3

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: dcschwartz@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, 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.

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 company11 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
Access in action

There's also PubMed
Access in action

There's also PubMed
Access in action

There's also PubMed

New Generations: Sequencing Machines and Their Computational Challenges

David C. Schwartz¹ and Michael S. Waterman²

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, 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 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


Clustering tooth surfaces into biologically informative caries outcomes.
Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA. jsh@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

Clustered Tooth Surfaces into Biologically Informative Caries Outcomes


This item requires a subscription to Journal of Dental Research.

Full Text (PDF)

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
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.

Services
- Email this article to a colleague
- Alert me when this article is cited
- Article Metrics (beta)
- Similar articles in this journal
- Similar articles in PubMed
- Download to citation manager
- Request Permissions
- Request Reprints
- Load patientINFORMATION
Another example
Another example

INTRODUCTION

Dental caries, which affects the great majority of adolescents and adults throughout the world, is a multifactorial disease caused by the effects of numerous environmental, behavioral, and genetic factors. Many risk factors have been identified, such as host genetics (Hovemitz et al., 1983); environmental exposures, including fluoride, cariogenic bacteria, and pH-altering agents; behavioral factors, including diet and oral hygiene; characteristics of the diet, 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, ethnicity, socio-economic status, and access to oral health care (Hunter, 1988). This complexity is further compounded by the presence of genetic factors; the disease process is likely initiated by an interaction of multiple risk factors that are both environmental and genetic in nature.
Another example
Twitter is useful
(for venting)

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


# GENETICS

## Review Invoice

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

## Charge Information

### 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

### Total Charges

<table>
<thead>
<tr>
<th>Total</th>
<th>$2,890.80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance Due</td>
<td>$2,890.80</td>
</tr>
</tbody>
</table>
# GENETICS

## Review Invoice

### Article Information

<table>
<thead>
<tr>
<th>Publisher:</th>
<th>Genetics Society Of America</th>
</tr>
</thead>
<tbody>
<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