My Desktop
Prepare & Submit Proposals
Proposal Status
Proposal Functions
Awards & Reporting
Notifications & Requests
Project Reports
Submit Images/Videos
Award Functions
Manage Financials
Program Income Reporting
Grantee Cash Management Section Contacts
Administration
Lookup NSF ID

Preview of Award 1062301 - Final Project Report

Cover |
Accomplishments |
Products |
Participants/Organizations |
Impacts |
Changes/Problems
| Special Requirements

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1062301
Project Title:	ABI Innovation: Barcoding-Free Multiplexing: Leveraging Combinatorial Pooling for High-Throughput Sequencing
PD/PI Name:	Stefano Lonardi, Principal Investigator Timothy J Close, Co-Principal Investigator
Recipient Organization:	University of California-Riverside
Project/Grant Period:	05/01/2011 - 04/30/2015
Reporting Period:	05/01/2014 - 04/30/2015
Submitting Official (if other than PD\PI):	N/A
Submission Date:	N/A
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	N/A

Accomplishments

*** What are the major goals of the project?**

The research plan for this NSF award was articulated around a novel sequencing protocol that combines next-generation sequencing instruments and 'smart' pooling. The proposed protocol hinges on the computational component, which deals with the preprocessing of short reads and post-processing of contigs.

The development of this computational component was the focus of the our effort in Year 1. In Year 2 we applied the new protocol on a BAC library for barley (*Hordeum vulgare*) and cowpea (*Vigna unguiculata*). In Year 3 (this year), we have improved the assemblies and prepared/submitted manuscripts.

The main steps of our combinatorial sequencing method are summarized next. More details can be found in the original proposal.

- A. Obtain a BAC library for the target organism
- B. Select gene-enriched BACs from the library (optional)
- C. Fingerprint BACs and build a physical map
- D. Select a minimum tiling path (*MTP*) from the physical map
- E. Pool the MTP BACs according to the shifted transversal design
- F. Sequence the DNA in each pool, trim/clean sequenced reads
- G. Clean and error-correct reads and generate unireads by greedy assembly
- H. Assign unireads to BAC clones (*deconvolution*)
- I. Assemble reads BAC-by-BAC using a short-read assembler
- J. Merge BAC assemblies (guided by the physical map)

*** What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?**

Major Activities: Accomplishments on steps A-E for the cowpea and barley genome were reported in Year 1.
In Year 2, we focused on developing/improving steps F, G, and H. In particular we sequenced and assembled ~4,000 cowpea BACs and ~15,000 barley BACs (as reported in Year 2 progress report).
In Year 3, we focused on developing/improving steps I and J, and publish manuscripts on the
findings of this project.

Specific Objectives: Specific accomplishments for Year 3:

- we designed and implemented a new method to improve read decoding and assembly quality based on the idea of data "slicing" (manuscript to appear in *Bioinformatics*)
- we carried out extensive evaluations of multiple de novo genome assemblers (Velvet, IDBA-UD, SPAdes) for BAC assembly and wrote a manuscript on the findings of the analysis of the 15,000 barley BACs (currently on biorXiv, to be submitted to *Genome Biology)*
- we designed and implemented a new meta-assembly method (called "Slicembler") for ultra-deep sequencing data (manuscript to appear in *Bioinformatics* and to be presented at *ISMB'15, Dublin Ireland*)
- we have completed the error-correction manuscript (submitted to *Workshop on Algorithms in Bioinformatics 2015* in Atlanta, GA)
- we are still working on the problem of merging BAC assemblies (Step J)

Significant Results:
- Our innovative approach to "slice" the data (divide & conquer on the input reads) dramatically improves read decoding and assembly
- Our new meta-assembler significantly improves assembly quality for

ultra-deep sequencing data
- Our novel algorithm for error-correction is extremely accurate compared to state-of-the-art

Key outcomes or
Other achievements:

*** What opportunities for training and professional development has the project provided?**

This project allowed us to train three PhD students (Computer Science) two of which are female. One MS student (Computer Science), and one undergraduate student (Computer Science) were involved in Year 1. All students have been trained in the domain of computational biology. Specifically,

- PhD student Denise Duma (female) after completing the error-correction project (and her PhD thesis) moved to Baylor College of Medicine, Houston, TX as a post-doc
- PhD student Hamid Seyed Mirebrahim designed Slicembler and was involved in the slicing project to improve read decoding; he is graduating at the end of 2015 and will be looking for a post-doc position
- PhD student Hind Alhakami (female) is currently working on the problem of integrating BAC assemblies with the whole genome shotgun assembly

*** How have the results been disseminated to communities of interest?**

We have produced several manuscripts:

- H. Mirebrahim, T. J. Close, S. Lonardi, "De Novo Meta-Assembly of Ultra-deep Sequencing Data Bioinformatics", to appear in *Bioinformatics*, 2015. In *Proceedings of Conference on Intelligent Systems for Molecular Biology and European Conference on Computational Biology* (ISMB/ECCB'15), Dublin, Ireland, 2015
- S. Lonardi, S. Mirebrahim, S. Wanamaker, M. Alpert, G. Ciardo, D. Duma, T. J. Close, "When Less is More: Slicing Sequencing Data Improves Read Decoding Accuracy and De Novo Assembly Quality", to appear in *Bioinformatics*, 2015
- D. Duma, F. Cordero, M. Beccuti, G. Ciardo, T. J.Close, S. Lonardi, "Scrible: Ultra-Accurate Error-Correction of Pooled Sequenced Reads", submitted to *WABI 2015*, Atlanta, GA
- María Muñoz-Amatriaín, Stefano Lonardi, ... Timothy J Close, "Sequencing of 15,622 gene-bearing BACs reveals new features of the barley genome", submitted to *Genome Biology*

Supporting Files

Filename	Description	Uploaded By	Uploaded On
ISMB15.pdf	Slicembler paper (ISMB 2015)	Stefano Lonardi	05/12/2015
paper.pdf	Error correction paper (submitted to WABI'15)	Stefano Lonardi	05/12/2015
slicing_embedded.pdf	Slicing paper (to appear in Bioinformatics)	Stefano Lonardi	05/12/2015
Manuscript_GB_BACs.pdf	Barley BAC paper (submitted to Genome Biology)	Stefano Lonardi	05/12/2015

Products

Books

Book Chapters

Conference Papers and Presentations

S. Lonardi, D. Duma, M. Alpert, F. Cordero, M. Beccuti, P. R. Bhat, Y. Wu, G. Ciardo, B. Alsaihati, Y. Ma, S. Wanamaker, J. Resnik, S. Bozdag, M.-C. Luo, T. J. Close, (2013). *Combinatorial Pooling Enables Selective Sequencing of the Barley Gene Space*. ISMB/ECCB 2013 Highlights Tracks. Berlin, Germany. Status = ACCEPTED; Acknowledgement of Federal Support = Yes

H. Mirebrahim, T. J. Close, S. Lonardi (2015). *De Novo Meta-Assembly of Ultra-deep Sequencing Data*. International Conference on Intelligent Systems for Molecular Biology and European Conference on Computational Biology (ISMB/ECCB 2015). Dublin, Ireland. Status = ACCEPTED; Acknowledgement of Federal Support = Yes

D. Duma, F. Cordero, M. Beccuti, G. Ciardo, T. J.Close, S. Lonardi (2015). *Scrible: Ultra-Accurate Error-Correction of Pooled Sequenced Reads*. Workshop on Algorithms in Bioinformatics (WABI 2015). Atlanta, GA. Status = SUBMITTED; Acknowledgement of Federal Support = Yes

Inventions

Journals

H. Mirebrahim, T. J. Close, S. Lonardi (2015). De Novo Meta-Assembly of Ultra-deep Sequencing Data. *Bioinformatics*. . Status = ACCEPTED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

María Muñoz-Amatriaín, Stefano Lonardi, ... Timothy J Close (2015). Sequencing of 15,622 gene-bearing BACs reveals new features of the barley genome. *Genome Biology*. . Status = SUBMITTED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

S. Lonardi, S. Mirebrahim, S. Wanamaker, M. Alpert, G. Ciardo, D. Duma, T. J. Close (2015). When Less is More: Slicing Sequencing Data Improves Read Decoding Accuracy and De Novo Assembly Quality. *Bioinformatics*. . Status = ACCEPTED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Licenses

Other Products

Other Publications

Patents

Technologies or Techniques

Thesis/Dissertations

Denisa Duma. *A Combinatorial Pooling Strategy for the Selective Sequencing of Very Large and Repetitive Genomes*. (2013). University of California, Riverside. Acknowledgement of Federal Support = Yes

Websites

Participants/Organizations

Research Experience for Undergraduates (REU) funding

Form of REU funding support:	REU supplement
How many REU applications were received during this reporting period?	Nothing to Report
How many REU applicants were selected and agreed to participate during this reporting period?	Nothing to Report
REU Comments:	Undergraduate student Matt Alpert has been working on the BAC assemblies in Year 1. After being admitted in our PhD program, he left for a job at Turtle Rock Studios.

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Lonardi, Stefano	PD/PI	1
Close, Timothy	Co PD/PI	1
Wanamaker, Steve	Technician	4
Mirebrahim, Seyed	Graduate Student (research assistant)	6

Full details of individuals who have worked on the project:

Stefano Lonardi
Email: stelo@cs.ucr.edu
Most Senior Project Role: PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: project supervision, manuscript preparation

Funding Support: NSF

International Collaboration: No

International Travel: No

Timothy J Close
Email: timothy.close@ucr.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: supervision, preparation of manuscripts

Funding Support: NSF

International Collaboration: No
International Travel: No

Steve Wanamaker
Email: s.wanamaker@sbcglobal.net
Most Senior Project Role: Technician
Nearest Person Month Worked: 4

Contribution to the Project: Sys administrator and programmer

Funding Support: NSF

International Collaboration: No
International Travel: No

Seyed Mirebrahim
Email: smire002@ucr.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 6

Contribution to the Project: PhD student supervised by the PI. Working on the assembly merging step.

Funding Support: NSF

International Collaboration: No
International Travel: No

What other organizations have been involved as partners?
Nothing to report.

What other collaborators or contacts have been involved?
Nothing to report

Impacts

What is the impact on the development of the principal discipline(s) of the project?

The objective of this project is to facilitate the sequencing of large, highly repetitive genomes like the genome of barley (5.3 GB -- twice the size of human) and cowpea (~650 MB). The availability of this new sequencing protocol and the availability of these two genome constitute one the major impact of this project. The development of novel algorithms for decoding, assembly, and error-correction of short reads constitutes the second major impact of this project.

What is the impact on other disciplines?

Cowpea also known as black-eyed pea, is a primary source of protein in the human diet in Sub-Saharan Africa, where it is grown for its foliage, and fresh and dry grains. Outside Africa, cowpea is grown in parts of Asia, Latin America, Southeastern USA and California. Despite its relevance to agriculture in the developing world, cowpea has received scant attention relative to other crops of major global significance. We believe that producing the primary DNA sequence of this important crop will have a significant scientific impact as well enable scientists to select and engineering specific traits that would help ensure high yields from sustainable agriculture into the future.

What is the impact on the development of human resources?

Training of graduate and undergraduate students in computational biology and genomics.

What is the impact on physical resources that form infrastructure?
Nothing to report.

What is the impact on institutional resources that form infrastructure?
Nothing to report.

What is the impact on information resources that form infrastructure?
Nothing to report.

What is the impact on technology transfer?

We are currently exploring whether our technology might have a market (commercial) value.

What is the impact on society beyond science and technology?

The long-term impact of sequencing another organism can be profound once the sequence becomes a routine component of practical problem solving, but in addition the process of sequencing an organism provides precious experience for young scientists who then become our continuity into the future. The organisms that have been sequenced so far represent a minuscule proportion of those which live on our planet, many of which we depend on for our food or against which we must defend to sustain human society. In the plant world for example, the outcome of sequencing an organism could lead to selection and engineering of specific traits that would help ensure high yields from sustainable agriculture into the future. The current food crisis has been fueled by an increase in food demands by developing countries, the spike in oil prices with impact on the cost of fertilizers, disruptions due to climate change and the push to produce biofuels. A global food crisis now severely effects at least 36 mainly developing countries and at least one billion people.

Changes/Problems

Changes in approach and reason for change
Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them
Nothing to report.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects
Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.

Special Requirements

Responses to any special reporting requirements specified in the award terms and conditions, as well as any award specific reporting requirements.
Nothing to report.