

TRANSCRIPT *The Next Generation*

WINTER 2019, ISSUE 2



From the CGRB Director's Desk

Hello everybody! With this second issue of CGRB's new newsletter, I would like to begin a series of short descriptions of CGRB's many areas of activity. Here, I'll talk about big data. Big data is typically characterized as data sets that are too large and/or too heterogenous to be understood without the use of computational tools. Examples include DNA sequence data from people, cancers, plants, animals, microbes, and ecosystems. Big data also includes geographical, weather, climate, environmental, healthcare, financial, commercial, social, and many other kinds of data. In addition to size and heterogeneity, big data sets may be poorly organized (unstructured) and rapidly changing (e.g. financial or Twitter data). The availability of big data is transforming numerous domains including agriculture, food systems, environmental management, social welfare, healthcare, education, transportation, commerce, finance, city planning, tourism, politics, and entertainment. The tools needed to analyze big data come from the fields of statistics, mathematics and computer science, often collected under the umbrella term "data sciences".

(continued on page 2)

In this issue:

Director's Desk
PAGE 1

Microscopy
PAGE 2

Beaver Genome
Update
PAGE 3

Spring Conference
PAGE 4

Spring Workshops
PAGE 4

BUG
PAGE 4

Out of the Past
PAGE 4

TechTips: Time
Saving Tricks
PAGE 5

CGRB In the News
PAGE 5

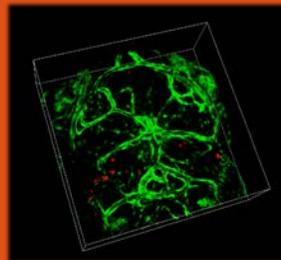
(Director's Desk continued)

Because this new field embraces diverse disciplines, research and especially training in data sciences can be challenging in a traditional university environment where the constituent disciplines are housed in multiple colleges. As a unit that spans all units of the university, and has collaboration as a core value, the CGRB has a major role to play in promoting and supporting data sciences research and training. The CGRB provides analytical tools, high performance computing infrastructure, and data sciences training workshops open to all. The CGRB has also led the establishment of a graduate minor in biological data sciences and a soon-to-be launched undergraduate major on the same topic. We also recently submitted a data sciences Ph. D. training grant proposal to NSF, as well as proposals to NSF and USDA to promote data sciences training for undergraduates and rural communities. Through our efforts in the big data arena, we aim to contribute to OSU's mission, renewed through SP4.0, to deliver exceptional research, discovery, innovation and engagement combined with globally relevant and affordable education.

- Brett Tyler

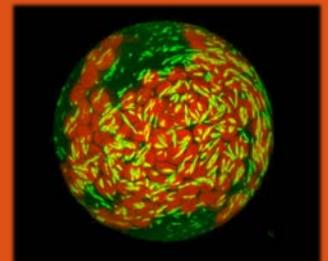
Microscopy

CGRB houses microscopes that are available to anyone for use after training. A Zeiss 780 confocal NLO microscope and a Zeiss Axiovert 200M epifluorescent/widefield microscope are overseen by CGRB while we also house a Keyence BZ-X800 fluorescent microscope overseen by Louisa Hooven in Horticulture. The Confocal system has an hourly rate for use while the widefield is free for use after training. Anne-Marie Girard also offers demo and assisted imaging sessions on either microscope. The confocal can be used on thicker samples to generate optical z-stacks, 3D reconstructions, or time courses. The multiphoton laser is useful for imaging deeper into the sample and can be less damaging. Fluorescence correlation spectroscopy, FCS, a correlation analysis of fluctuation of the fluorescence intensity is also available on the Zeiss 780. Zeiss has a wide range of tutorials on different aspects of microscopy. Here is one that [compares widefield to confocal](#).



3D rendering embryonic zebrafish
John Gamble, Greenwood lab

Arabidopsis protoplast.
Kai Tao, Tyler lab



Available in CGRB Core Labs:
New Robotic preps for Illumina HiSeq

Illumina Nextera DNA Flex Library Prep

Lexogen QuantSeq 3' mRNA-Seq FWD
Library Prep

CONTACT

Mark Dasenko or Anne-Marie Girard
CGRB Core Labs ALS 3012, 541-737-3413

Beaver Genome Project Update

by Brent Kronmiller, PhD

The Beaver Genome Project (BGP), a collaborative OSU project across eight departments, sequenced the beaver genome in order to provide a genetic resource for beaver biology and gave OSU the opportunity to be the first PAC-12 University to sequence their mascot. Through a crowd-funding campaign, the BGP successfully raised \$30,000 to sequence the beaver genome and to obtain a tissue transcriptome atlas. The funding campaign involved the wider Corvallis and OSU community in beaver biology and the scientific capabilities of OSU and the CGRB.

The beaver is a keystone species in riparian ecosystems. By felling trees and building dams, beavers are one of few animals that alter their habitat; creating lasting changes for themselves and other species. Beavers thrive in wetlands due to their unique behavioral, musculoskeletal, metabolic, and dietary adaptations. Through hunting, the North American beaver, *Castor canadensis*, reached near-eradication in much of their range. Through re-introduction, the beaver has regained much of its original habitat.



Filbert, a resident of the beaver exhibit at the Oregon Zoo, donated a DNA sample for the BGP. The CGRB's Illumina HiSeq 3000 completed 3 billion sequence reads of 150bp in length. This raw sequence data was assembled into a 4.1Gb genome (N50 300Kb). A custom beaver specific repeat database was created and used to repeat mask 30% of the genome. We collected sixteen beaver tissues and Adelaide Rhodes profiled the RNAseq to create a beaver transcriptome atlas. The tissue RNA was sequenced, assembled, and combined with the genome assembly for gene predictions, 19,124 genes were identified

Members of the Jaiswal lab investigated orthology and comparative genomics. Phylogenetic analysis of whole-genome orthologous proteins from 19 species shows beaver is found in the rodent group alongside kangaroo rat.

They also investigated gene responsible for the beaver's unique adaptations. Several genes responsible for digit development and preventing webbing are missing in our predicted gene set.

Amita Kashyap in Stephen Ramsey's lab identified long noncoding RNA (lncRNA) genes, recognized as an important class of regulators of tissue-specific gene expression. They identified 40 lncRNAs that are homologous to known lncRNAs in other species and 147 potential novel lncRNAs that did not have any detectable homology to known genes in other species. The 187 noncoding RNAs that they identified include both tissue-specific and ubiquitously expressed transcripts, and the team is now focusing on functional interpretation of these RNAs by analyzing their secondary structure and tissue-specific expression patterns.



Filbert, *Castor canadensis*, Oregon Zoo



Welcome Bo

CGRB welcomes our new Operations Manager, Bo Price. Bo comes to us from UABC. He has nearly 30 years experience in various aspects of financial accounting and reporting, including 23 years in the private sector and over 6 years with OSU. He also has considerable experience with software and IT systems.

SPRING CONFERENCE

Friday
April 19, 2019
CH2M Hill Alumni Ctr

All individuals with an interest in computational biology are invited to attend and participate. In addition to OSU researchers, this all day conference will feature the following speakers:



Susan Holmes
Professor
Dept of Statistics
Stanford University



Katia Koelle
Associate Professor
Dept of Biology
Emory University

The Spring Conference/Colloquium will also feature Lightning Talks and a Poster Session. There will be cash prizes awarded in each category.

REGISTRATION IS NOW OPEN

For additional details and to register, check out cgrb.oregonstate.edu/spring-conference

WORKSHOPS

Spring Term 2019

Introduction to Unix/Linux
Command-Line Data Analysis

Metabarcoding **NEW**

Metagenomics (shotgun sequencing) **NEW**

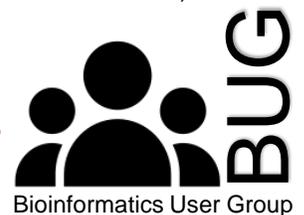
online
Intro to R and RStudio

For additional details and registration information see cgrb.oregonstate.edu/training/workshops

BUG Spring Term 2019

April 3	Yao Yao (Stephen Ramsey Lab)
April 17	Anne Thessen and Kent Shefchek (TISLab)
May 1	Yanming Di (Statistics)
May 15	Greg Goralogia (Steve Strauss Lab)
May 29	Felix Vaux (Kathleen O'Malley Lab)
June 12	Michael Nagle (Steve Strauss Lab)

BUG is at noon in ALS 3005



Out of the Past

The Applied Biosystems Model 370 DNA Sequencer is now operational and is being evaluated. To help us in this evaluation we would like you to send us DNA of known sequence out at least 500 bases. We will run the extensions, sequence the DNA and compare the results. If interested, contact Anne-Marie Girard for a brief consultation on the constraints of automated DNA sequencing.

[Historical note: the ABI 370 ran 16 samples at a time. It gave at most 500 bases of sequence from each sample.]

techtips

CGRB's Time Saving Tricks

by Andrew Black, PhD, Bioinformatics Trainer

We are all busy. So here are a couple of super easy tips to free up some time and increase productivity in a Linux / Unix environment. These small tricks that will help you work faster and more efficiently, both on your local computer and/or cluster!

1. Tab completion

Opposed to typing every letter in a path, use the 'tab' button to autocomplete this. This will save you a huge amount of time, help you find the names of files / directories, and prevent errors!

```
[~]$ /A/really/long/path/to/<tab>
```

2. Cycle through your recent commands

To cycle through your most recent commands (and re-execute or modify) use the "up" arrow

```
[~]$ <most recent command>
```

3. Search your history

To search through your history (and re-execute or modify), hit "control-r" and type in your search word for the command you are looking for. Use 'control-c' to exit!

```
[~]$ cp D199832_R1.fastq ./
bck: D199
```

4. To run a specific command from your history:

First run "history" to print out your session commands:

a) [~]\$ history

```
...
71      10:44      gzip -d fastq.gz
72      10:47      mkdir -p ~/test/directory
73      10:48      nano ./somefile.txt
...
```

b) Now, re-execute a specific command:

```
[~]$ !73
```

c) Which will access the following command:

```
[~]$ nano ./somefile.txt
```

5. Clear the entire line of text

You want to delete your entire line of text, but it takes ages to hold down the 'delete' button. Save time by clearing it with "control-u"

```
[~]$ SomeLineOfCode <Ctrl-u>
```

```
[~]$
```

6. Move your cursor to the beginning of the line

It's amazing how long it can take to move your cursor to the beginning of your line of code with the "left" arrow. Use "Ctrl-a" instead!

```
[~]$ SomeLineOfCode| <Ctrl-a>
```

```
[~]$ |SomeLineOfCode
```

```
↑
```

CGRB in the news



Feb 12, 2019: CGRB teamed with TechData, IBM, OpenPOWER and NVIDIA on a [GPU Architecture demo](#) used at IBM THINK.

Jan 16, 2019: Chris Sullivan, CGRB Assistant Director for Biocomputing, was awarded [IBM Champion 2019](#).

Jan 9, 2019: OSU and CGRB were featured in AMD President and CEO Dr. Lisa Su's keynote address at CES in Las Vegas.