**About**

The Center for Genome Research and Biocomputing facilitates genome-enabled and data-driven research in the life and environmental sciences at OSU and across the state. The Center offers leadership, training and services to faculty, staff and students through expert staff, core laboratories, computational facilities, seminars and technology workshops, and conferences. It also provides a focal point for researchers to establish contacts, initiate collaborations, and apply new technologies in their own laboratories.

Research in the CGRB and faculty affiliate laboratories seeks to improve health, better utilize natural and agricultural resources, understand our global environment, and develop new bio-based products and energy sources. Over 120 OSU faculty, holding primary appointments in academic departments of the Colleges of Agricultural Sciences; Engineering; Forestry; Pharmacy; Science; Veterinary Medicine; Public Health and Human Sciences; or Earth, Ocean and Atmospheric Sciences, are affiliate members of the Center.

**Functions of the CGRB**

The CGRB functions and facilities include:

1. Staffed Core Laboratory facilities that provide a variety of services in genomics, functional genomics, imaging and genotyping;
2. Staffed biocomputing facilities with an extensive high performance computing infrastructure, which includes a managed cloud and shared resources;
3. Computational scientists and bioinformaticists to facilitate or collaborate in computationally intensive research;
4. Entry level and intermediate training workshops, and one-on-one training, in bioinformatics and biological computing, provided by Ph.D. bioinformatics trainers.
5. Seminar Series that features leading scientists in molecular and genomic biosciences and computational biology, and technology training workshops that introduce cutting-edge technologies;
6. Annual Fall and Spring Conferences that brings together faculty, staff and students from across OSU;
7. Campus-wide leadership in genomics, bioinformatics and computational biology.

A key part of the Center is the CGRB Core Facilities that provide services, technical expertise, collaborative functions and share-use equipment for molecular bioscience research at Oregon State University. The Core Facilities are a fully staffed facility that serves as a focal point for acquisition and development of new instrumentation and technologies. A professional staff of nine provides service in four areas:

* ***Genomics*** – DNA sequencing, high throughput sequencing (Illumina), genotyping and fragment analysis;
* ***Biocomputing and Bioinformatics*** – advanced computational resources for data mining, data analysis and database development;
* ***Imaging and Image Analysis*** – a confocal laser scanning microscope facility for high resolution analysis of wide variety of specimens.

In addition, the CGRB provides shared instrumentation, including real-time PCR, scanners, robotics and computational facilities for use by walk-in users.

## Annotation

Annotation, or the prediction of genes that encode proteins (CDS) and other genomic features such as signal peptides, transmembrane regions, ribosomal binding sites, transfer RNAs and conserved motifs/domains, is not easily automated and still needs to be customized based on the organism of interest and available databases. The bioinformatics group can assist in annotation by setting up and running database searches on the command line against the most common resources (NCBI, Pfam, RefSeq, SMART, PRINTS, COGS, Blast2GO, KEGG, etc. - many of these databases are hosted on our local servers for more efficient searching and higher throughput) and collect them into user-friendly tools such as SQLite and Microsoft Excel for downstream analysis. We can also assist in the development of training sets for gene predictions using GeneMark, GeneMaker, FGENESH, Glimmer, tRNAscan-SE, SNAP and Augustus. For unique or previously undescribed features, we can build custom tools for annotation efforts.

## Assembly

Bioinformatics for non-model species often requires de-novo sequencing, either from genomic DNA or transcriptomic cDNA. We provide services and one-on-one guidance to help the community navigate this incredibly broad space, with introductions to genome assemblers like Velvet, SOAPDenovo, CLC-Bio, and others as they are developed. Similarly, we have experience with DeNovo transcriptome assemblers such as Trinity, Oases, SOAPdenovo-trans, and Newbler (for 454 and similar data).

**Bioinformatics Users Group (BUG)**

As much as we can help, we recognize that for most topics, members of the broad CGRB community are the "experts in the field," and more often than not we know who those experts are! We will be sure to connect you with the lab or person that has previously cleared the path for the analysis or data you are facing.   One way we connect people is through the [Bioinformatics Users' Group (BUG)](http://cgrb.oregonstate.edu/bioinformatics-users-group-bug), which meets for an hour, every two weeks.

BUG consists of life scientists, bioinformaticians, computer scientists, mathematicians, engineers, statisticians, and researchers of all types who meet to discuss topics related to these fields of study.

* No experience necessary to participate
* Informal: discussions and interactive-talks
* Short workshops
* Bring/request your own topics of interest

Previous topics include Hidden Markov Models, SNP Calling toolkits, Metagenomics, Structured Query Language (SQL), De-novo genome assembly tools, Project management, and many more.  
  
Our current meeting schedule is every other week on Wednesdays at Noon.  [Calendar](http://calendar.oregonstate.edu/month/cgrb/)

To join the BUG mailing list please [subscribe](http://lists.oregonstate.edu/mailman/listinfo/cgrb-bug) or contact the [CGRB Bioinformatics Trainers](mailto:%20teaching@cgrb.oregonstate.edu).

## Command-Line Unix/Linux

The CGRB infrastructure is heavily based on the Linux operating system and is accessed primarily via remote login on the command-line. Computational jobs are run using the Sun Grid Engine/SGE system.  We can help train users to work with these systems through [classes and workshops](http://cgrb.oregonstate.edu/workshops-and-classes) and [one-on-one training](http://cgrb.oregonstate.edu/training/one-one-training). We can assist if you need help accessing your files and navigating the file system, installing and running software, modifying permissions, writing quick scripts or getting programs running.  We can help!

## Metagenomics

The CGRB Bioinformatics staff can assist in the development of whole shotgun sequencing efforts for ecosystem-level inquiries, environmental clone libraries for functional studies, amplicon variant detection for population analysis and single reference gene-based analysis for community ecology.  Our team has experience in specialized metagenomic tools for assembly (MetaVelvet, Meta-IDBA), annotation (Glimmer-MG, FragGeneScan) and taxon assignment (Metawatt, MetaBin, TANGO) to increase the information recovery from metagenomic environmental samples.

## Microarray Data

Although eclipsed by RNA-seq technology, many labs still have (or plan to produce) microarray data and publications continue to come out with conclusions supported by microarrays. Worry not, we haven't forgotten how to work with these data! We can help normalize and analyze your microarray samples, as well as analyze multi-factorial designs and check sample consistency with PCA.

## Phylogenetic Inference & Population Genetics

Evolution and relationships among genes and organisms can draw from multiple data sources.  Our bioinformaticists can assist in DNA or protein database construction and multiple sequence alignment for phylogenetic inference using distance-based, maximum likelihood and Bayesian modalities.  We can also assist in computational tests of models of evolution and phylogenetic reconstruction, tree topologies, and molecular clock analysis.  Furthermore, we can apply standard or customized statistical tools to population genetic data in order to detect selection pressure, quantitative trait loci, linkage disequilibrium and recombination events.

## Programming

Because research keeps us on the cutting edge and existing software may not accomplish your analysis needs, writing programs is an integral part of modern bioinformatics. We can teach you how to do it or code it for you.  We provide [classes and workshops](http://cgrb.oregonstate.edu/workshops-and-classes) in programming for languages such as R and Python.  We can work one-on-one and guide self-directed learning in other languages like Perl, C, and markup languages like HTML, LaTeX, and Markdown.  If you need someone to do it for you, we can write new programs or alter existing software.

## Sequence Analysis

One of the most basic needs in bioinformatics is comparative sequence analysis. We provide services and training in a wide variety of topics related to sequence analysis, including database searching (e.g. with BLAST, BLAT, and HMMER), pairwise and multiple alignment (e.g. via BioPython modules and tools like Muscle and ClustalW), and reference-guided alignment (e.g. BWA, TopHat/Bowtie).

## Transcriptomics & RNAseq

High throughput RNA sequencing (RNA-Seq) provides information on the location, structure and quantity of genes expressed. RNAseq data from eukaryotes or prokaryote can be mapped to a reference genome or used in de novo (without a reference) transcriptome assembly using assemblers such as transAbyss, Trinity and Velvet/Oases. RNASeq is also useful for the quantification of alternative splicing, detection of allelic variation and for the improvement of genome assembly. Most RNASeq data sets are used for differential expression analysis, a powerful tool that allows for the detection of expression levels among various conditions, cell types and developmental stages. Our bioinformaticists work with researchers to identify the appropriate amount of replication and sequencing depth to recover robust statistical information that supports the biological evidence on the difference in expression in genes of interest, or across the whole genome.

## XSEDE

[XSEDE](http://www.xsede.org) is a collection of computing resources that scientists can use to interactively access and share computing resources, data and expertise. It consists of supercomputers, high throughput computing, storage, cloud computing, software and support for scientific computing.  
  
XSEDE is NSF funded and is available to all Oregon State researchers.  Allocations are available based on a proposal system. A startup allocation can be requested any time. Based on the performance of the code and appropriateness of the computations a proposal can be submitted for a full research allocation (there are 4 proposal periods per year).  XSEDE also offers specialized computing gateways, including bioscience gateways, that can be used at anytime without going through a proposal process.

Oregon State has an XSEDE campus champion, [Steve Wolbers](mailto:stephen.wolbers@oregonstate.edu%20) ([Research Computing](http://is.oregonstate.edu/research-computing) Manager working in Information Services) who can assist in consulting about XSEDE and the available resources, in getting started with the program, and with the proposal process.

### [The Facility](https://shell.cgrb.oregonstate.edu/)

The [CGRB Biocomputing and Bioinformatics facility](https://shell.cgrb.oregonstate.edu/) provides wide-ranging resources, expertise and support for computational needs of the molecular biosciences community at Oregon State University. The facility offers a robust computing environment for high-level computational biology and a versatile intellectual resource for interdisciplinary collaboration.

The first Genome cluster was a compute farm in the Beowulf tradition (2001); i.e., built with commodity hardware and freely available software. Growing to include more than 60 CPU's, Genome served both as a test bed for CGRB developers and system administrators and as an increasingly powerful tool for OSU faculty, staff and students. Learning from setbacks and capitalizing on successes, Genome has steadily grown to include more than 4000 processors. The new version of the Genome is currently has none of the original commodity hardware left. Space and BTU constraints motivated the migration to rack-mount multi-processor nodes and the need to address large amounts of RAM instigated the move to 64-bit hardware. Right now, about 90% of Genome is composed of rack-mount multi-processor nodes that contain 4, 8, 16, 24, 32, 40, 48, 64, 80, 120 and 160 core IBM, AMD or Intel based multi-core processors.  
  
The Center for Genome Research and Biocomputing maintains an extensive and well-managed infrastructure consisting of a distributed service architecture, a greater than 4000-processor computer cluster and a secure private 1G/10G/40G network. Each machine has internal hard drive disk space, but is also connected over 3.5PB of NFS shared disk space. The CGRB encourages high-volume users to contribute to the computational infrastructure. Users are charged $32, $64, and $96 per/month for maintaining each processing machine, web/database server, and file server, respectively. The nodes are provided at the highest priority to the specific project and provided upon request at a lower priority to other CGRB researchers. Thus, subsets of cluster nodes are dedicated to specific research projects, but they function as part of a unified cluster when needed for intensive jobs. This priority-based scheduling has proven quite successful, both in terms of end-user satisfaction and in the execution of systems administration activities. Computational requirements are constantly re-evaluated and new hardware is integrated as needed. Researchers who do not own equipment and can not complete their work within the resources provided to general users can rent resources from the CGRB as needed to complete the work on time. Research rental hardware includes processing machines (including machines with 1TB of RAM), web/database servers and file space to store data. Please see [Biocomputing Service Fees](http://cgrb.oregonstate.edu/biocomputing/biocomputing-service-fees) for pricing and [email](mailto:support@cgrb.oregonstate.edu) with questions.