

# Introducing NCGR Informatics

John A. Crow, Ph.D.

Vice President for Informatics

National Center for Genome Resources

[jac@ncgr.org](mailto:jac@ncgr.org)

## NCGR Research

Sponsored research  
*PIs, postdocs, visiting scientists*

## NCGR Informatics

Academic and business partnerships  
*Bioinformatics, software development, IT*

## NCGR Sequencing Services

NCGR scientists and external collaborations  
Illumina sequencing & genotyping, PacBio sequencing



N C G R

National Center for Genome Resources



## The teams

- Computing and Networks
- Software Development
- Bioinformatics

Formally established early 2011.

# Computing and Networks

- Administration of HPC assets
- Database servers and administration
- Disk arrays and long term storage
- High performance internal networks

John Utsey  
Forrest Black  
Kathy Myers

# Software Development

- Project-based databases and web resources
- Development of internal processing pipelines
- Internal LIMS
- Evaluation of new software technologies

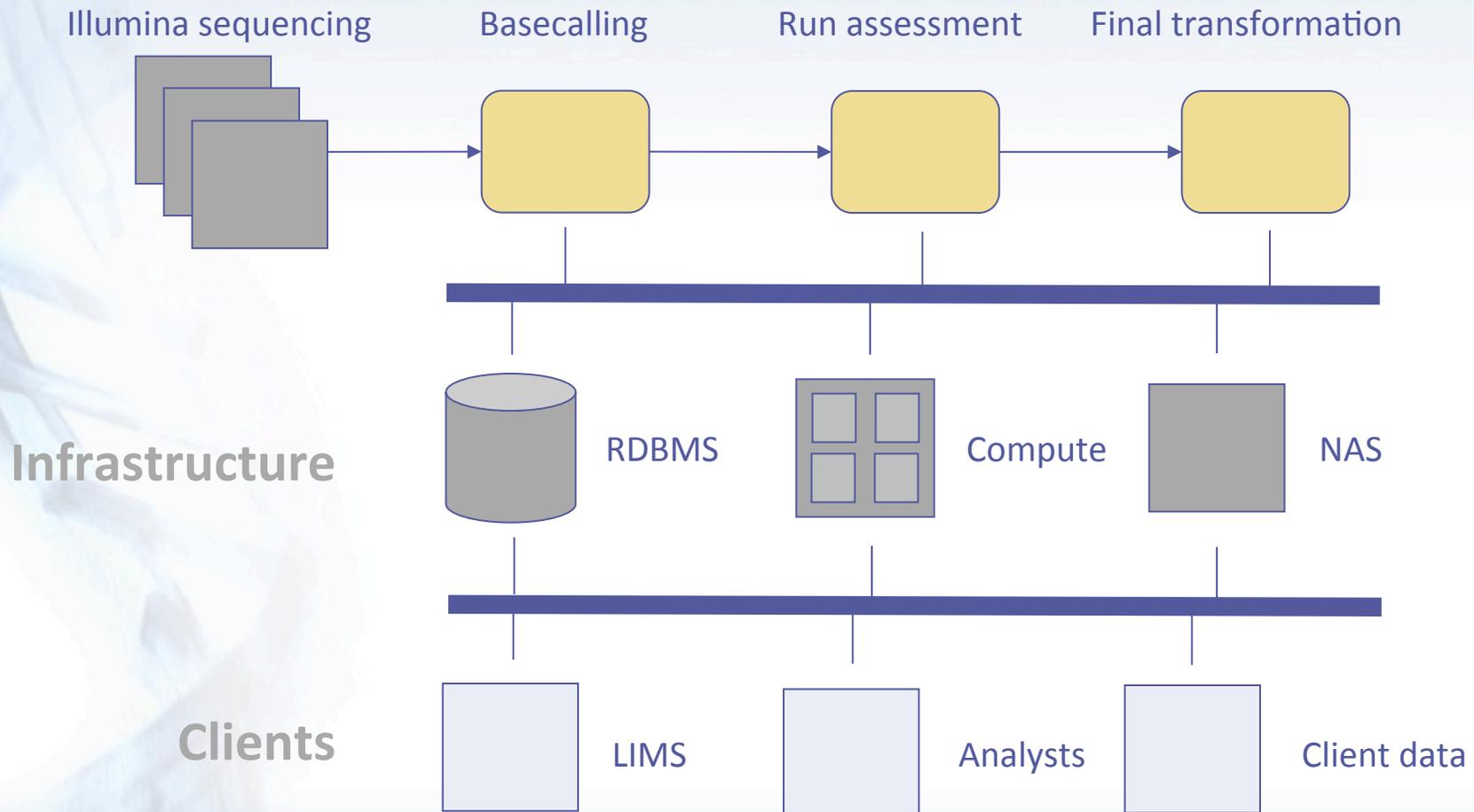
Ken Seal  
Alex Rice  
John Crow



NCGR

National Center for Genome Resources

# Sequencing run postprocessing pipeline



# Grindstone

Home | Sequencing: Runs Flowcells Pools Libraries Samples | Projects Contacts

John Crow | Logout | 15 minutes

## Run 120426\_SNPSTER\_0909\_64RRCAAXX\_PE

[Edit](#) | [1st Cycle Report\(s\)](#) | [Runs index](#)

Instrument	Flowcell	Run Operator	Started	Completed	Cycles
SNPSTER	<a href="#">64RRCAAXX</a>	Pooja Umale	2012-04-26	2012-05-03	2x54 paired end

<a href="#">Lab Tracking Worksheet</a>	<a href="#">First Cycle Report(s)</a>	<a href="#">Analysis 2012-05-03</a> <a href="#">[details]</a>
<a href="#">New worksheet</a>	<a href="#">909_FCR_Read1.pdf</a>	<a href="#">GERALD Summary: 2012-05-03</a>
		<a href="#">Quality Graph</a>

Comments: [New Comment](#)

[Run done.](#) Peter Ngam 2012-05-03 08:05am

Lane	Library	Concentration	Project
1	 <a href="#">AL_lib_idx10</a>	7.0 pM	<a href="#">May - NIH - Medicinal plants</a>
2	 <a href="#">AR_lib_idx8</a>	8.0 pM	<a href="#">May - NIH - Medicinal plants</a>
3	 <a href="#">CCL_lib_idx3</a>	8.0 pM	<a href="#">May - NIH - Medicinal plants</a> <a href="#">pearl</a>
4	 <a href="#">CEPHA_plant_Drupacea_lib_idx1</a>	8.0 pM	<a href="#">May - NIH - Medicinal plants</a> <a href="#">pearl</a>
5	<a href="#">PhiX-Control</a>	2.0 pM	<b>control</b>
6	 <a href="#">CF-L_lib_idx1</a>	7.0 pM	<a href="#">May - NIH - Medicinal plants</a> <a href="#">pearl</a>
7	 <a href="#">TAXME_entire_tree_lib_idx3</a>	8.0 pM	<a href="#">May - NIH - Medicinal plants</a> <a href="#">pearl</a>
8	 <a href="#">TrHr_lib_idx9</a>	8.0 pM	<a href="#">May - NIH - Medicinal plants</a> <a href="#">pearl</a>

Grindstone  
Internal LIMS



# LIS

## Legume Information System



Select Language

Powered by Google Translate

Home | About | News | Search | Maps | Species | Collaborators | Resources

### Entry Points

#### Species Access and Data

Information, resources, and datasets for many legume species.

Go



#### Seqple Sequence Search

Do multi-sequence queries against the reference legume genomes and other databases (UniProt, expression database, others).

Go

Pseudomolecule  
Hit positions 446  
Hit score 364,

#### Leggleg Gene Family Pages

Searchable gene families, alignments, and gene trees for the legumes and outgroup species.

Go



#### CViT Genome Search & Sy

Search for sequences or genes

### Welcome

The mission of LIS is to facilitate discoveries and crop improvement in the legumes — critical components of global agriculture. Please send us your questions, suggestions, and ideas for improving this site! [lis\\_feedback@ncgr.org](mailto:lis_feedback@ncgr.org)

### Meetings

Molecular Breeding of Forage and Turf 2012,  
Salt Lake City  
Jun 03-07, 2012

National Association of Plant Breeders  
Aug 06 - Sep 08, 2012

**Legume Information System**  
<http://comparative-legumes.org>



National Center for Genome Resources

# Bioinformatics

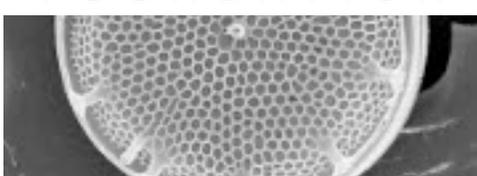
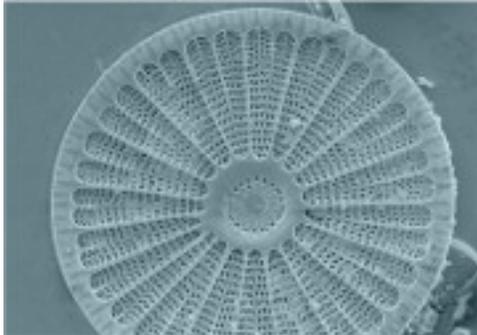
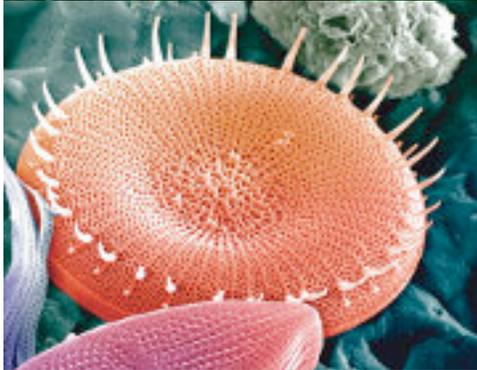
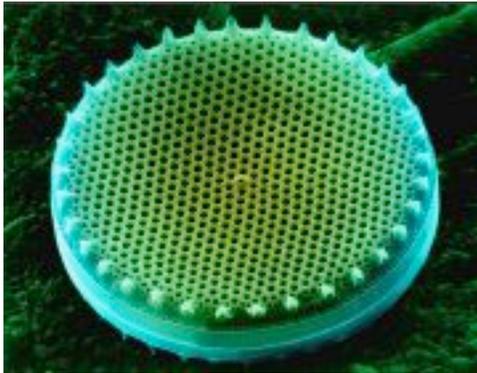
- Project-based data analysis and interpretation
- Development of analysis methodologies
- Experimental design

Andrew Farmer  
Thiru Ramaraj  
Robin Kramer  
Connor Cameron



N C G R

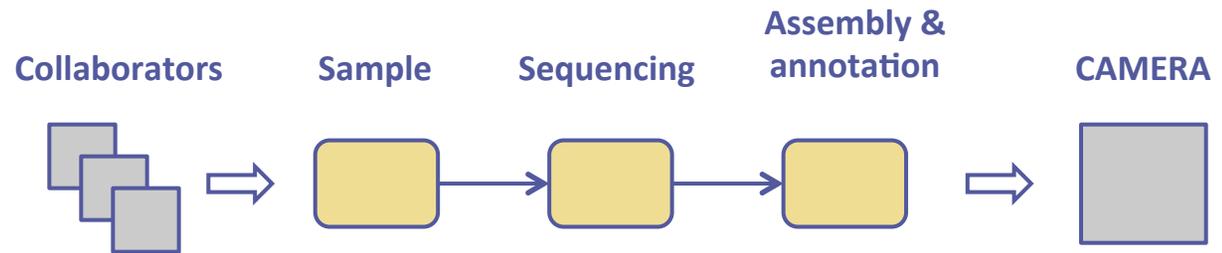
National Center for Genome Resources



# Marine Microeukaryote Transcriptome Sequencing Project

Whole transcriptome sequencing of 750 microeukaryotes  
Callum Bell (NCGR), Arvind Bharti (NCGR)

Ongoing: December 2010 – present



Development of improved pipeline for high throughput RNA-Seq assembly (Robin Kramer, Connor Cameron)

<http://marinemicroeukaryotes.org>

# Medicinal Plants Consortium

Washington State University – Danforth Center – University of Illinois, Chicago – NCGR

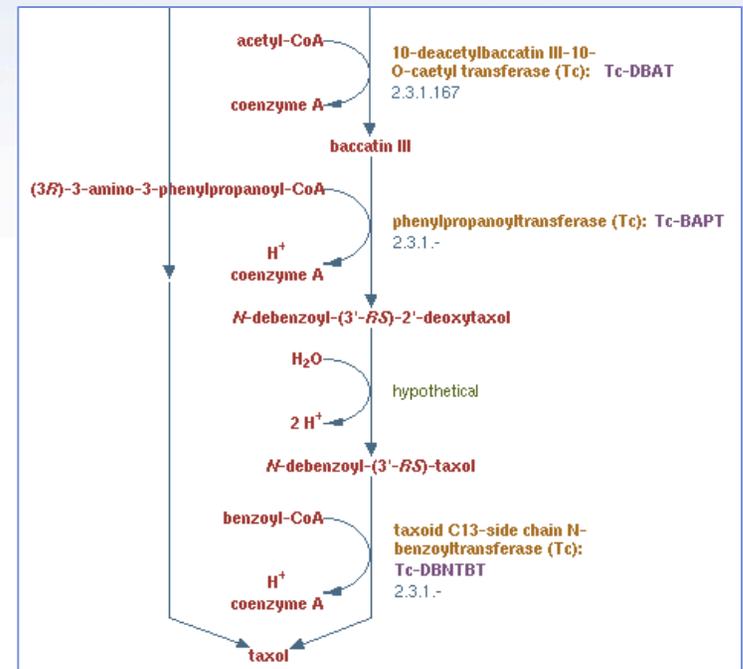
Whole transcriptome sequencing of 25+ species of medicinal value

**Taxus spp.** – production of paclitaxol (Taxol) for anticancer treatment (breast, ovarian, lung)

**Papaver somniferum** – “opium poppy”

**Digitalis lanata** – production of digoxin (Lanoxin) for treatment of atrial fibrillation, atrial flutter.

Transcriptome assembly, annotation, analysis and completion of biosynthesis pathways, identification of coexpression networks

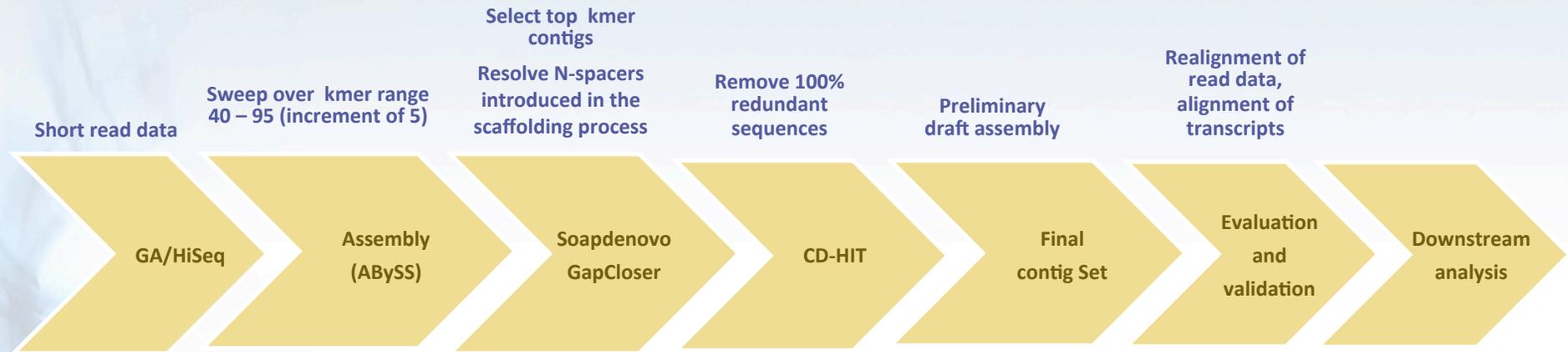


<http://medplants.ncgr.org>



**Medplants**   
Medicinal plant transcriptomics

# Baseline genomic assembly



Joann Mudge, Thiru Ramaraj, Robin Kramer, Arvind Bharti

## Modifications per project

Incorporation of additional sequence data (454, Sanger)

Iterative resc scaffolding

Manual improvement and correction

Phased assemblies: sequence → assemble → assess & repair → sequence → ...

*Gossypium arboreum* genome assembly using a sequence-based physical map (with Texas Tech University)

Alfalfa (*Medicago sativa* L.) genome sequencing (with the Noble Foundation)

Sequencing of the chocolate (Cacao) genome (with USDA and Mars, Inc.)

# National Center for Genome Resources

Santa Fe, New Mexico USA

<http://ncgr.org>

[info@ncgr.org](mailto:info@ncgr.org)