# The UC Davis Genome Center, Next-Generation Sequencing And Plant Breeding

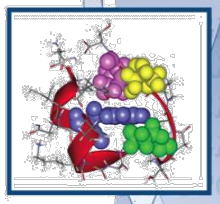
Richard Michelmore
The Genome Center, UC Davis

http://genomecenter.ucdavis.edu
http://michelmorelab.ucdavis.edu

UCDAVIS
GENOME AND BIOMEDICAL SCIENCES FACILITY

GENOME CENTER

A global approach to biology



## Wednesday, November 9th

| Time (Pacific time) | Event |
|---|---|
| **12:00 – 1:00 PM**<br><br>Location: Plant & Environmental Sciences Building (Room 3001)<br><br>Click here to register | **Bonus presentation**<br><br>  <br><br>**Speaker**:<br>**Dr. Kent Bradford**<br>Director, Seed Biotechnology Center (SBC), UC Davis<br><br>Dr. Bradford will report on the Gene Flow and Co-existence workshop held in Washington, DC on September 7 and 8, 2011. He will summarize discussions among industry and academia experts on current and emerging strategies to minimize gene flow and maintain seed purity across agricultural communities. |
| **2:00 – 3:45 PM**<br><br>Location: Plant & Environmental Sciences Building (Room 3001)<br><br>Click here to register | **Seed Central SCIENCE**<br>*Research & Technology presentations*<br><br>**Topic**: Strategies for using modern genomics in plant breeding<br>**Speakers**:<br><br>  <br><br>**Dr. Richard Michelmore**<br>Director<br>The Genome Center, UC Davis<br><br>**Dr. Allen van Deynze**<br>Director of Research<br>Seed Biotechnology Center (SBC)<br>UC Davis<br><br>  <br><br>Dr. Michelmore and Dr. van Deynze will share the Genome Center's and the SBC's strategies for integrating modern genomic-assisted approaches to plant breeding practices. |
| **4:00 – 4:45 PM**<br><br>Location: Plant & Environmental Sciences Building (Room 3001) | **Seed Central BRAINSTORMING**<br>*Brainstorming session*<br><br>**Topic**: The potential for establishing on the UC Davis campus collaborative, pre-competitive research facilities focused on research and technologies for using modern genomics in plant breeding. |
| **5:00 – 7:30 PM**<br><br>Location: Buehler Alumni and Visitors Center<br><br>Click here to register | **Seed Central FORUM**<br>*Networking event with featured speaker*<br><br>  <br><br>**Featured speaker (6:00 - 6:45)**:<br><br>**Dr. Jorge Dubcovsky**<br>UC Davis Professor and HHMI-GBMF researcher<br><br>Dr. Dubcovsky's research illustrates the power of modern biotechnology approaches in delivering new tools to breeders. In his keynote address, he will share the successes of the USDA-NIFA<br><br>**Sponsored by**<br><br>ENZA ZADEN |

# THE CENTRAL RESEARCH DOGMA

Basic research  Translational research

**DATA** → **KNOWLEDGE** → **SOCIETAL BENEFIT**

Transitioning from a data poor to data (over-)rich reality
Need to adjust research, teaching, and application strategies
Access to large amounts data is no longer limiting;
People's (researcher's) time and attention are.
Need to be more concerned with false positives
and data quality than with false negatives

# The Transition from a Data-Poor to a Data-Rich Reality

**GENOTYPE + ENVIRONMENT => PHENOTYPE**

Controlled input

Variable component

Consequential output
multiple aspects & levels

New measurement opportunities:

High-Throughput
Sequencing
& Genotyping

Detailed
metadata
collection
technologies
GIS
What to
measure?

RNA: arrays, -> sequencing
Proteins: MS -> arrays?
~500 -> ~200,000+ proteins
Metabolites: MS profiling, -> ID
endogenous & exogenous cpds
Live imaging: Whole organism
-> single cells

# The UC Davis Genome Center: UCD's Response to the Opportunities and Challenges of the Genomics Revolution

Biological paradigms being worked out in model species
Need to use genomic information for medical and agricultural benefits
Comparative functional genomics increasingly informative
Big strength at UC Davis = diversity of organisms studied
UC Davis = one of largest & most diverse biology campuses in the world

Rapidly changing technologies
Biology becoming more computational, data poor to data rich
Genetic components of biological/medical studies essential

New genomics technologies expensive and optimal ones uncertain
Centralized access to enabling technologies on as-needed, at-cost basis
Ability to generate and manipulate very large datasets
Can spend money much faster, even if cost per data point very low

**http://genomecenter.ucdavis.edu**

UCDAVIS genome center

**Objectives**:

To ensure diverse research campus-wide remains current.

To house ~8 bioinformatics & 8 wet lab, technology-driven faculty.

To provide genomic technologies at-cost, as-needed through service cores:

DNA Technologies, Expression Analysis, Proteomics, Metabolomics, Bioinformatics

**Rationale:**

No longer possible for a single investigator to do everything
Need multi-disciplinary capabilities at-cost, as-needed.
Not “Big Biology”.  Extension of individual labs.
Enable innovation at individual level.
Good ideas should not be technology limited

Technology antenna for genomics on campus

**http://www.genomecenter.ucdavis.edu**

# The Transition from a Data-Poor to a Data-Rich Reality

**GENOTYPE + ENVIRONMENT => PHENOTYPE**

Controlled input

Variable component

Consequential output
multiple aspects & levels

New measurement opportunities:

High-Throughput
Sequencing
& Genotyping

Detailed
metadata
collection
technologies

GIS

What to
measure?

RNA: arrays, -> RNAseq

Proteins: Mass Spec -> arrays?
~500 -> ~200,000+ proteins

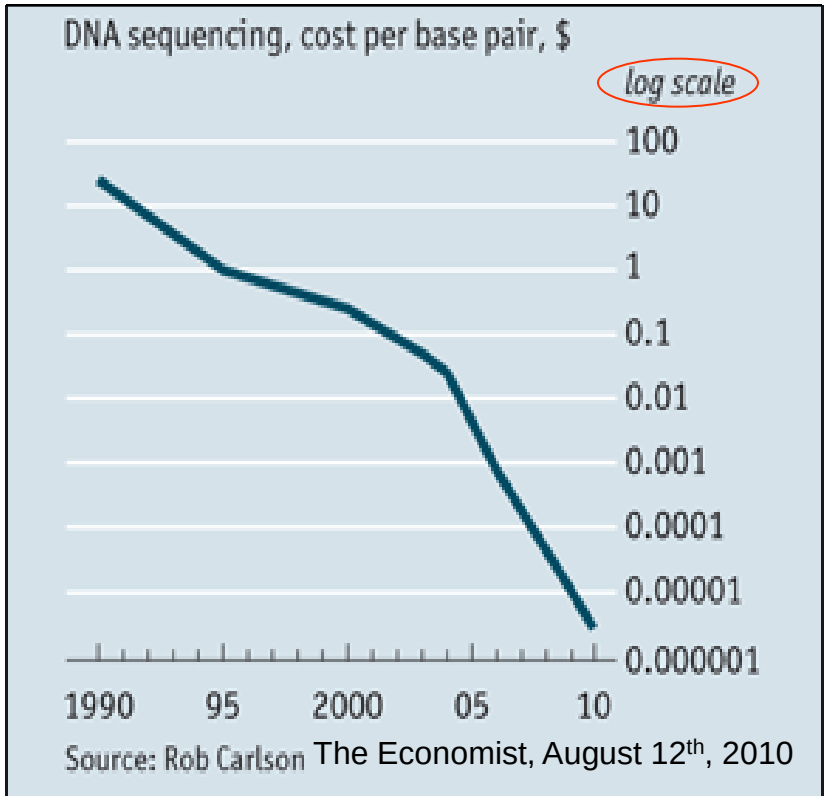
Metabolites: MS profiling, -> ID
endogenous & exogenous cpds

Live imaging: Whole organism
-> single cells

Impacted by next-gen
sequencing

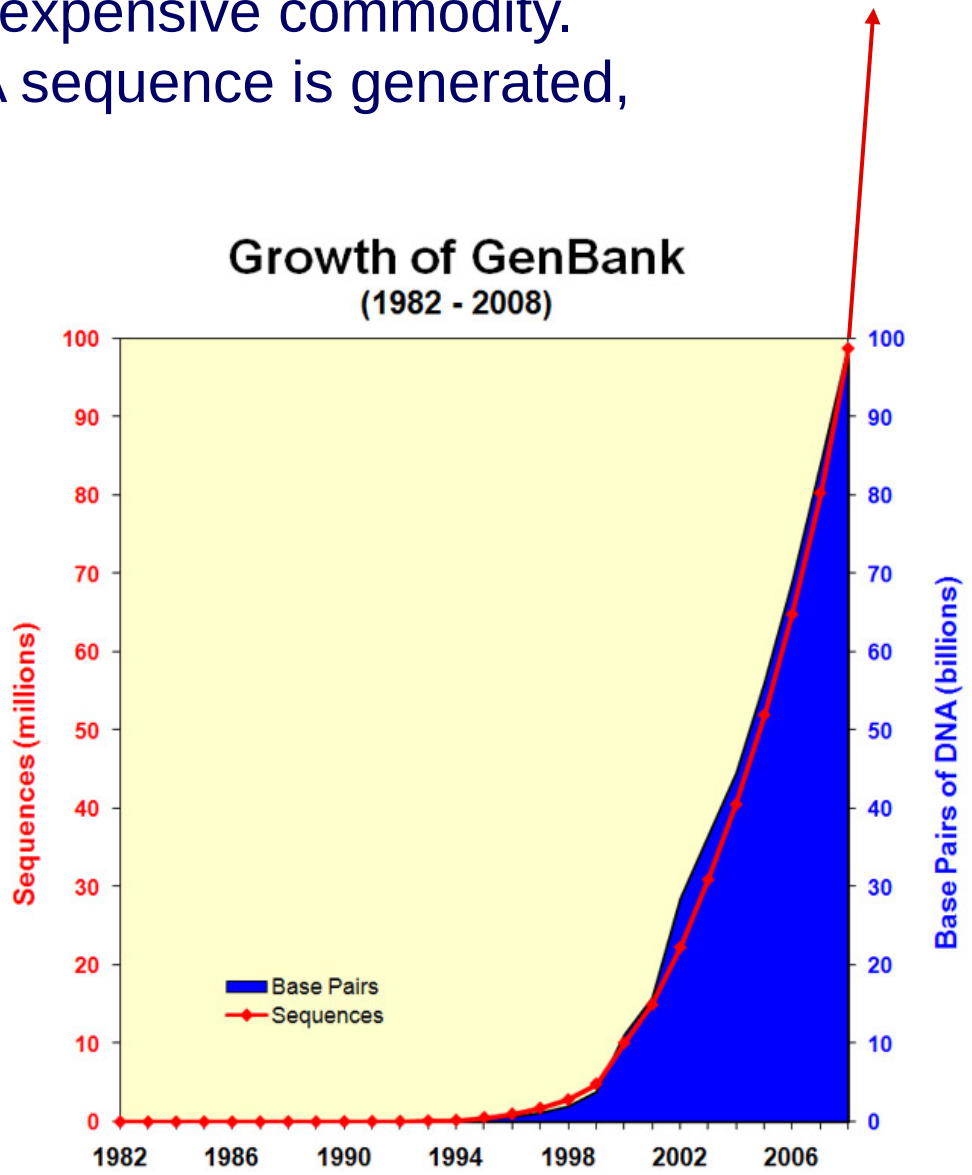
# DNA sequence becoming an inexpensive commodity. New paradigms as to how DNA sequence is generated, handled and valued.

Decreasing cost of sequencing
**(1990 – 2010)**




The Economist, August 12th, 2010

**Growth of GenBank**
**(1982 - 2008)**




Gary Shroth (Illumina): “A single lab with one HiSeq can generate as much sequence as was in GenBank in 2009, every four days”.

# Modes of Interaction & Recharge Categories

1) Recharges per sample for standard analyses. Results ~guaranteed. GC assumes risk. Actual costs of consumables, labor, & service contracts. Minimal equipment and administrative overhead.

2) Research/exploratory mode. Recharges for units of project time, at cost. Results not guaranteed: only 'will do best'.. Researcher assumes risk.

3) Annual training fee for 24/7 access to non-production machines for high-volume users.

4) Seed/pilot projects ($2,000). Calls every ~6 months. Preliminary data for grant proposals

Priority: Campus > Off-campus > Commercial Clientele

(cost + OH) (cost + 53% OH) (cost + 80% OH)

# LATEST GENERATION SEQUENCING AND GENOTYPING AVAILABLE IN DNA TECHNOLOGY AND EXPRESSION CORES

## Massively parallel DNA sequencing

- 2 Illumina Genome Analyzers
- 1 Hiseq 2000, 1 Miseq
- 1 Roche 454 Junior
- 1 Pacific Biosystems RS

## Uses

- Transcriptome: Gene discovery
- Resequencing: Diversity analysis
- Highly efficient SNP discovery
- Expression analysis (RNAseq)
- Whole genome sequencing
- Metagenomics
- Segregation analysis: assembly validation


## GoldenGate SNP genotyping

- iScan, BeadArray & BeadExpress

Ryan Kim, Core Manager

# Examples of Organisms Genotyped in the DNA Technologies Core using Illumina GoldenGate SNP Assay

Example Project: David Neale, Plant Sciences, CA&ES. High Throughput Genetic Analysis of Trees. USDA CAP $5.9M, $1M to GC DNA Tech.

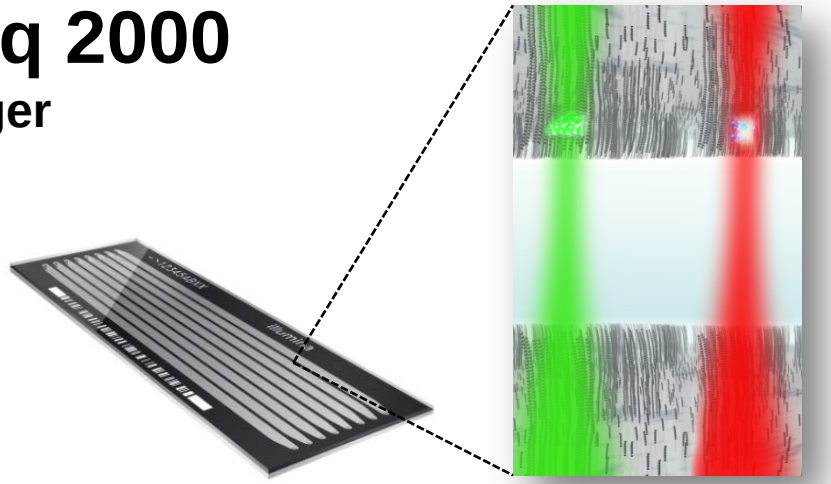
10,000 pine trees = ~all potential trees used for breeding in USA being scored for 7,600 genetic markers (often in candidate genes) per individual.

Scored for plant characteristics e.g. growth and wood quality throughout US.

Analyzed for metabolic characteristics by Metabolomics Core.

# Illumina HiSeq 2000

**Game Changer**



Dual surface imaging

Fast scanning and imaging

Two flow cells run in parallel

Capable of 600 Gb per run -> 1,000+ Gb

Run time 11 days for 2 x 100 nt

55 Gb/day

6 billion paired-end reads

<$5,000 per human/lettuce genome

<$200 per transcriptome

Challenges: Library preparation & data analysis


Gary Shroth (Illumina): “A single lab with one HiSeq is able to generate more sequence as was in GenBank in 2009, every four days”.

# THIRD GENERATION SEQUENCING

PACIFIC BIOSCIENCES™

http://pacificbiosciences.com

Single Molecule Real Time (SMRT™) sequencing
Recording natural DNA synthesis by DNA polymerase as it occurs
Single molecule resolution
Simple amplification-free sample prep
Long reads, average read over 1kb, Poisson distribution to 3 – 4+ kb
Fast, 1 to 3 bases incorporated per second
Sample prep to data analysis in less than a day
Low overall costs
160,000 Zero Mode Waveguides (ZMWs) per SMRT cell
~33% of ZMWs have only one polymerase
15% error rate (indels)
Complementary to Illumina
Not for counting large numbers of tags or sequencing large genomes

UCDAVIS
UNIVERSITY OF CALIFORNIA

华大基因
BGI

**The BGI@UC Davis Partnership:**

**Transforming Data into Knowledge for Societal Benefit**

West Coast sequencing facility in UCD School of Medicine.
Immediately: 3 + 2 CLIA Hiseq2000s. 2012: 10 to 20 Hiseqs.

Research collaborations: human & animal health, global food security, bioenergy, biodiversity, and environmental sustainability.

Educational exchanges.

Access to sequencing and computational resources:

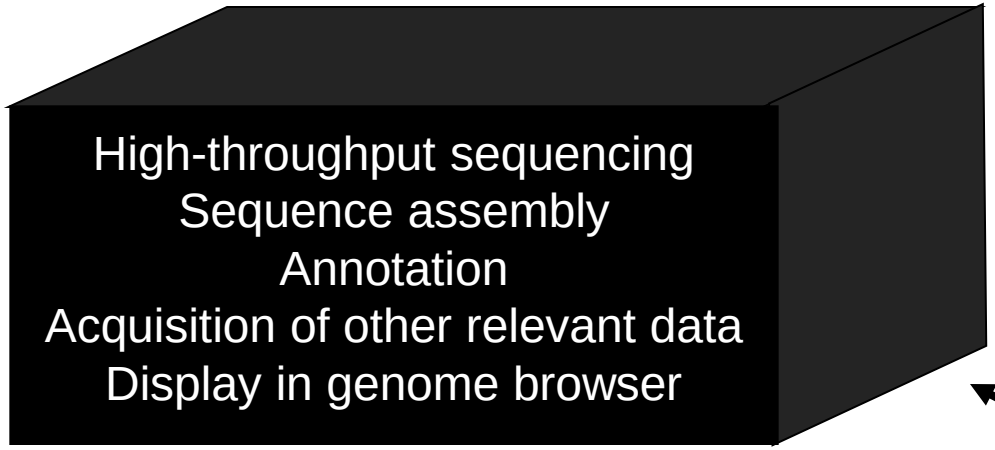
5 Tb/day = 1,600 human 1x, 312 wheat cultivars, 1 M bacteria

500 Tb in 2010 (= 10x NCBI over past 20 years), 100 Pb in 2011

Portal for other institutions.

# UC Davis Sequencing & Gene Expression Service Cores

Tissue, DNA or RNA samples brought to DNA Technology Core




*Integrated activities of DNA Technology & Bioinformatics Cores and BGI@UC Davis*

华大基因
BGI

Researcher queries samples versus existing information over web

## In near future: DNA sequence = an inexpensive commodity generated on a variety of platforms

$1,000 ($100?) human genome coming =>

- $1,000 genome for many animals and plants
- $100 genome for fungi
- $10 genome for bacteria *en masse*

Metagenomics: sequencing of communities

- biomes (humans = 100x more bacteria)
- novel & unculturable organisms
- characterization of diversity & unique genes

Not just genomic DNA sequence:

- DNA modifications
- epigenomics & copy number variation (CNV)
- expression analysis (RNAseq not arrays)

Enormous amounts of sequence data

Need for major data handling capabilities

Vital role for bioinformatics

**The Challenge and Opportunity:**

**How to utilize the deluge of sequence data?**

# SEQUENCING APPLICATIONS: REVOLUTIONIZING BIOLOGY

## Genomic sequencing

*De novo*

- Microbial, animal and plant diversity
- Novel & unculturable organisms
- Biomes (bacteria = 100x human)
- Novel genespace

Re-sequencing

- SNP and CNV discovery, TILLING
- Gene cloning, novel allelic diversity
- Genome Wide Association Studies (GWAS)
- High resolution population genetics

Mapping, genotyping by sequencing

BSAseq

## Gene regulation

Transcriptome sequencing for gene models and splicing

RNAseq for expression analysis

Small and non-coding RNAs

Ribosome profiling

CHIPseq for DNA binding sites

DNA modifications and epigenomics

# Gene discovery & genotyping by sequencing of non-model organisms e.g. plant pathogens

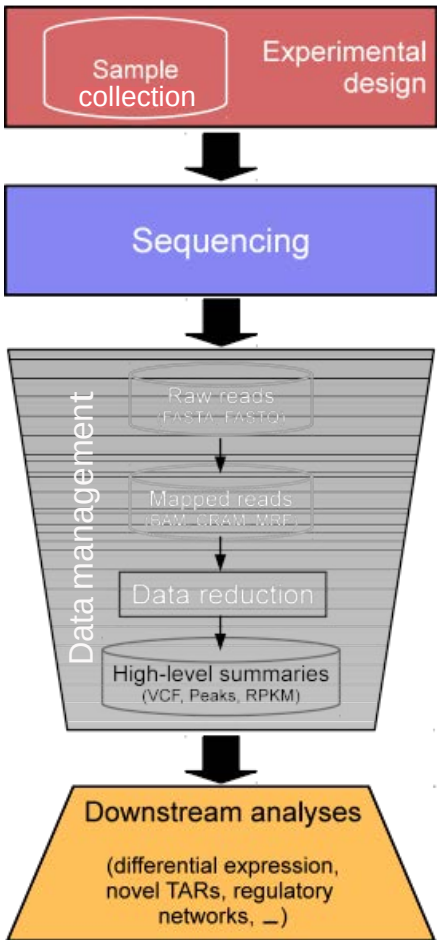
E.g. Cloning of 10 genes from wheat rust, *Puccinia striiformis* f.sp. *tritici* (collaboration with J. Dubcovsky)

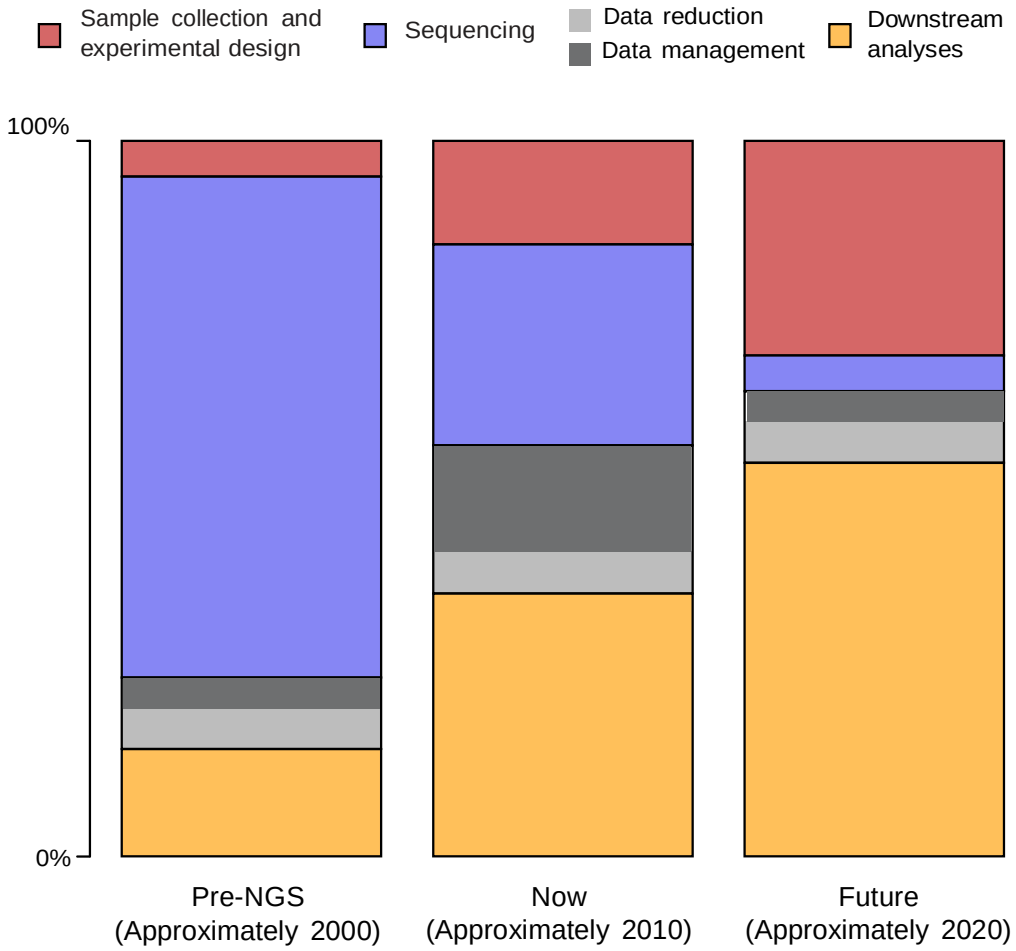
Whole genome sequencing (<100Mb genome)
Two lanes Illumina GA to ~60x.
Quicker, cheaper, more informative than gene-by-gene.
Permanent resource for other studies.
Genotyping by low-coverage sequencing possible.

Cantu *et al*., 2011. Plos One

# The real cost of sequencing: higher than you think!

Sboner *et al*. *Genome Biology* 2011, **12**:125



COMPUTATIONAL CHALLENGES FOR CAMPUS (& ELSEWHERE):

Need to greatly increase computational resources.

Need local CPU and GPU compute clusters housed in a dedicated, efficient facility with critical mass of diverse staff.

Need to promote synergy between big data users on campus.

Support for efficient use of resources, e.g. adaptations to GPUs.

Need massive amounts of storage with reliable back-up.

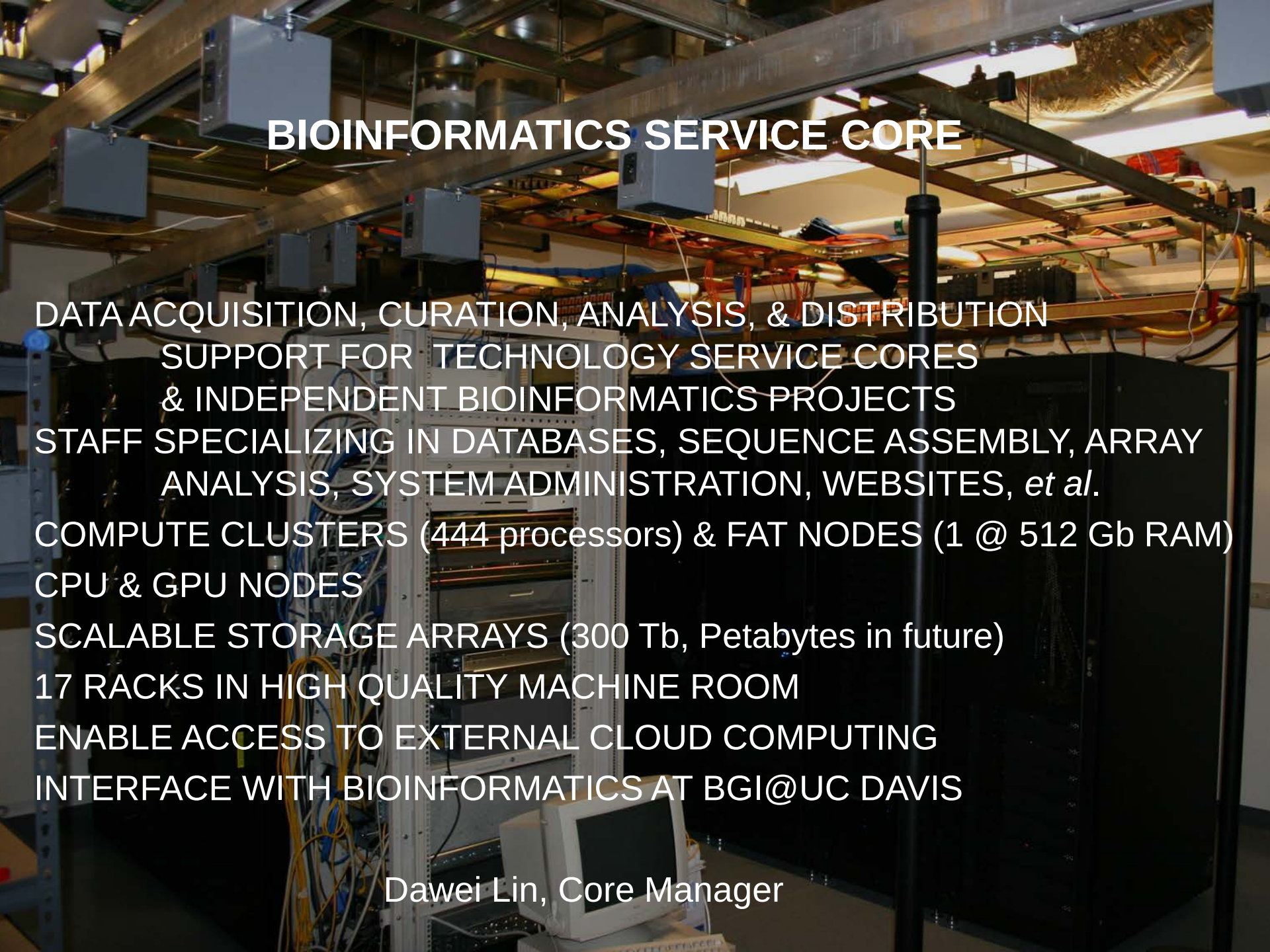
Need access to off-campus cloud and super-computing facilities.

Need efficient methods of moving large amounts of data to off-campus resources.


terabyte $10^{12}$
petabyte $10^{15}$
exabyte $10^{18}$
zettabyte $10^{21}$

Petabytes in 2008… exabytes in 2011…?? in 2014...?? in 2017?

We need the proper equipment and we had better get it right.



# BIOINFORMATICS SERVICE CORE

DATA ACQUISITION, CURATION, ANALYSIS, & DISTRIBUTION
SUPPORT FOR TECHNOLOGY SERVICE CORES
& INDEPENDENT BIOINFORMATICS PROJECTS
STAFF SPECIALIZING IN DATABASES, SEQUENCE ASSEMBLY, ARRAY ANALYSIS, SYSTEM ADMINISTRATION, WEBSITES, *et al*.
COMPUTE CLUSTERS (444 processors) & FAT NODES (1 @ 512 Gb RAM)
CPU & GPU NODES
SCALABLE STORAGE ARRAYS (300 Tb, Petabytes in future)
17 RACKS IN HIGH QUALITY MACHINE ROOM
ENABLE ACCESS TO EXTERNAL CLOUD COMPUTING
INTERFACE WITH BIOINFORMATICS AT BGI@UC DAVIS

Dawei Lin, Core Manager

# Data-Poor to Data-Rich Environment

Biology becoming increasingly computational

Vast data sets from sequencing, genotyping and phenotyping
- Acquisition, curation, interrogation, integration, distribution
- Connectivity, data structures, computing speed and architecture, storage (archiving and retrieval)

Challenges: Data -> Knowledge
- Identification of signal from noise
  - Relevant from irrelevant data: search tools
  - Good vs. poor relevant data: metadata and curation
- Evaluation of significance of good, relevant data:
  - Query and statistical tools
  - Educated professionals and recipients
- Utilization of knowledge
  - Integration with existing work flows
  - Development of novel products and workflows
  - Dealing with the implications of (uncertain) inferences

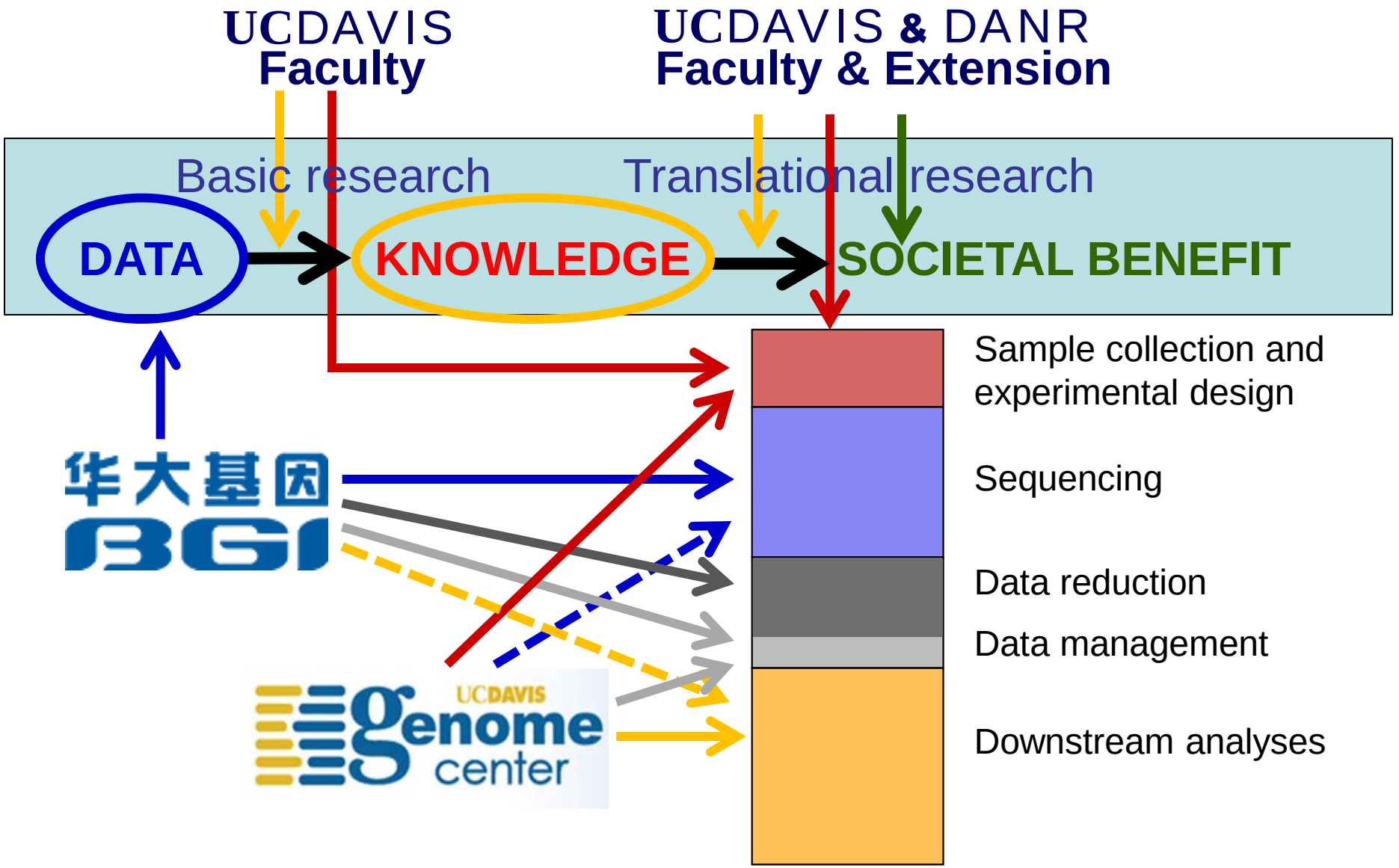
# Implications of the Genomics Revolution

Very large amounts of sequence information for many organisms
Genome sequences of individuals will become increasingly available
Genetic predispositions for many normal and pathological traits
The molecular basis of numerous normal
and pathological states will be understood
Intervention strategies will lag behind knowledge of
genetic predisposition and molecular understanding
High resolution and very large datasets on genotypes and phenotypes
at multiple levels: protein, metabolite, whole organism.
Need to consider phenomena at the systems level.

## Challenges:

Including a genetic component into many types of studies
Sorting the signal from the noise
Converting data to knowledge
Staying current, training & sustaining sufficiently informed practitioners
Conveying the useful information to the patient/public
Dealing with the implications of (uncertain) inferences
[when intervention strategies not available (yet)]
Societal/ethical/legal issues of confidentiality and (mis)use

# INTEGRATION OF BGI AND UC DAVIS STRENGTHS



# GENOMIC OPPORTUNITIES FOR BREEDING COMPANIES

Basic research Translational research

**DATA** → **KNOWLEDGE** → **SOCIETAL BENEFIT**

Pre-competitive research Commercial application

Collaborative, pre-competitive, knowledge-generating projects.

Development of ultra-dense, gene-based genetic maps with agriculturally phenotypes.

Genome sequencing.

Analysis of crop diversity.

Identification of candidate genes for breeding targets.

Marker-assisted selection.

Monitoring pathogen populations to direct resistance gene deployment.

# Genetic Mapping & Marker Development

Old paradigm (slow and inflexible):

- One-by-one marker development.
- Utilization of core set of reference markers.

Current paradigm (faster but specific to populations):

- Sequence transcriptome of parents to identify 10,000s of SNPs.
- Develop informative SNP panels for specific sets of crosses.
- Run SNPs on segregating individuals.

Latest paradigm (fast, flexible, & highly informative):

- Sequence segregating individuals.

Rate limiting steps:

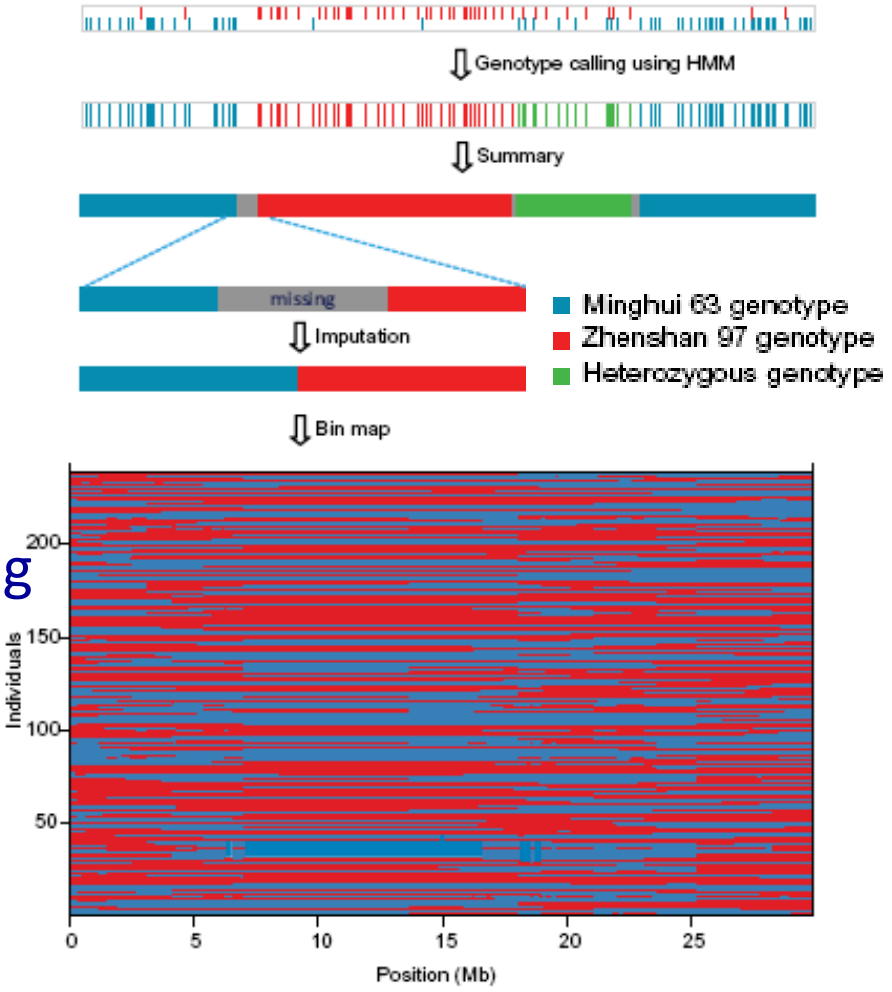
- Informative populations.
- Accurate phenotyping.
- Library preparation.
- Sequencing not limiting.
- Data analysis.

# Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Xie et al., 2010.* PNAS 2010.

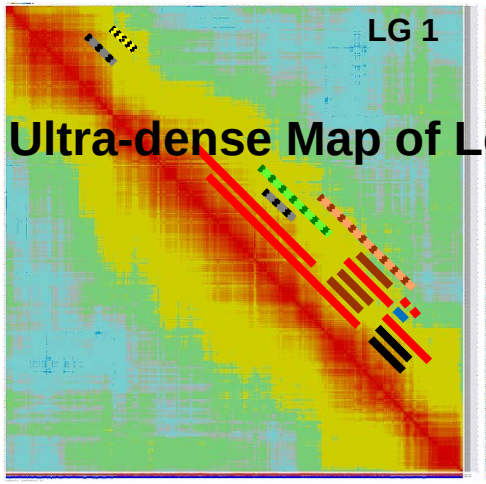
238 rice RILs each sequenced to 0.055x, 13x in aggregate.
Barcoded and multiplexed.
2x 36 nt paired-end reads,
20.6 Mb total single run.

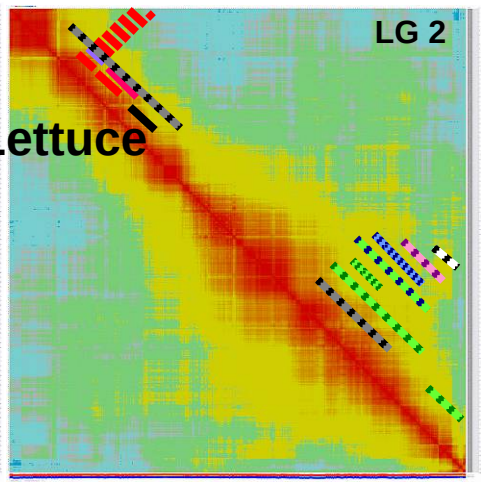
Genotypes inferred from RILs using maximum parsimony of recombination & HMM.

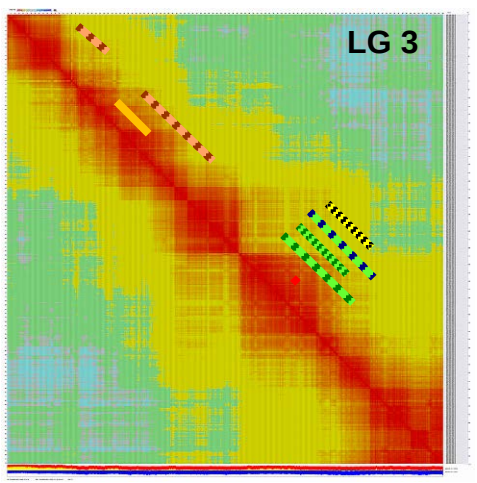
New capabilities => any species tractable in a single run.

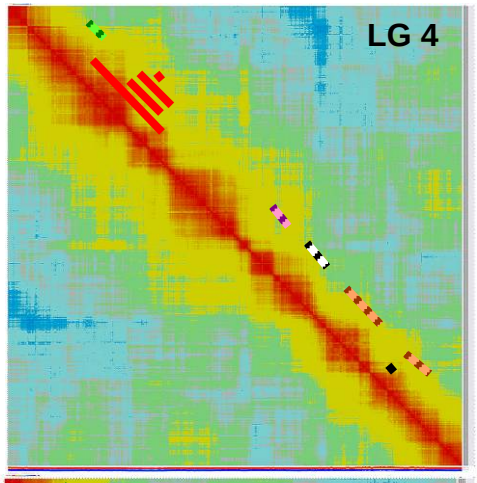


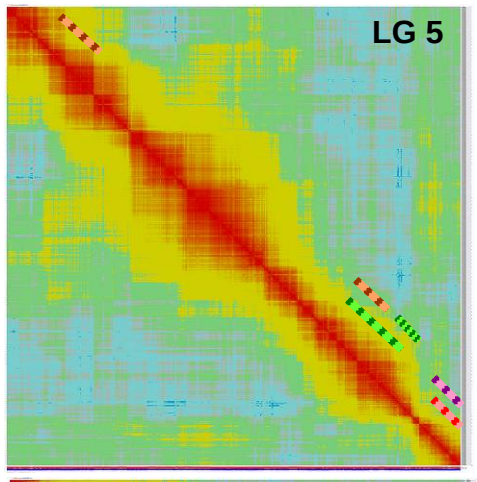
# Ultra-dense Map of Lettuce

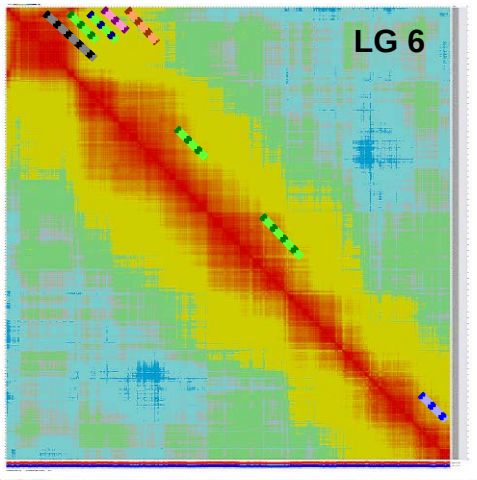


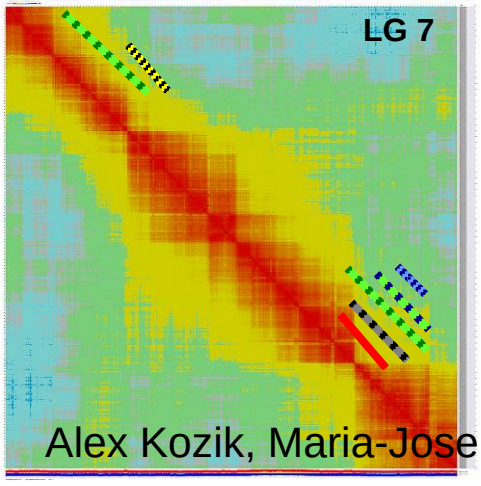



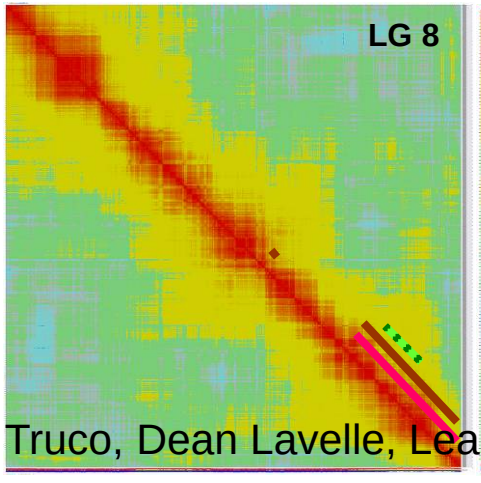



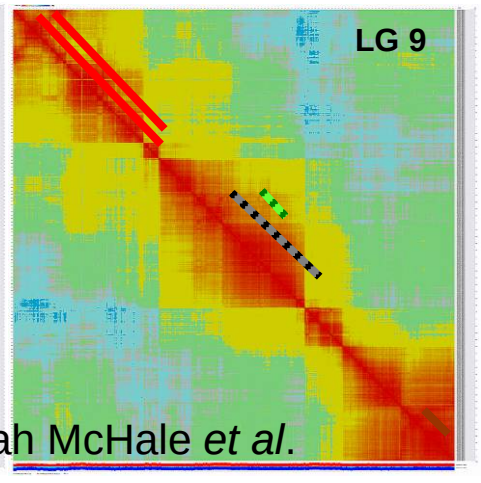



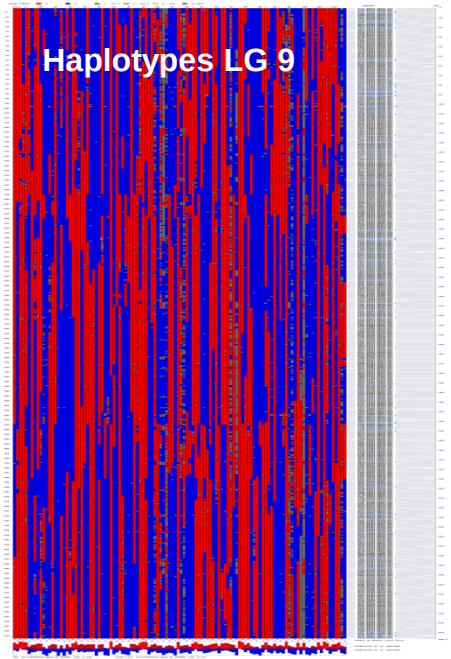




**Resistance phenotypes:**

- Downy mildew
- Virus
- Anthracnose
- Root aphid
- *Plasmopora lactucae-radicis*
- Corky root

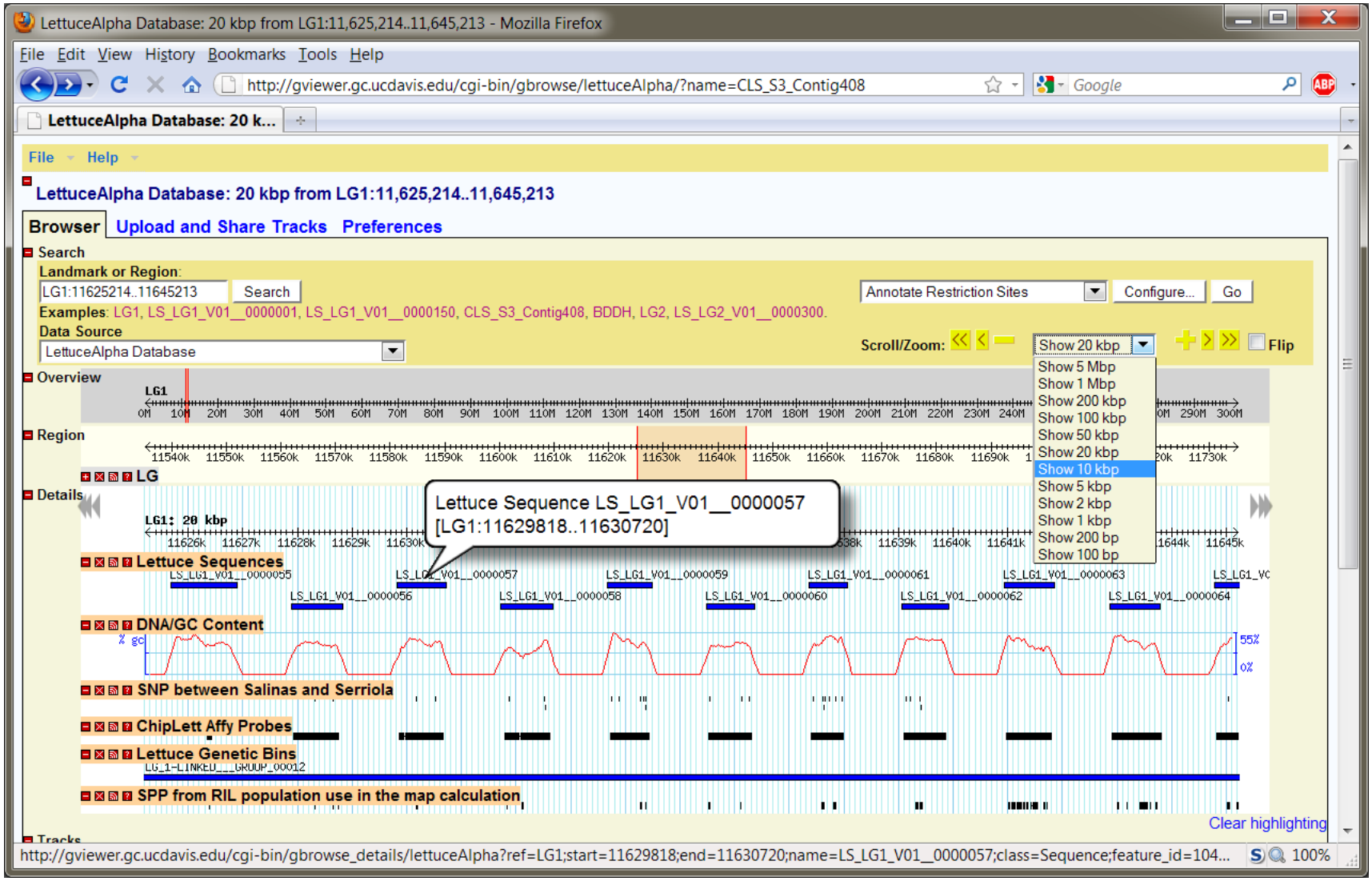
**Developmental Phenotypes**

- Bolting, Budding, Flowering
- Leaf Development
- Root Development
- Heading
- Rosette Formation
- Side Shoots
- Fasciation
- Branching
- Tipburn
- Spines
- Seed Shattering




Alex Kozik, Maria-Jose Truco, Dean Lavelle, Leah McHale *et al*.

# GBrowse Display of Lettuce PseudoSequence




**13,788 mapped ESTs separated by ‘N’s scaled to map distance for 9 chromosomes**

http://gviewer.gc.ucdavis.edu/cgi-bin/gbrowse/lettucePublic/ Sebastian Reyes Chin-Wo

# PARTICIPANTS IN THE LETTUCE GENOME SEQUENCING CONSORTIUM

BGI (CN) + 10 Companies:
Agrisemen (NL),
Enza Zaden (NL),
Gautier Semences (FR),
Isi Sementi (IL),
Monsanto Vegetable Seeds (USA)
Rijk Zwaan (NL),
Syngenta (USA),
Taki & Co (JP)
Tozers (UK),
Vilmorin (FR)

Agreements signed: August 1, 2010

First draft to participants: January, 2010. Annotated draft: March 2011.

Final report to participants: August 1, 2011.

Public data release: 12 months after release to participants.

# Lettuce Genome Sequencing

Assembled into 13,352 scaffolds > 2 kb, 552,061 > 100 bp, N50 = 460 kb.

2.5 Gb of total 2.7 Gb genome = 93% assembled.

Genomic sequence data integrated with transcriptome, gene-space,
to generate annotated draft genome sequence.

73% repeats, mostly retrotransposons and unknown repeats.

44,229 gene models, 79% genes functionally annotated.

Displayed using Gbrowse.

Assemblies validated using genetic data from ultra-dense map.
ordered into chromosomal linkage groups.
integrated with phenotypic and syntenic information.

Annotation and gene models refined at jamboree.



Zhiwen Wang, Song Chi, Xu Xun
& Sebastian Reyes Chin-Wo *et al*.

# Lettuce Genome Annotation Training and Jamboree

September 28th to
October 19th, 2011


**Participants**

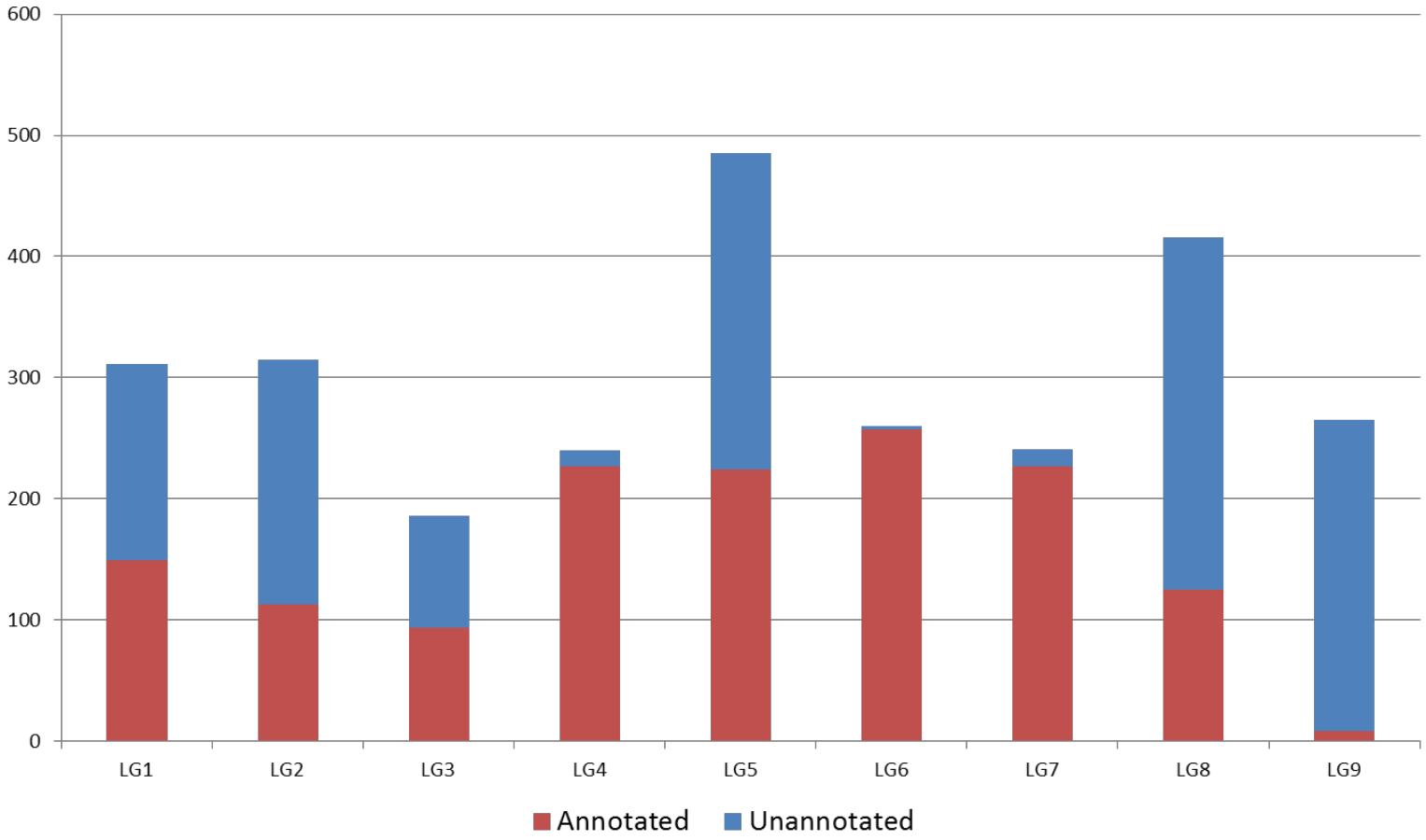
Breeding companies: 11
BG1: 2
Michelmore lab: 12
Non-UCD academics: 9
Non-UCD students: 7
Countries: 7
Compositae species: 10+

# JAMBOREE ANNOTATIONS

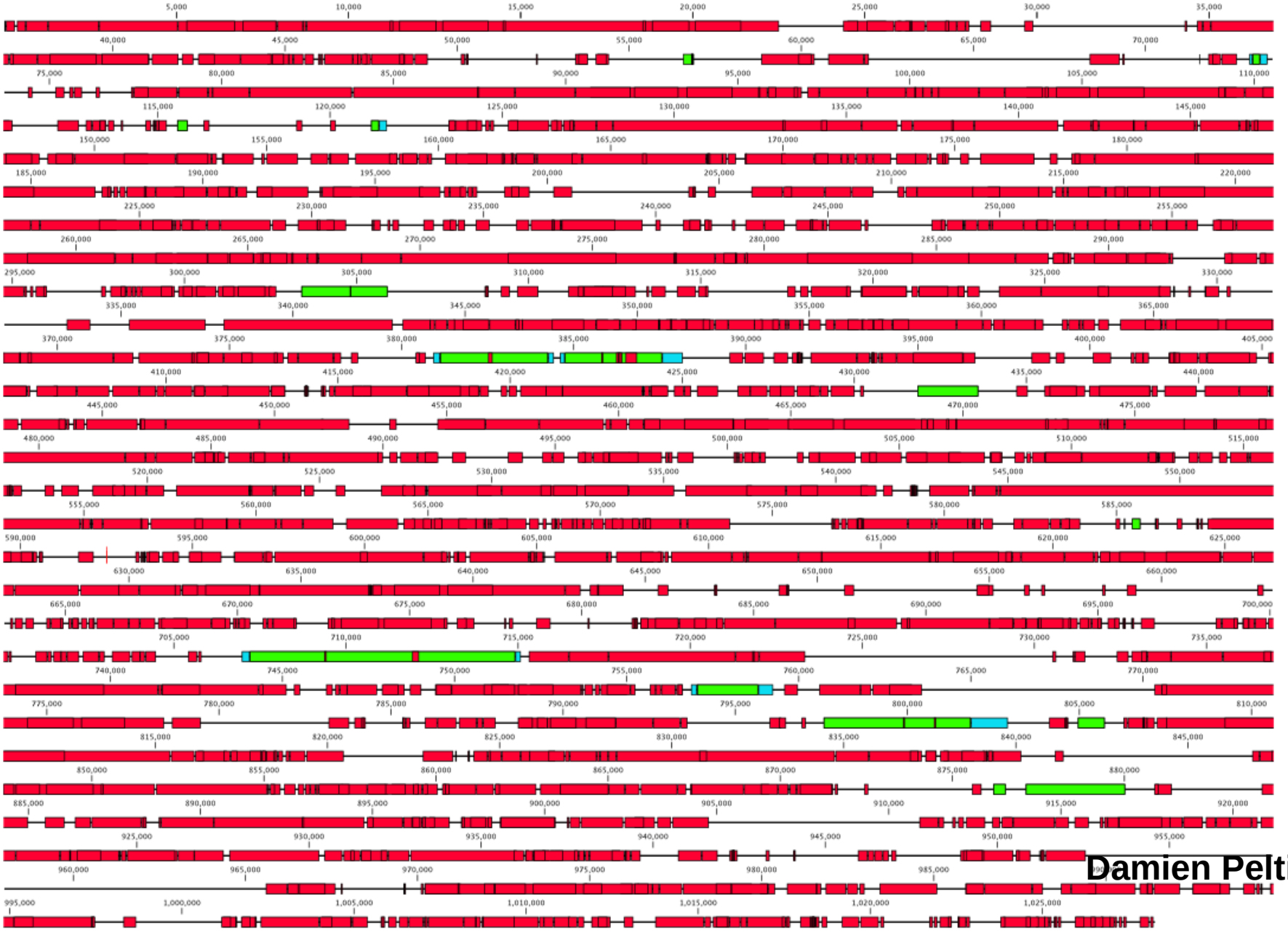
Eight teams of 3 to 4 people, each assigned a chromosome

8,333 gene annotations (~18% of 45,000 total)

~1,600 scaffolds (52% of mapped scaffolds, ~10% of total)



# Lettuce genes are located in a “sea” of repeated sequences



Damien Peltier

# Validation of Whole Genome Assemblies: Cosegregation of Unigenes in Scaffolds

11,279 mapped unigenes align with 3,201 scaffolds

Average 3 unigenes per scaffold (1 to 30)

3,023 (94.5%, ) all unigenes co-segregate = correct assembly

87 (2.7%) unigenes map to 2 co-segregating groups = simple chimeras

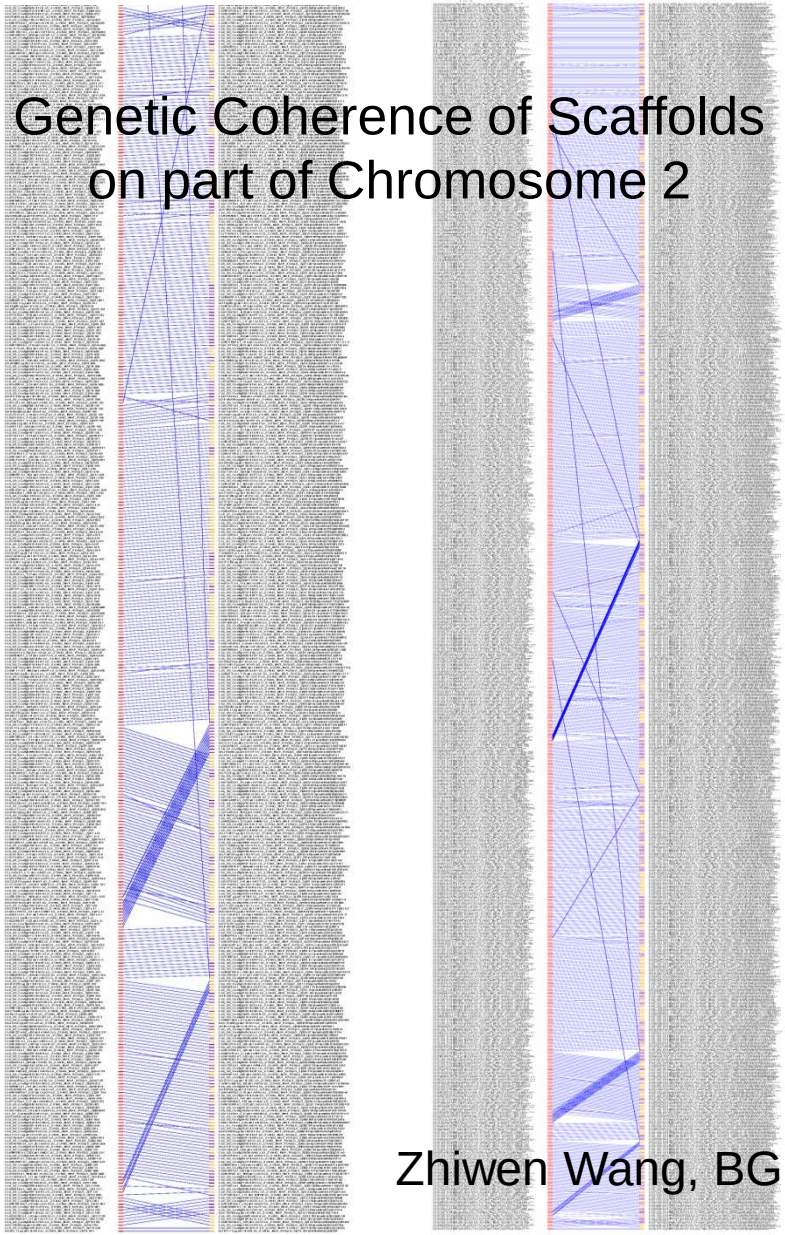
79 (2.5%) unigenes map to 1 cosegregating group + single anomaly

9 (0.3%) unigenes map to 3 or more linkage groups = mis-assembly

~60% assembly (~55% total) in chromosomal linkage groups

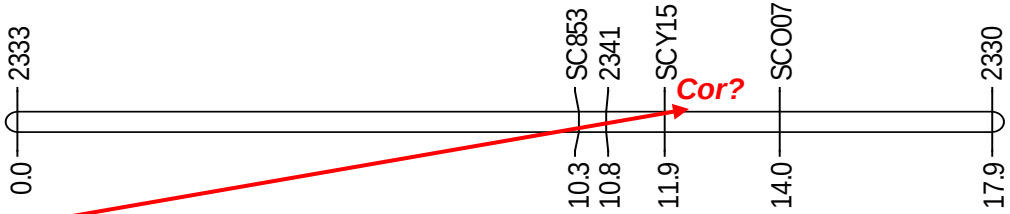
Sebastian Reyes Chin-Wo, UCD

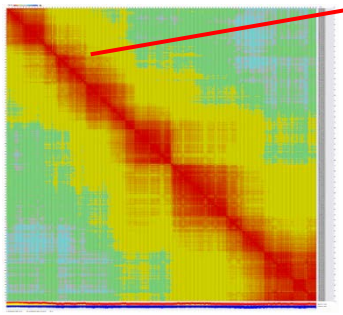
Genetic Coherence of Scaffolds on part of Chromosome 2



Zhiwen Wang, BGI

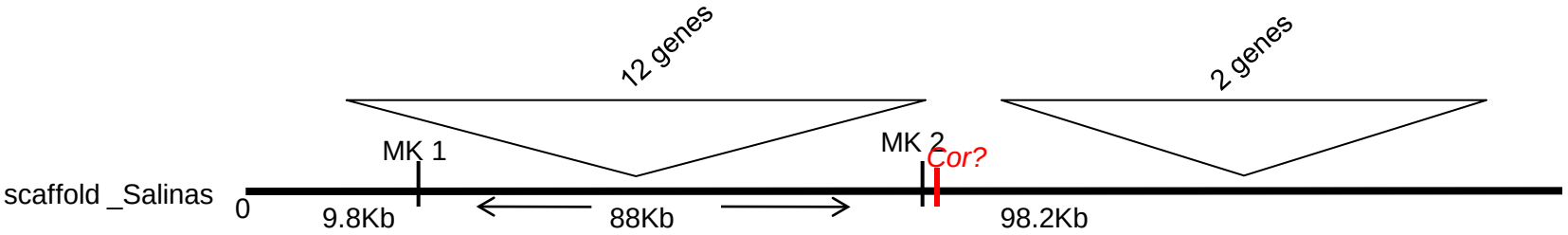
# HIGH-RESOLUTION GENETIC MAPPING OF *COR* REGION



- High resolution lettuce-chip map
- Numerous recombinants in region
- Lettuce transcriptome
- Lettuce gene space
- Lettuce genome sequence: ~650 kb

Identification of candidates in *cor* region




Disease screens of $F_5$ families

Green Lakes Diana

Future directions:

- Fine map relative to candidate gene(s) recombinants & BSAseq
- Assess allelic variation
- Functional analysis: RNAi
- Complementation analysis

*Manju Govindarajulu, Oswaldo Ochoa*

# COMPONENTS OF A GENERIC PLANT GENOMICS PROJECT

1. Generation of a high-quality, annotated reference genome sequence.
   - Single genotype, draft assembly and automated annotation.
   - Then manual annotation and finishing. Continual long-term process.
2. Development of ultra-high density genetic maps.
   - Genespace or WGS of ~100 descendants from one or more crosses.
   - Genotyping by sequencing. Chromosomal ordering of assemblies.
   - Candidate genes for traits.
3. Characterization of allelic variation.
   - Genespace or WGS of 500 to 1,000 diverse (randomly-mated) individuals.
   - Genome-wide association studies (GWAS).
   - Fine mapping of traits to candidate genes.
4. Development of a gene expression atlas.
   - RNAseq from multiple tissues, developmental stages and conditions.
   - Functional inferences on candidate genes.
5. Phenotyping and validation of marker associations and candidate genes.
   - Multi-locational testing of key genotypes and populations.
   - Mutant and transgenic analysis.
   - Dissection of complex traits. Systems biology.
6. Epigenetic analysis.
   - Requires technology development for routine global analysis of crop species.
7. Development and application of new technologies and analytical tools.
8. Database for curation, analysis and distribution.

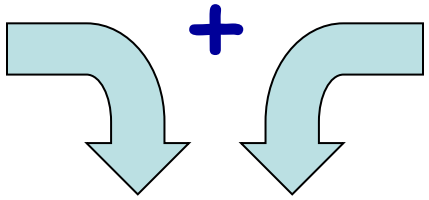
# IMPACT OF HIGH THROUGHPUT SEQUENCING & MASSIVELY PARALLEL GENOTYPING ON RESISTANCE STRATEGIES

- Global rather then gene-by-gene analysis
- Saturation of identification of candidate genes
  - Recognition, signal transduction, response
  - SNPs in causal genes
- Characterization of germplasm
  - Full genome resequencing of 10 – 100 genotypes
  - Gobal genotyping of 1,000 - 10,000 genotypes
  - Natural variation in 1°, 2°, & 3° genepools
  - Vast numbers of resistance genes available
- Characterization of pathogens
  - (A)virulence factors
  - Pathogen variability
- Gene deployment
  - Marker assisted selection of causal genes
  - Pyramids of multiple genes, conventional & transgenic
  - Heterogeneity between genotypes in space & time
  - Fragment selection pressure on pathogen populations
  - Manage pathogen evolution

# PATHOGEN POPULATION GENETICS SHOULD DRIVE DEPLOYMENT OF RESISTANCE GENES

Continual sampling of pathogen
Virulence phenotyping
Gene-space sequencing (10s)
SNP genotyping (1000s)

Resistance gene discovery pipeline
Germplasm screens
Mapping, molecular markers
Molecular characterization



Deployment of effective resistance genes
Pyramiding, MAS or effector-driven selection
Allo- and sympatric diversity
Temporal adjustment of resistance genes deployed
Transgenic approaches for novel resistance strategies

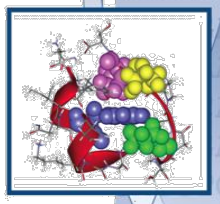
# Genome Center Courses

## Fall, 2011

UCDAVIS GENOME AND BIOMEDICAL SCIENCES FACILITY

GENOME CENTER

A global approach to biology



**Monday, August 29 – Friday, September 2**
Proteomics and Metabolomics (Brett Phinney and Oliver Fiehn)

**Tuesday, September 6 – Friday, September 9**
Sequencing Library Preparation. (Henny O'Geen and Ryan Kim)

**Monday, September 12 – Friday, September 16**
Next-Gen Sequence Analysis (Dawei Lin)

**Monday, September 19 – Tuesday, September 20**
Cloud Computing for Bioinformatics (Dawei Lin)

**Wednesday, September 21 – Friday, September 23**
PERL for Bioinformatics (Ian Korf)

**Monday, September 26 – Tuesday, September 27**
Data Analysis and Visualization Using R (Dawei Lin)

**Wednesday, September 28 – Friday, September 30**
Genome Annotation (Barry Moore and Ian Korf)

**Details available at [http://courses.genomecenter.ucdavis.edu](http://courses.genomecenter.ucdavis.edu)**

# GENOME CENTER CONTACTS FOR ASSISTANCE

Dr. Ryan Kim.
Director, DNA Technologies & Expression Cores.
rwkim@ucdavis.edu

Dr. Dawei Lin.
Director, Bioinformatics Core.
lhslin@ucdavis.edu

Dr. Brett Phinney.
Director, Proteomics Core.
bsphinney@ucdavis.edu

Professor Oliver Fiehn,
Faculty lead, Metabolomics Core.
ofiehn@ucdavis.edu

**http://genomecenter.ucdavis.edu**

## Current RWM Lab Members

María José Truco
Oswaldo Ochoa
Tadeusz Wroblewski
Lutz Froenicke
Katie Caldwell
Dean Lavelle
Joan Wong
Marilena Christopoulou
Keri Cavanaugh
Manjula Govindarajulu
Juliana Gil
Miguel Macias Gonzales
Pauline Sanders
Natalie Pelter

## Databases & Bioinformatics

Alex Kozik
Huaqin Xu
Bertrand Perroud
Sebastian Reyes Chin-Wo
Christopher Beitel
Damien Peltier
Davide Scaglione
Belinda Martineau

## Collaborators

BGI: Zhiwen Wang, Song Chi, Xun Xu,
Bicheng Yang, Sanwen Huang,
GCP: Loren Rieseberg, Steve Knapp,
Kent Bradford, Rick Kesseli,
John Burke, David Still, Zhao Lai
Allen van Deyzne & the LetChip team
Ian Korf
Ryan Hayes *et al.* (USDA Salinas)
Jean Greenberg & lab (U. Chicago)
Managers & staff of the GC Cores
Many others for specific materials
& populations

## Funding

NSF Plant Genome & 2010
UDSA NRI, AFRI, SCRI
Seed Company Gifts
UC Biostar Program & Bioseeds Co.s
Novozymes Inc. Endowed Chair
Ca. Leafy Greens Research Board
The Lettuce Genome Sequencing
Consortium

http://michelmorelab.ucdavis.edu