The iPlant Collaborative

Matt Vaughn
Texas Advanced Computing Center

iPlant Collaborative™

NSF's PSCIC Program

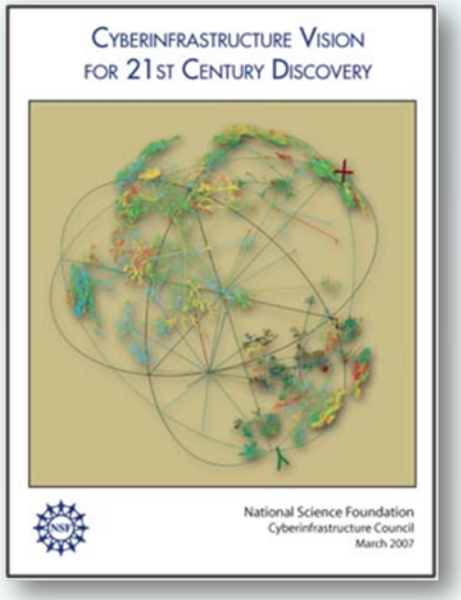
PSCIC Goals:

- ***"to create a new type of organization*** *- a cyberinfrastructure collaborative for plant science"*
- *"to enable new conceptual advances through integrative, computational thinking"*
- *"to address an evolving array of* ***grand challenge questions*** *in plant science: the driving force and organizing principles for the collaborative"*

The iPlant Collaborative

Cyberinfrastructure for the Plant Sciences

- $50M NSF Funded Project – in 5th year
- Recommended for second 5 year term
- iPlant is a cyberinfrastructure platform
- The platform is extensible by users
- NSF recommended scope beyond plants
- iPlant supports plant & animal breeding
- iPlant will bridge the genomics – breeding gap



Brief History of iPlant

- Funded by NSF – February 1st, 2008
- Kickoff Conference at CSHL – April 2008
- Grand Challenge & CI Workshops – Sept 2008 – Jan 2009
- GC White Paper Review & Recommendations – Mar 2009
- Project Kickoffs – May 2009 & August 2009
- Requirements gathering – ongoing since May 2009
- Start of software development – Sep 2009
- First prototypes delivered to public – Apr 2010
- DE release with user-driven tool integration – Jul 2011
- "iAnimal" portal clone & announcement – Jan 2012

iPlant Collaborative™

What is iPlant?

- **iPlant's mission is to build the CI to support plant biology's Grand Challenge solutions**
- **Phase I – Community Input**
- **Phase II – Building the CI Foundation**
- **Phase III – Enabling Discovery & Applications**

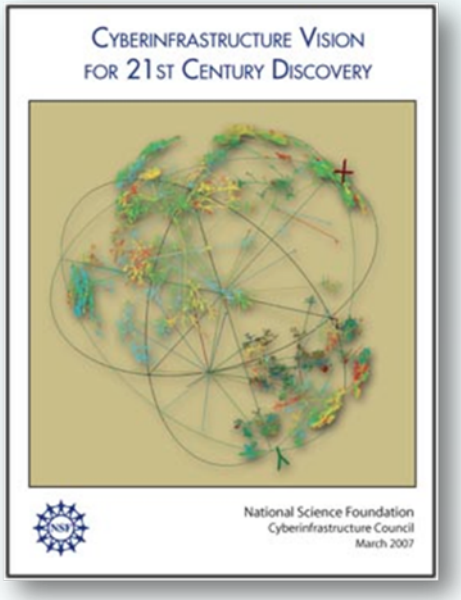
Need to:

- **Integrate more tools and workflows**
- **Add more collaboration support**
- **Promote testing theories & discover**
- **Support downstream applications**

iPlant Collaborative™

NSF Cyberinfrastructure Vision

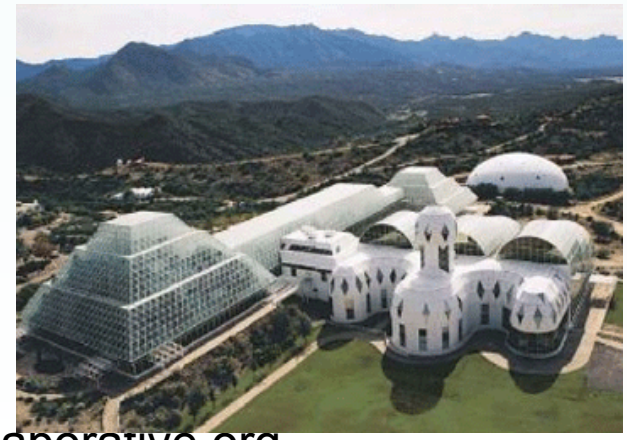
- **High Performance Computing**
- **Data and Data Analysis**
- **Virtual Organizations**
- **Learning and Workforce**



Ref: “Cyberinfrastructure Vision for 21st Century Discovery”, NSF Cyberinfrastructure Council, March 2007.

Grand Challenge & CI Workshops

- Mechanistic Basis of Plant Adaptation (Sep 2008)
- Impact of Climate Change on Plant Productivity: Prediction of Phenotype from Genotype (Sep 2008)
- Developing common models for molecular mechanisms, crop physiology, and ecology (Nov 2008)
- Assembling the Tree of Life to Enable the Plant Sciences (Nov 2008)
- Computational Morphodynamics of Plants (Dec 2008)
- Botanical Information & Ecology Network (Dec 2008)
- CI Workshop (Jan 2009)



Grand Challenge Projects + Added Efforts

- Plant Tree of Life – iPToL – May ’09
  - + Taxonomic Intelligence (TNRS)
  - + Scientific Networking Website (MyPlant)
  - + Perpetually Updated Trees
  - + Species Distribution Maps (Seed Project)
- Genotype to Phenotype – iPG2P – Aug ’09
  - + Image Analysis Platform (PhytoBisque)
  - + GLM by GPU and FPGA
  - + Integrated Breeding Platform (GCP/Gates)
  - + Comparative Genomics Platform (CoGe)
  - + Semantic Web Development

iPlant Collaborative™

iPToL Working Groups

- Building Big Trees
- Analyzing Trait Evolution
- Gene-Species Tree Reconciliation
- Visualization of Big Trees and Related Info
- Data Integration in Phylogenetics
- Data Assembly
- Community / Network Building

iPG2P Working Groups

- Ultra High Throughput Sequencing
- Genome Assembly
- Statistical Inference
- Visual Analytics
- Data Integration
- High-throughput Image Analysis
- Modeling Tools

Can Plant Research Use CI Now?

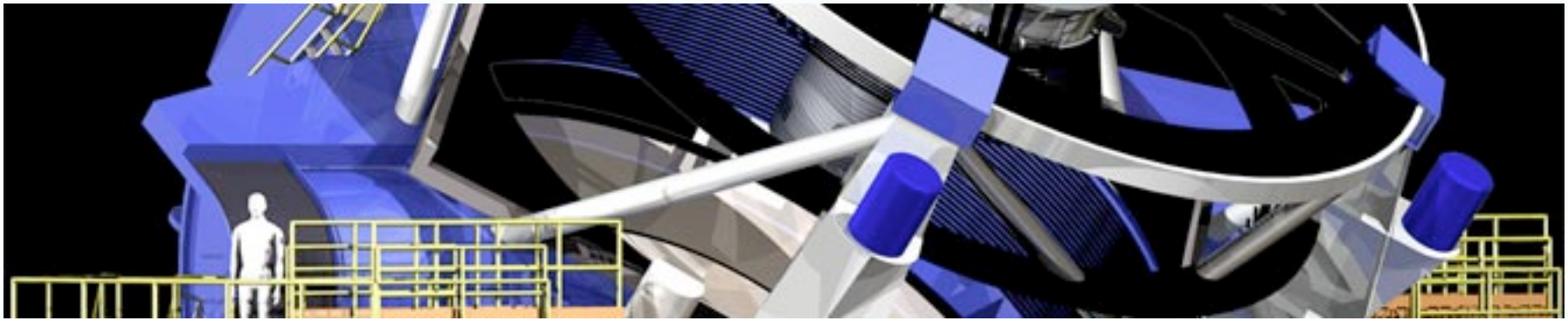
Candidate problems with existing data:

- Hybrid vigor, inbreeding depression
- Water & nutrient use efficiency
- Identification of yield components
- Interaction of plants & microbes
- Ecosystem dynamics
- Protein structure & interactions

Most NSF CI Supports Physical Science

Example: The Large Synoptic Survey Telescope (LSST)

- 8.4 meter mirror
- 3,200 megapixel digital camera
- Surveys entire sky every few days
- Will generate >30 terabytes of images per day
- Must store and share large volumes of data



Biology's "Big Data" Instruments

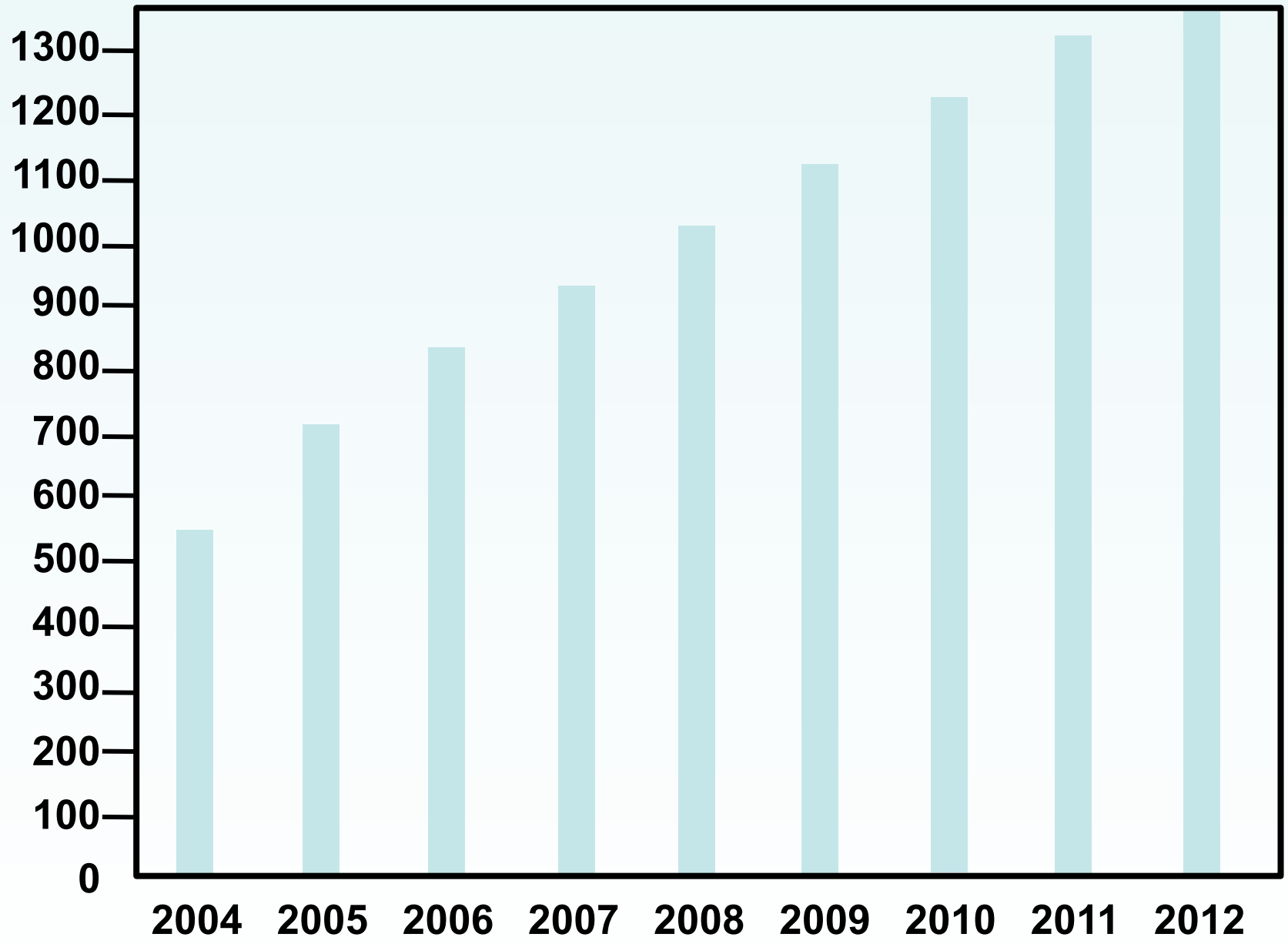
Ultra-High-Throughput Sequencers

Example: Illumina HiSeq 2000

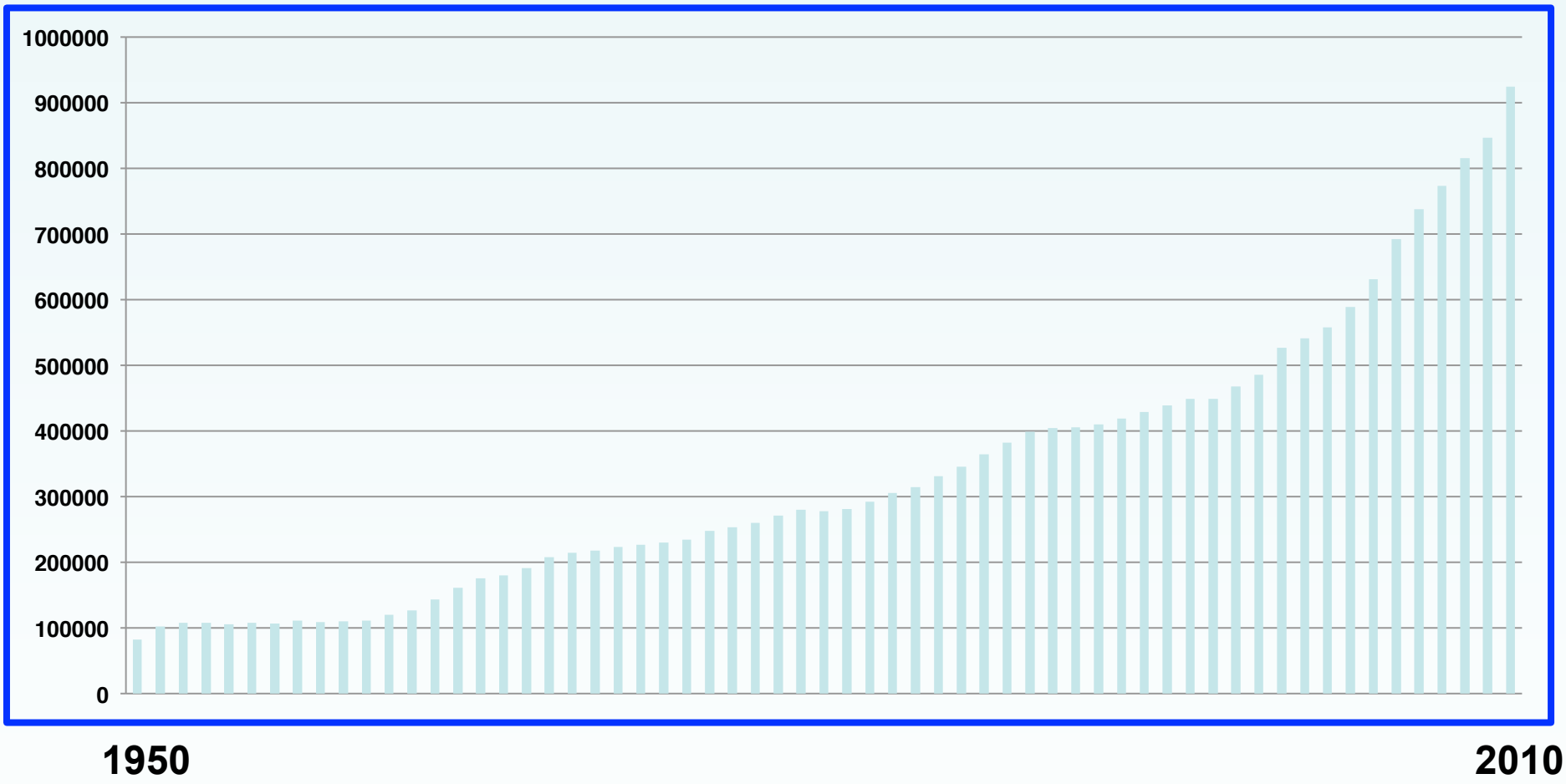
- >1 terabyte sequence data / 11 days
- >7 x 10^9 nucleotides / run
- Will create >1k analysis jobs/day
- Analysis –the new bottleneck

.....
AGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACG
CCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGC
CTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGT
TCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCA
AATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAAT
GCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGA
CGCCTGTTCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGG
CCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTG
TATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTGCAAATGACGCCTGTATCAATGCTAGGCCTTG
CAAATGACGCCTGTTCAATGCT .....

NAR Databases & Tools Over Time



PubMed Publications Over Time



Accounts for ~70% - Currently >2,500/day

The iPlant Collaborative

A virtual organization

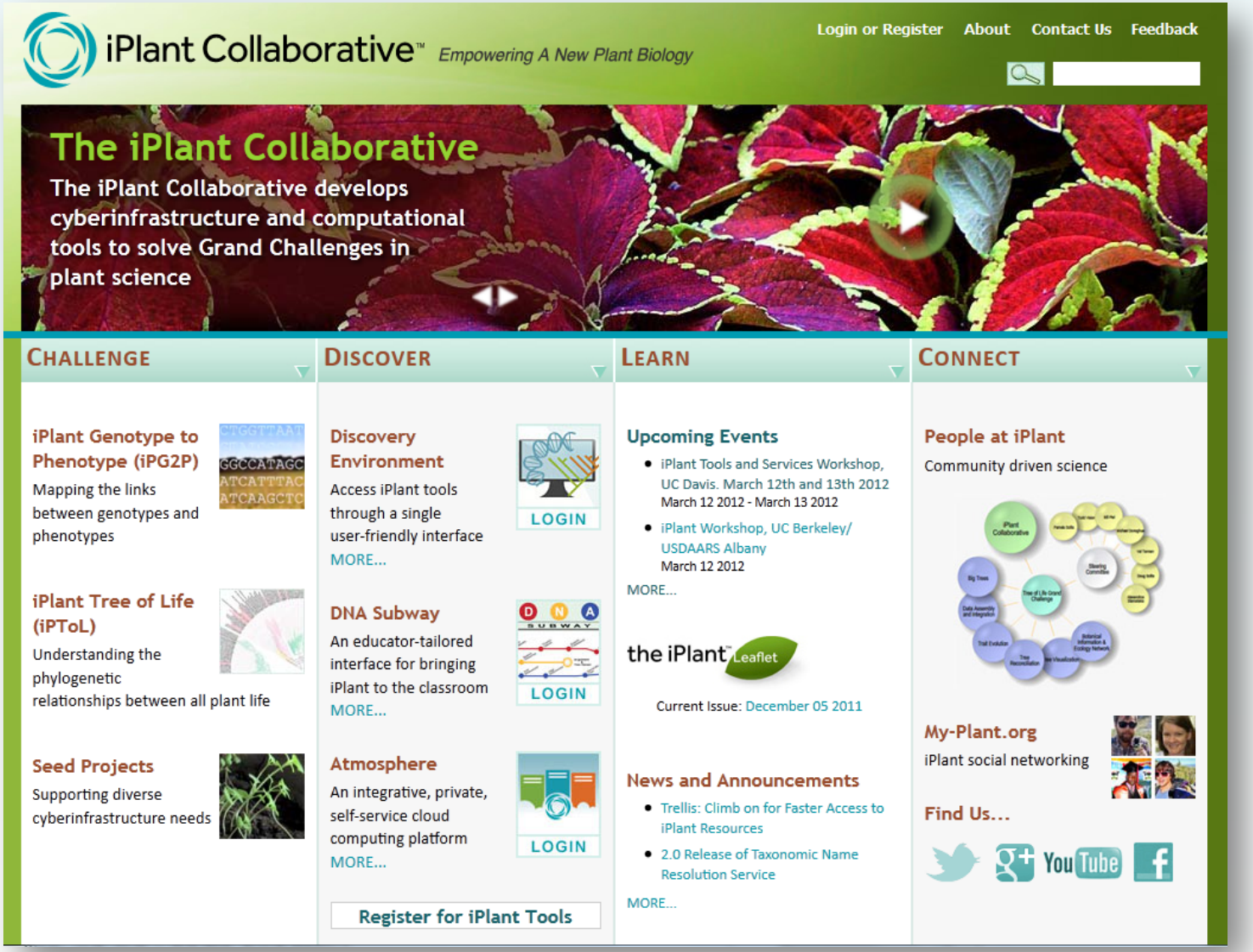
www.iplantcollaborative.org

What iPlant has to offer:

- Data Management Resources
- High-Performance Computing Resources
- Tool Integration System
- Application Programming Interfaces
- Cloud Computing Resources
- Genotype To Phenotype Science Enablement Portfolio
- Tree of Life Science Enablement Portfolio
- Image Analysis Platform
- Molecular Breeding Platform (with IBP)

The iPlant Collaborative

Web site – entry point to tools & documentation

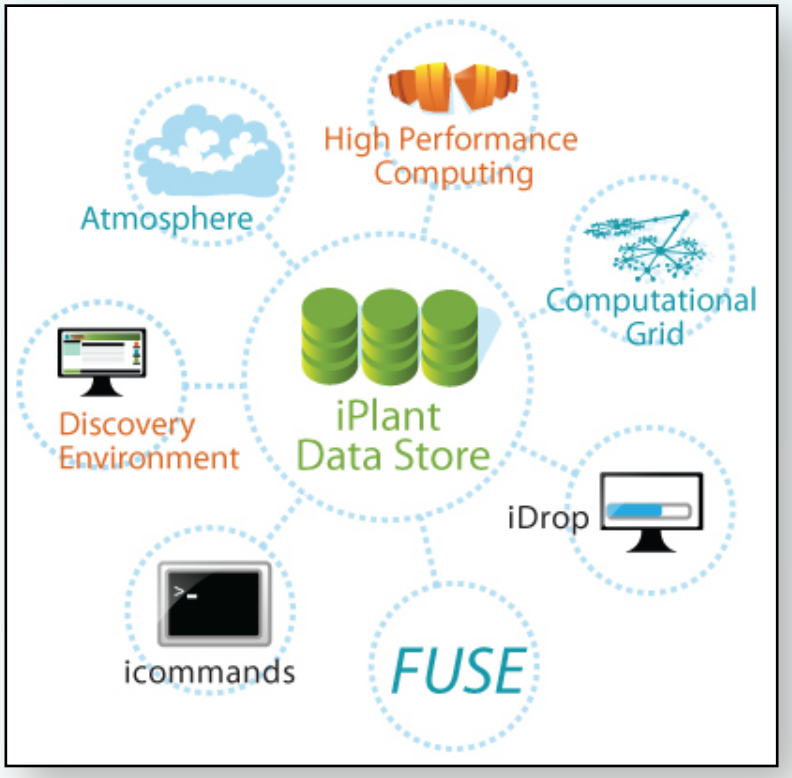


www.iplantcollaborative.org

Customized cloud platform for computing on your terms !

The iPlant Data Store

"Cloud Storage"… but it's not Amazon



Fast data transfers via parallel, non-TCP file transfer

- Move large (>2 GB) files with ease

Multiple, consistent access modes

- iPlant API
- iPlant web apps
- Desktop mount (FUSE/DAV)
- Java applet (iDrop)
- Command line

Fine-grained ACL permissions

- Sharing made simple

Access and a storage allocation is automatic with your iPlant account

www.iplantcollaborative.org

The iPlant Data Store

- Access to >100,000 Terabytes of disk & tape (100 Petabytes)
- Fast transfer
- Storage near HPC

iPlant Data Store Performance

Data Transfer Rate from UC Berkeley to UA (iPlant DS)

- **Dec 5th, 2011:**
- **100GB: <30 min**

iPlant Access to HPC via XSEDE

Scalable Computation for High Throughput Analysis

- Leveraging XSEDE
- TACC, SDSC, PSC, EBI
- >500,000 Compute Cores
- Up to 4TB shared memory machines

TACC Lonestar

TACC Stampede

PSC Blacklight

TACC Corral

SDSC CI

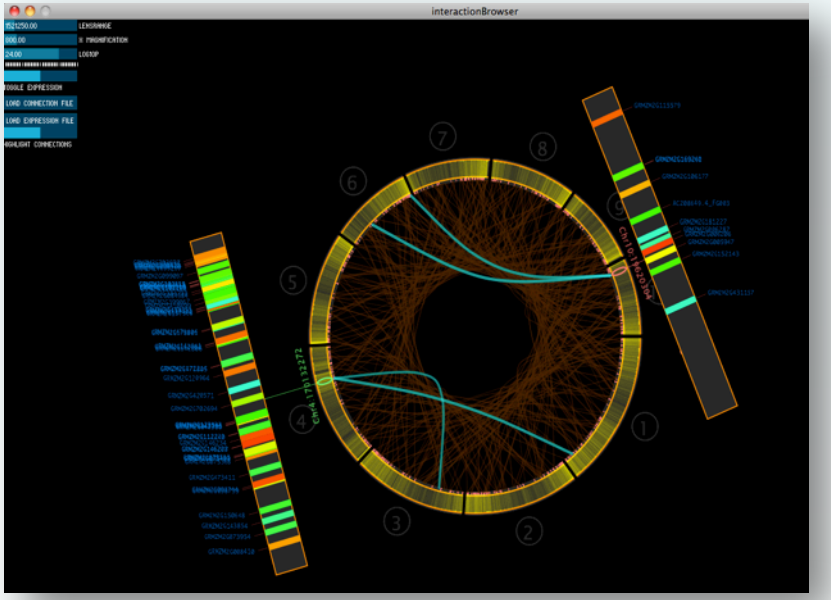
EBI Web Services

Accelerating Analysis – an Example

- **Code Parallelization efforts yields major results**
  - **Genome Wide Association Study accelerated from 1,600 years to 4 hours**

Problem:

  - **Two months of communication required**
  - **Few weeks of development**



iPlant Cyberinfrastructure Strengths

- Extensible, flexible platform architecture
- Not limited to plant science (iAnimal, iArthropod)
- Diverse community collaborations
- Experienced staff working in a distributed fashion
- Unified access to iPlant (single sign-on)
- Genotype to Phenotype & Phylogenetics tools
- Various levels of support, novice to expert user
- Developing semantic web effort

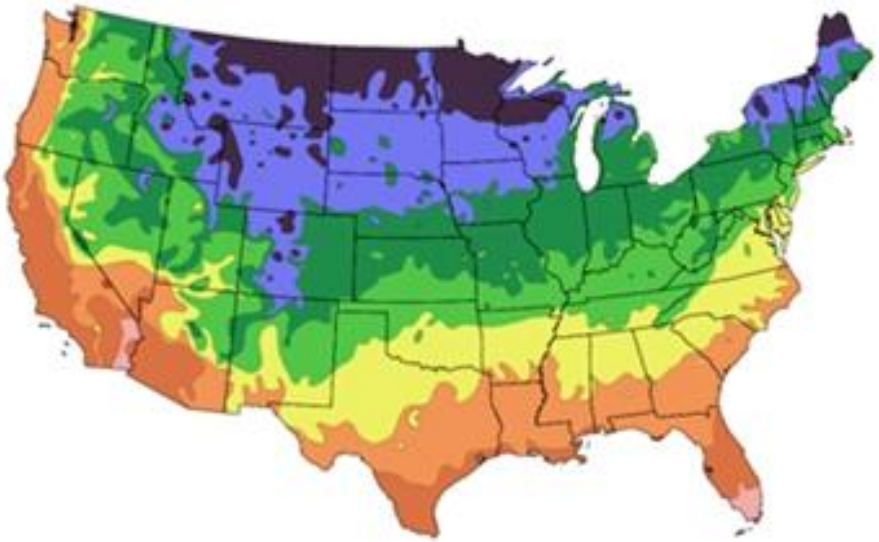
iPlant Collaborative™

The Future of iPlant

iPlant Collaborative™

The Fundamental Problem…

1990 Map



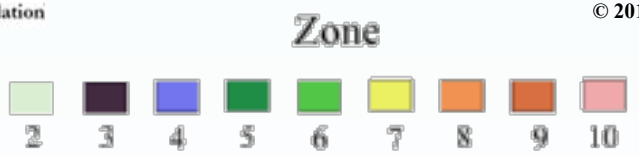
After USDA Plant Hardiness Zone Map, USDA Miscellaneous Publication No. 1475, Issued January 1990

© 2006 by The National Arbor Day Foundation

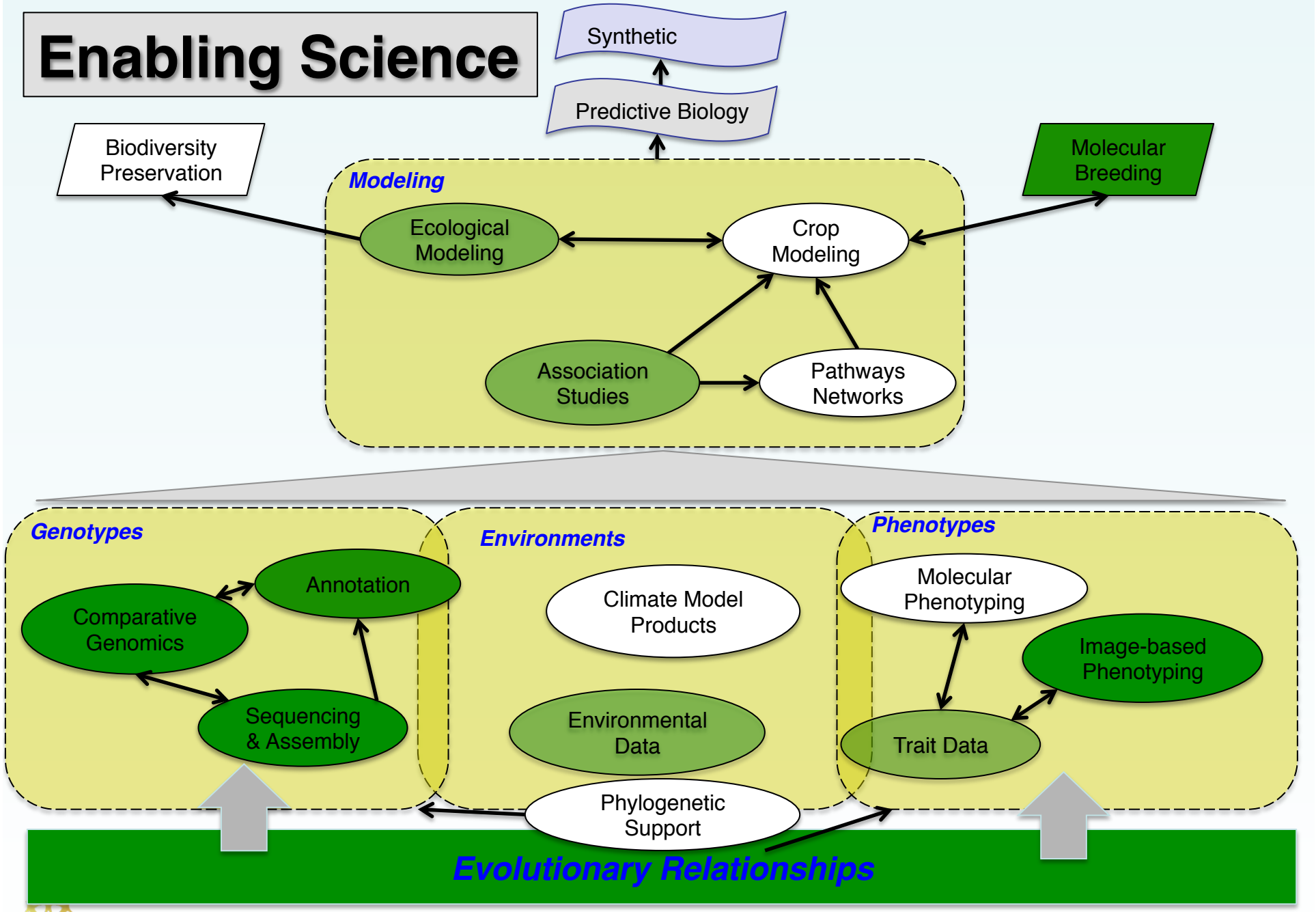
2012 Map

Re-colored version of the 2012 USDA Plant Hardiness Zone Map (available at: http://planthardiness.ars.usda.gov/PHZMWeb/)

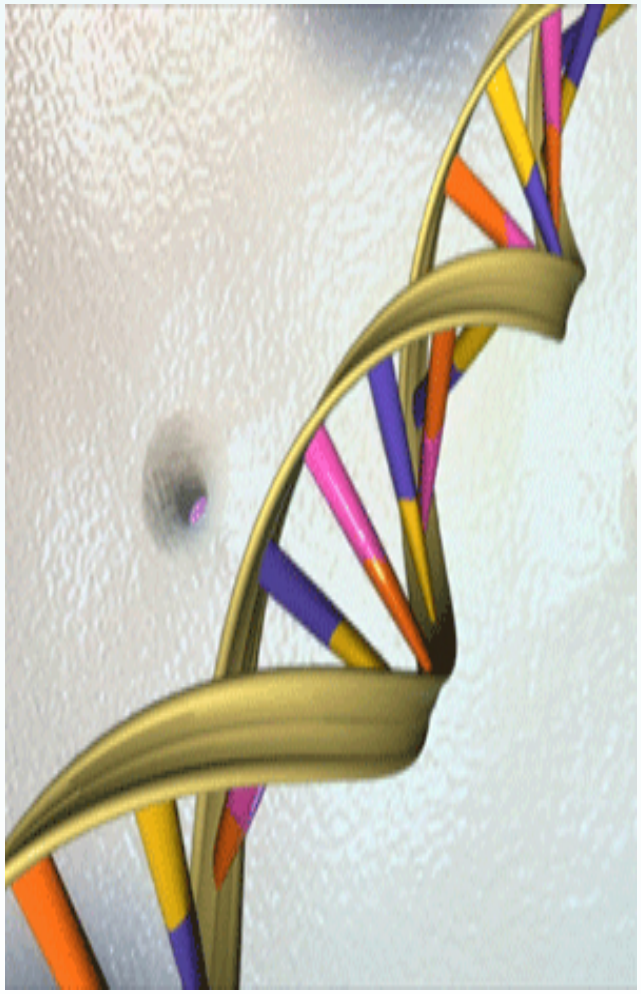
© 2012 National Wildlife Federation



Enabling Science

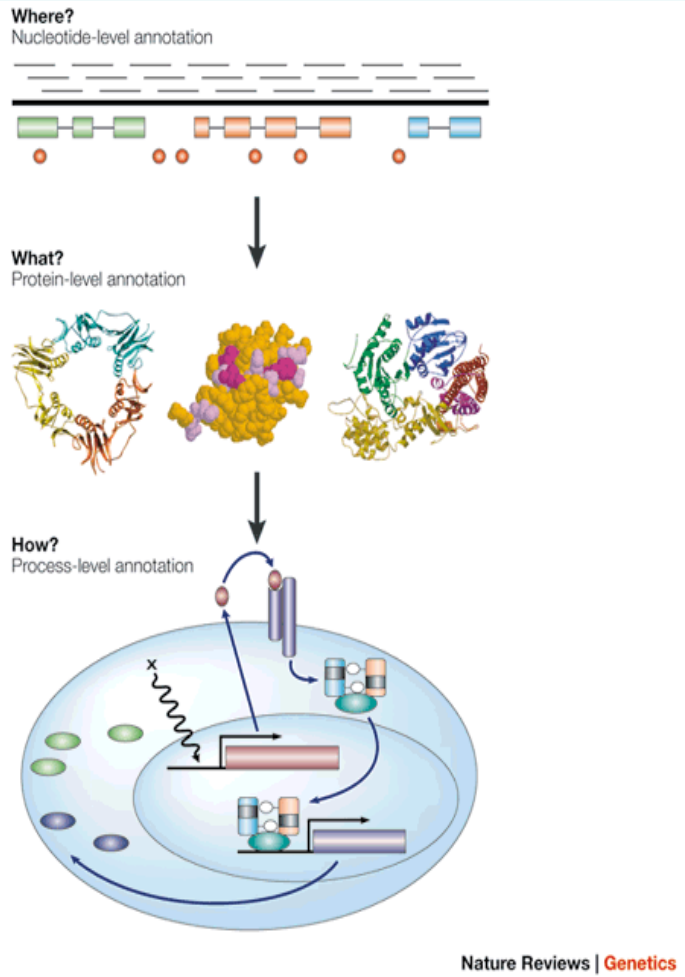


Sequencing and Assembly



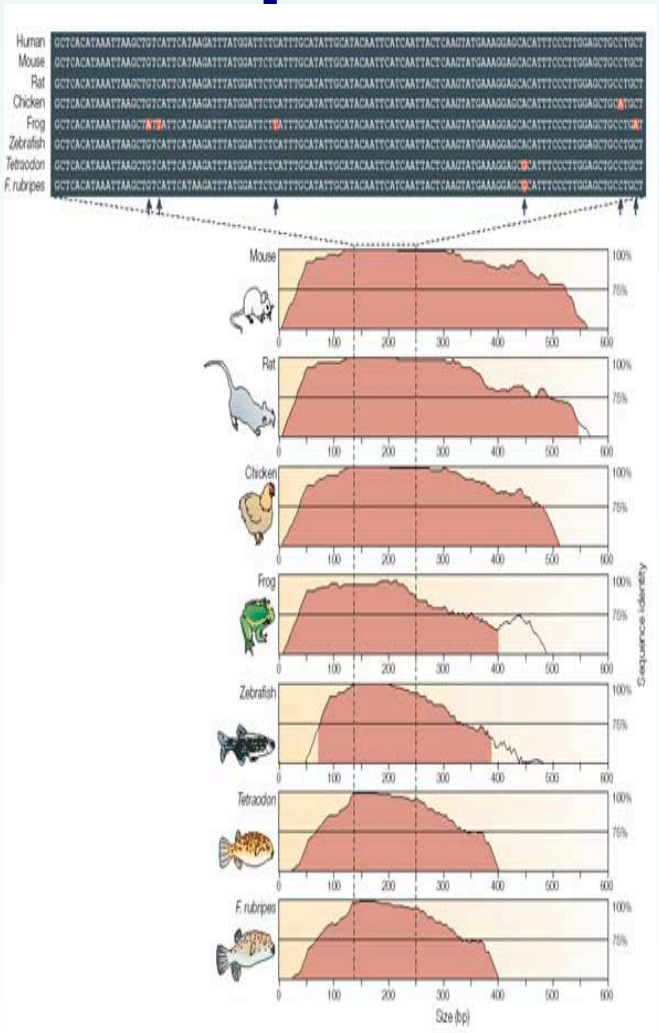
- Support for longer reads and larger datasets
- Incorporation of high priority algorithms
- Scaling of algorithms & workflows on XSEDE
- Algorithm development for analysis of assembly correctness and completeness
- Genome complexity and library construction modeling
- Tutorials and documentation updates
- Publication of plant genome-specific Assemblathon
- Availability of data via iPlant Data Commons
- Data exchange/federation with existing platforms

Genome Annotation



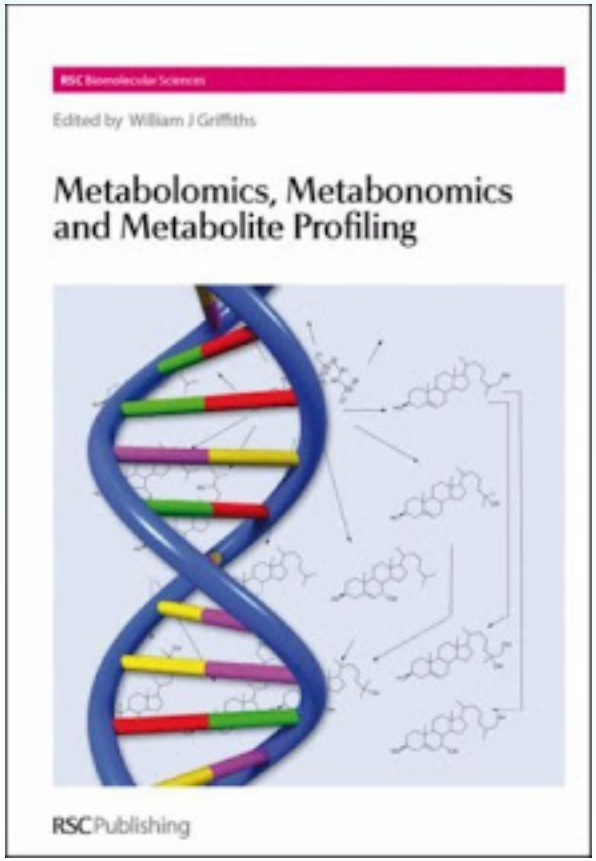
- Full scale annotation pipeline - protein genes
- Pipelines for functional annotation of protein coding genes
- Pipelines for annotation of non-coding genes, regulatory elements, consequence of variation and potentially chromatin state
- Visualization of annotation
- Information resources to elevate field of plant genome annotation
- Support materials for outreach and training.

Comparative Genomics



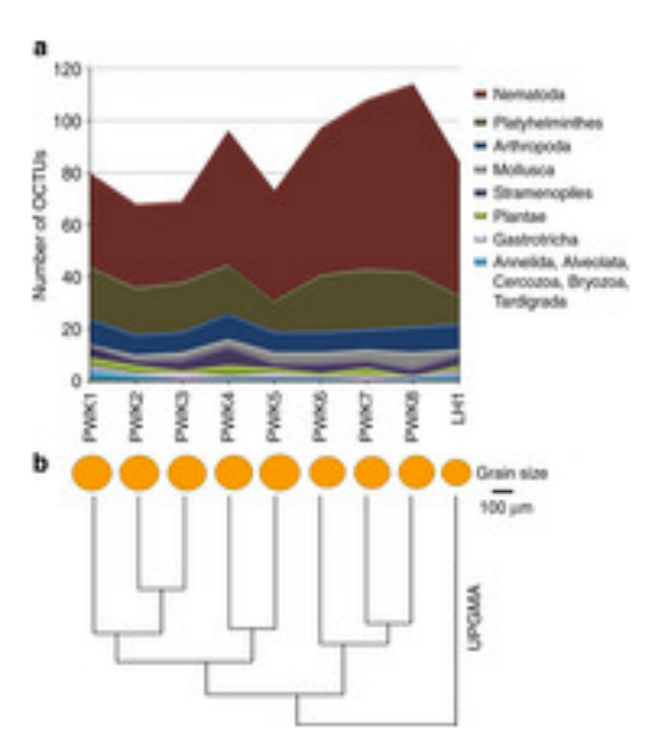
- Toolkit available with at least one tool for local and whole genome alignment and synteny analysis
- Tutorials and documentation
- Interoperability with CoGe platform
- Interoperability with other service providers
- Distributed Annotation System (DAS) proserver and associated tools
- Homology resolution service

Molecular Phenotyping



- Ability to deposit, store, disseminate and share metabolomic data (including NMR and MS spectra)
- Toolkit to process metabolomic data so they can be used with other types of biological data
- Documentation and tutorials
- Publication that highlights best practices

Trait Data



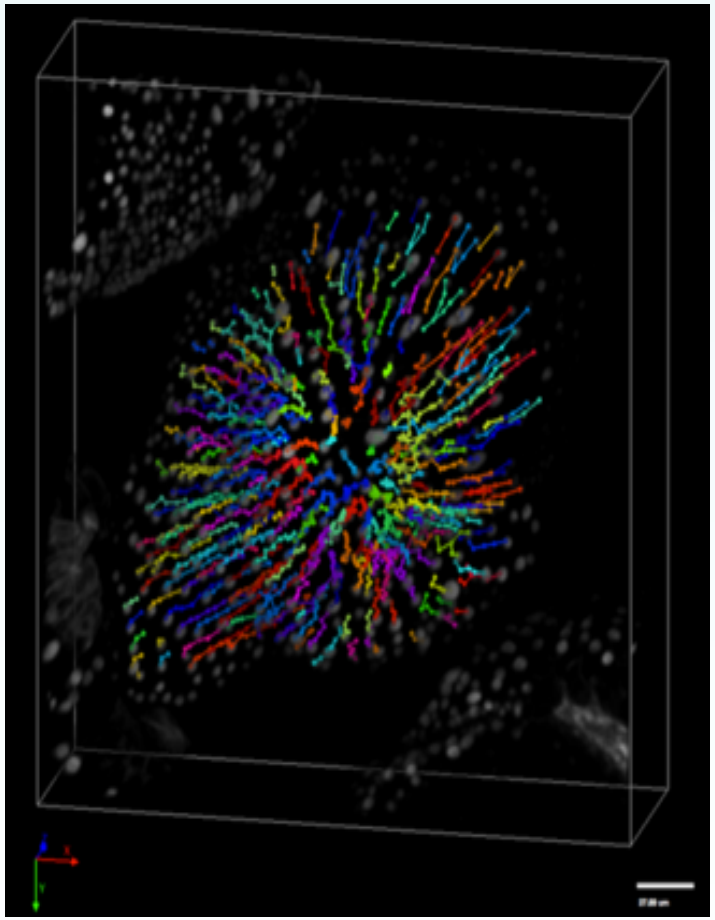
Ecological Trait Data

- BIEN database available via iPlant CI
- Semantically described web services from Dendrome
- Collaboration with agribusiness for data acquisition
- Data discovery, ingestion and federation
- Text mining workflows

Molecular Trait Data

- Data storage based on standards
- Integration with existing repositories
- Toolkit for analyses and quality control
- Documents and tutorials
- Publications

Image Based Phenotyping

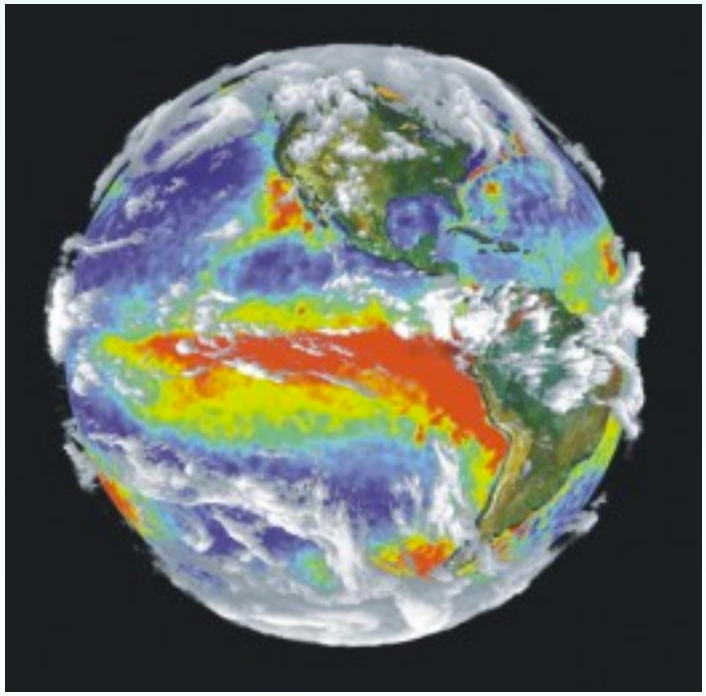


- Robust application integration toolkit
- Scalability
- Algorithm integration
- Training materials
- Integration with iPlant data store (metadata integration)
- GUI and usability improvements
- Data availability in iPlant Data commons
- Data exchange, ingestion and federation with existing platforms
- Atlas style reconstruction

Environmental Data

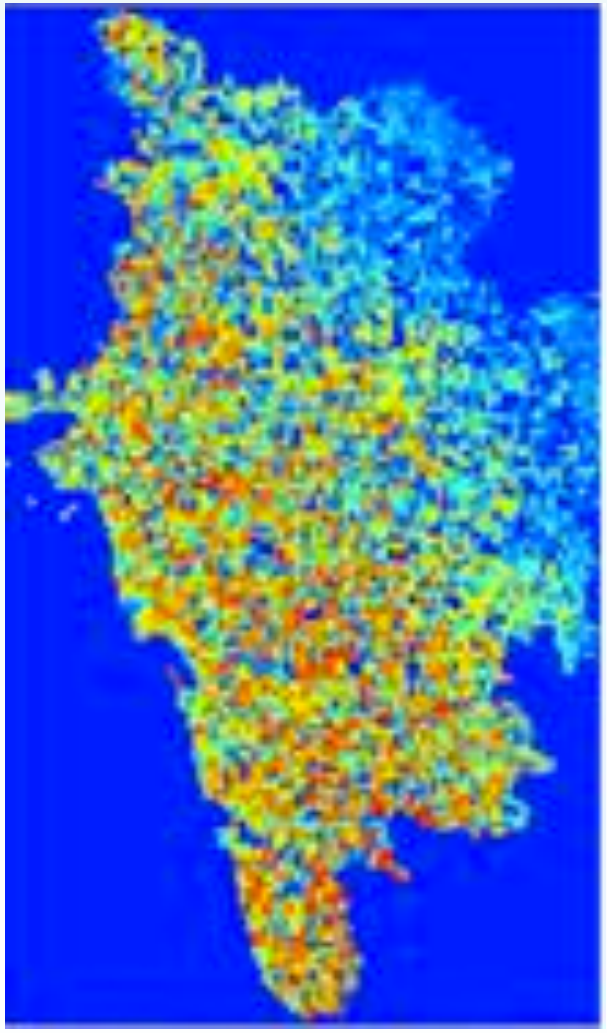
- CI to support visualization, exploration, sharing, sub setting and downloading of data
- New global environmental layers designed for biological research and readily usable by scientists
- Semantically described global environmental layers
- NEON data replication
- NEON data integration /exchange

Climate Modeling Products



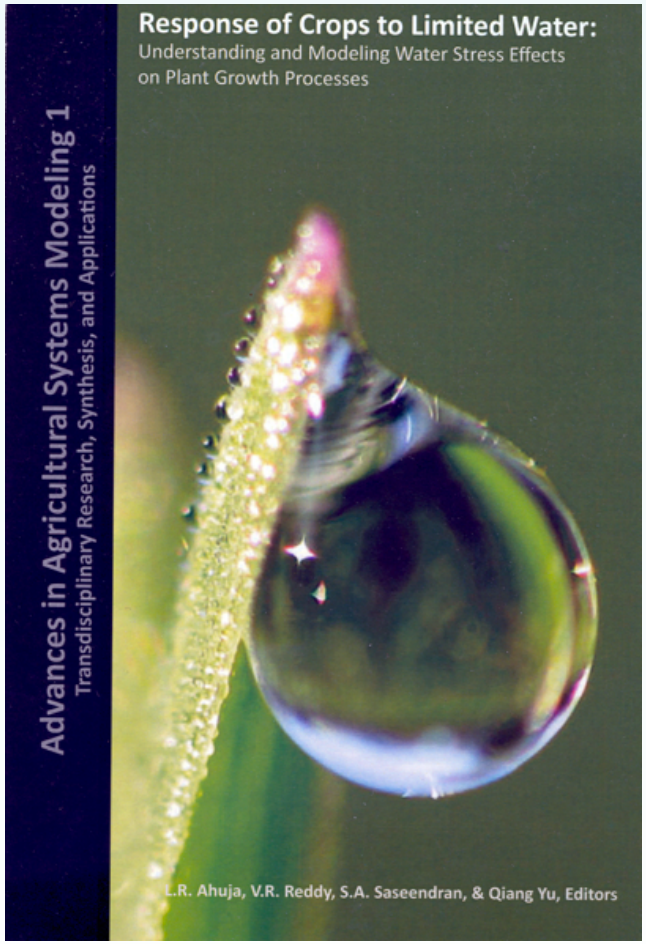
- Tools and workflows that allow climate model products as inputs to biological models
- Graphical tools for data selection
- Semantically described climate model product availability
- Training materials and workshops
- Publications

Ecological Modeling



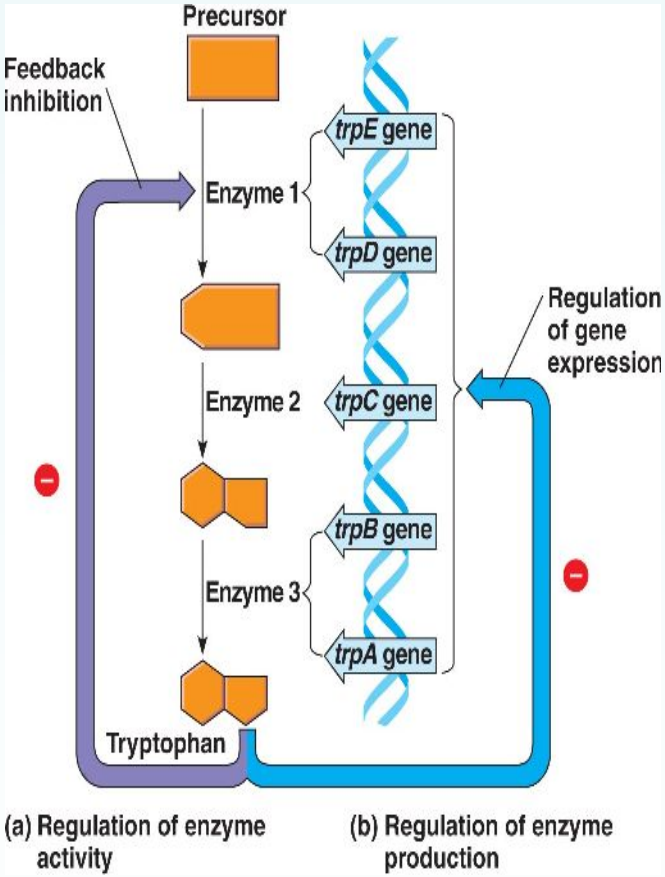
- Scalable, interoperable models
- Tools for visualizing and comparing outputs
- Models depicting plant species distribution in the Americas under various climate conditions
- Range maps based on mechanistic niche models that include traits and biotic interactions
- Models that can predict physiological state of plant given geographic location under current or future climate scenarios
- Ability to perform broad comparisons of the ecophysiological traits of crops vs. wild types
- Documentation and EOT support materials
- Publications

Crop Modeling



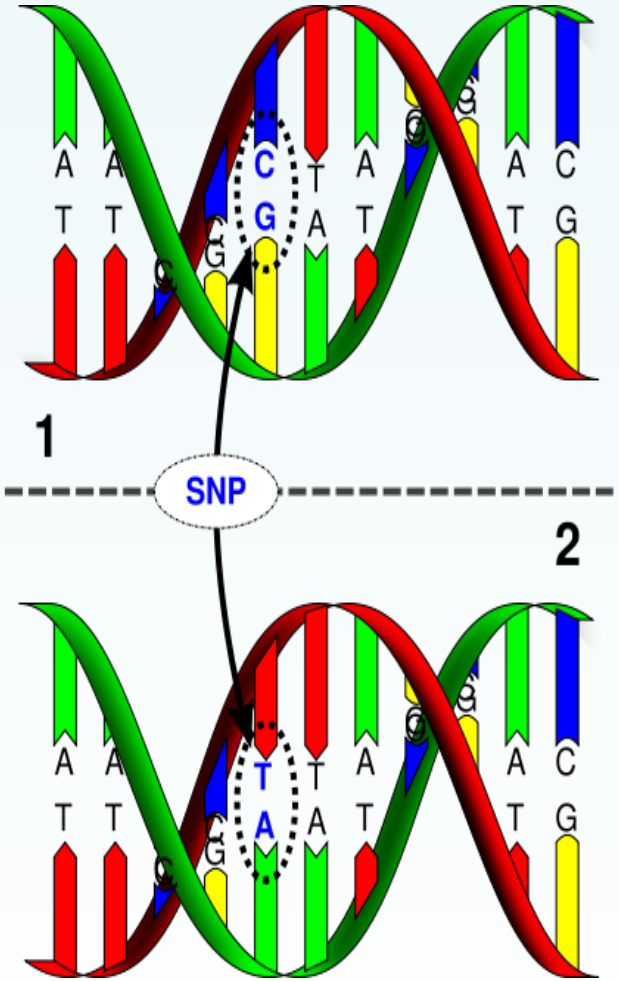
- Requirements analysis
- Community engagement (AgMIP)
- Data accessibility via Data Commons
- Support materials for EOT
- White papers

Metabolic/Regulatory Pathway Modeling



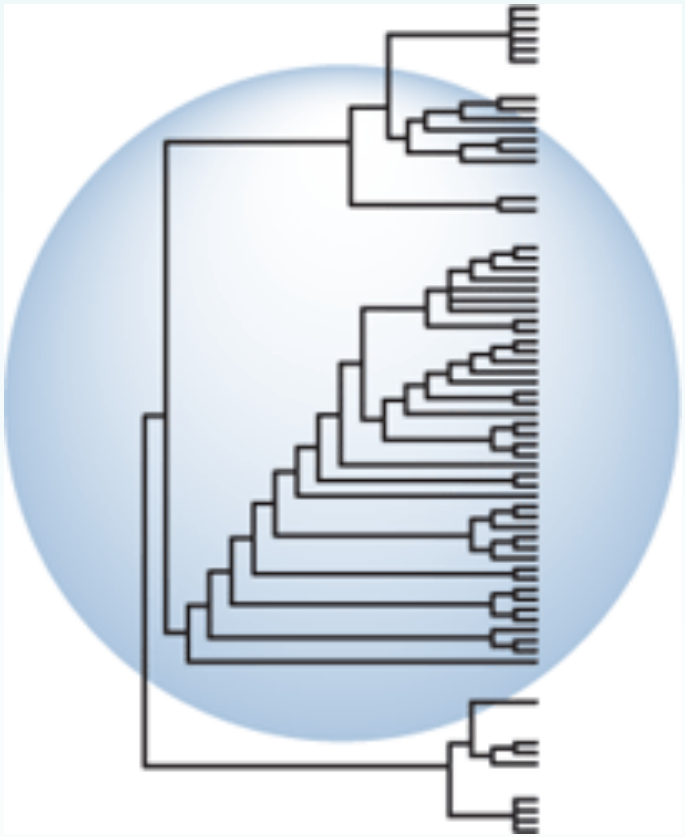
- Cytoscape integration
- InParanoid scaling
- Homolog resolution tools
- Visualization interfaces
- Graphing interfaces
- Kbase collaboration
- Additional algorithm support and validation

Association Studies



- Facilitate development of standards for data storage/access
- Identify data conversion needs from stored data to association modeling
- Full-scale genotype analysis pipelines for extracting SNP molecular markers, structural variants, and any other type of variants
- Full-scale pipelines using different statistical methods for association modeling
- Data publication and storage in “iPlant Data Commons”
- Support materials for training and outreach

Phylogenetic Support



- Interconnected tools that allow users to obtain, analyze, and visualize a phylogenetic tree by submitting a list of taxa or a trait matrix.
- Perpetually updating tree of green plants.
- Extension of rPlant to non-phylogenetic applications.
- One publication describing the iPToL infrastructure.
- Yearly reviews of iPToL infrastructure.
- One set of documents and tutorials to support EOT activities and a revised set at a later date.
- One publication that highlights best practices; revised publication at a later date.

A Few Success Stories

iPlant Collaborative™

Asian Wild Rice Distribution

The Research

- Genetic studies documented geographic subdivision of Asian wild rice (*Oryza rufipogon), the progenitor of* cultivated Asian rice.
  - Cause unknown.
- Use species distribution modeling (SDM) to examine environmental factors associated with the spatial and temporal distribution of *O. rufipogon.*
- Compare estimated distribution during Last Glacial Maximum (LGM) to genetic data.

Problem

- Analysis requires large datasets

iPlant Workshop at BSA, July 2011

- Pu Huang (Washington U.) attended.
- Learned about Atmosphere, iPlant's cloud computing platform.

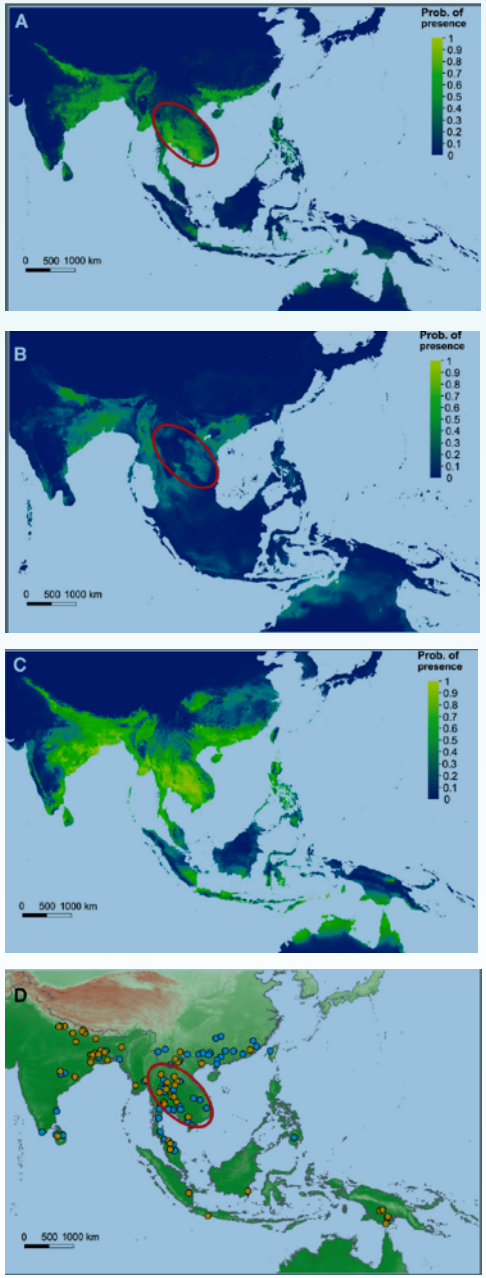
Results

- Present distribution of *O. rufipogon* (Fig. A).
- Projected paleodistribution at LGM was separated into disconnected east and west ranges (Fig. B).
  - Consistent with current geographic pattern of genetic variation, with two genetic groups that intergrade (Fig. D).
- Annual precipitation contributes most to SDM estimates.
- SDM projections for year 2080 indicate an increasing probability of presence and range expansion (Fig. C).
  - Indicates global warming is less threat to this endangered species than other human-mediated factors.

<u>Scalable science</u>

- 325 records of O. rufipogon sample locations from two sources.

iPlant enabled Huang and Schaal to successfully pursue this research.



(A) present, (B) Last Glacial Maximum, (C) Future 2080, (D) Genetic variation.

P Huang and B.A. Schaal, Am. J. Botany 99(11). 2012.

www.iplantcollaborative.org

Animal Genomics

The problem

- Multiple emergent large-scale projects in agricultural animal genomics, genetics, and stress physiology using next-gen sequencing
- National Animal Genome Research Program (NRSP-8) : *Minimal* agency-level resources for scalable computing, storage, and collaboration
- Coordinator for NRSP-8 contacted iPlant based on word-of-mouth

Our approach

- Extend iPlant support and resources to this program
  - iPlant Data Store
  - Foundation API
  - iPlant Discovery Environment
- Educate community members to develop and implement scalable versions of their own pipelines and algorithms

Results

Rapid adoption of iPlant Data Store

- 1000 Bull Genomes
- Water Buffalo SNP-chip
- Swine and Chicken Heat Stress Genetics
- Bovine, Sheep, & Horse Genome Projects
- ~40 TB data (and growing)

Advanced training: Three-day onsite "Introduction to Developing for iPlant" at TACC in July 2012 for animal genomics bioinformaticians

Scalable science:

- 192 CPU BWA alignment pipeline
- 768 CPU GATK-based genotyping pipeline
- 32 CPU RNAseq mapping pipeline
- More on the way…

We have radically transformed the process of animal genomics for these communities…

"The ability to transport 2 TB of data overnight using the iRODS system was particularly helpful because previously, we had been mailing hard drives which is not an optimal solution to sharing big data"

"We've successfully used iPlant to map buffalo sequencing data from [multiple] breeds to the bovine genome and the [water] buffalo genome for SNP and INDEL detection. This took only a few days, where it would have taken more than a month previously. That allowed us to help the buffalo community quickly create a SNPchip on a short timeframe and allowed us to more quickly provide variants for use in defining genetic diversity in water buffalo."

"Among the most helpful aspects of using iPlant has been the ability to more efficiently conduct collaborative research... Our collaborators have been able to use tools at iPlant to conduct RNAseq analyses and variant calling. [iPlant resources] have helped individuals with very limited programming experience do bioinformatics quickly so that they can spend more time working on understanding the biology related to their areas of study."

OneKP

The problem

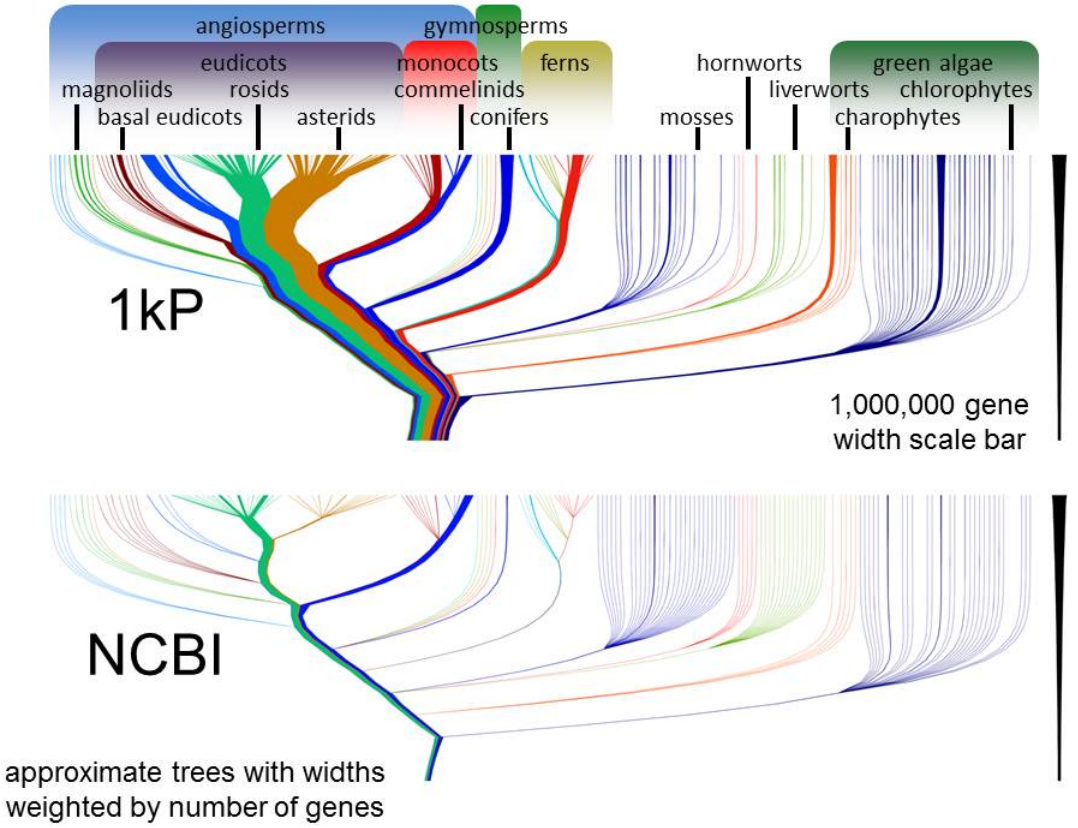
- OneKP: consortium formed to sequence the transcriptomes of 1000 phylogentically diverse plant species.
- Needs: storage, access to compute resources and expertise, distribution.

Our approach

- Assign personnel with expertise in the required fields to the project
- Cover storage and computational needs

Results

•iPlant is replicating the entire dataset including raw reads, assemblies and analysis results
•Annotated 86 million contigs against NCBI's RefSeq using BLASTX
•Identified the open reading frames and estimated the protein sequences resulting in 19,556,877 potential genes
•Will increased the number of plant genes in GenBank by a factor 100.



- Scrubbed all names to match NCBI taxa names (20% could originally not be matched)
- iPlant will be offering BLAST and search services against the OneKP results in the next DE release
- The optimized BLASTX and translation pipeline as available to the community through the Discovery Environment

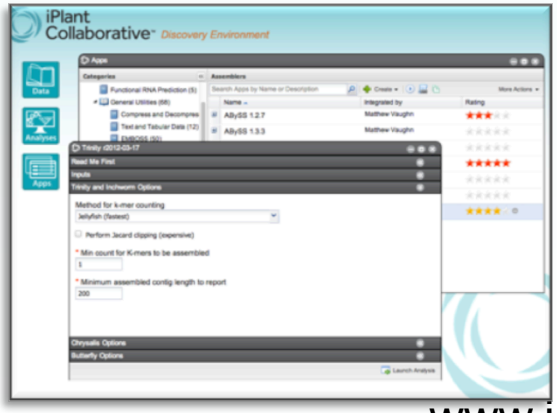
Assembly and Annotation

The problem

- Full-scale genome and transcriptome sequencing is affordable and accessible
- Assembly and knowledge extraction remains challenging
- Extremely computationally intensive. Complex, low-efficiency software. Command-line only.

Our approach

- Provide HPC resources
  - >100k CPUs
  - multi-TB RAM
  - petascale storage
- Optimize workflows and algorithms
- Provide access via Discovery Environment



Results

•Diverse species assembled/annotated: Rice, diploid switchgrass, *Ceratopteris*, several *Solanaceae*, mulberry, maize accessions, *Thellungiella*, barley, wheat, and soybean

•Laboratory groups engaged: >30, including Cornell, Iowa State University, University of Florida, JCVI, Penn State University, CSIRO, and Purdue

•Applications deployed to HPC: ALLPATHS, Velvet, Oases, ABYSS, Newbler, SOAPdenovo, SOAPdenovo-Trans, Trinity, Celera Assembler

•HPC applications available via DE: Velvet, ABYSS, Newbler, SOAPdenovo, Trinity, InterproScan

•Current deployment and optimization efforts: Trinity, InterproScan, MAKER

•HPC systems used: PSC Blacklight, TACC Ranger, TACC Lonestar, SDSC Trestles

•Usage statistics:

- 7,000 HPC jobs; 1.5 million computing hours in Y1 of this initiative
- > 1000 HPC-backed assembly/annotation jobs run by iPlant DE users in 8 months

Linking Datasets

The problem

- Biology is a data-driven science.
- 1380 online molecular biology databases (NAR 2012)
- Errors in data result in failure to link data records and erroneous scientific conclusions.
- Scientists spend too much time searching across repositories for needed data, and resolving data errors and inconsistencies.
  - Species names, gene names, geographic place names, latitude & longitude

iPlant's Approach

- Develop a Taxonomic name Resolution Service.
- Invest in Semantic Web technology to create semantically-described data repositories.
- Develop a Homology Resolution Service.
- Develop a Geographic Resolution Service for place names and latitude/ longitude.
- iPlant Data Commons

Results

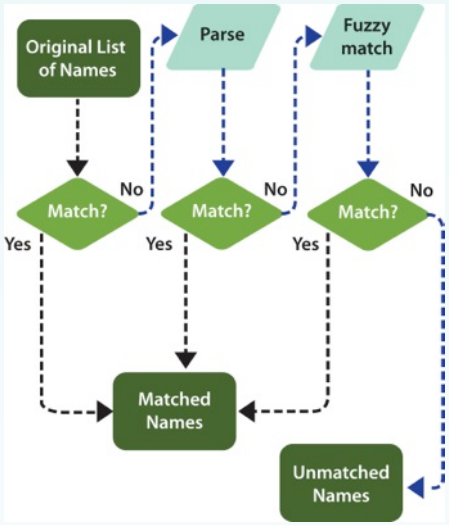
Taxonomic Name Resolution Service (TNRS)

- Web Service resolves large batches of names against four taxonomic sources
- Averaging 1000 hits per month, 57% returning visits, used in 80 countries
- BIEN database contains 200,915 species, 30% fewer than before resolving the names with TNRS.
- TNRS resolved 27,317 plant names in NCBI and ITIS, increasing name intersection by 15%.

Simple Semantic Web Architecture and Protocol (SSWAP)

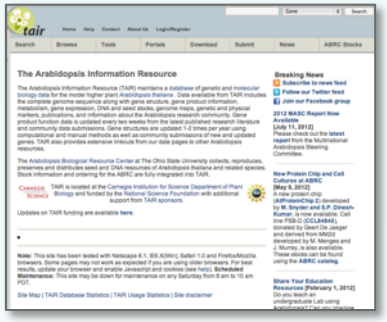
- 11 Dendrome TreeGenes web services are semantically enabled (contig keywords and records, tree and region samples, amplicons, alignments, ESTs, primers, SNPs, MultiFasta reads, Ameriflux data)
- Semantic reasoning underpins drag and drop workflow creation

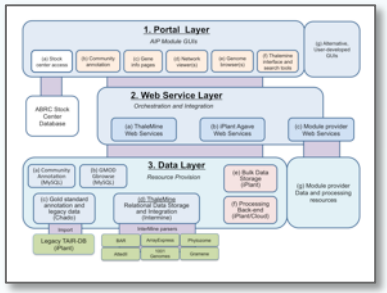
TNRS

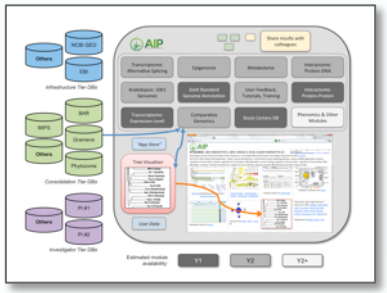


iPlant is delivering technologies to link biological datasets.

TAIR







The problem

- End of life: After 10 years of service, funding for TAIR (The *Arabidopsis* Information Resource) is ending
- Vast community: TAIR is estimated to serve > 10,000 users
- Complex architecture: TAIR has grown organically over its lifespan and is comprised of a heterogeneous collection of hardware, operating system, databases, scripting languages, and services
- End of support: No resources for maintenance beyond funding period
- Emergent data: Need for continuing development of *Arabidopsis* data resources

Our approach

- Use iPlant Atmosphere persistent virtualization to clone the entire TAIR infrastructure onto a few virtual machine hosts
- Migrate network addresses to virtualized TAIR when complete, for seamless transition
- Partner with the *Arabidopsis* community in designing and crafting the TAIR replacement

Result

•TAIR -> iPlant migration underway with expected completion in early 2013

•D. Stanzione & R. Dooley advise IAIC (International Arabidopsis Informatics Consortium) RCN

•Vaughn is co-PI on NSF ABI proposal to develop scalable, extensible *Arabidopsis* Information Portal (AIP) atop iPlant CI

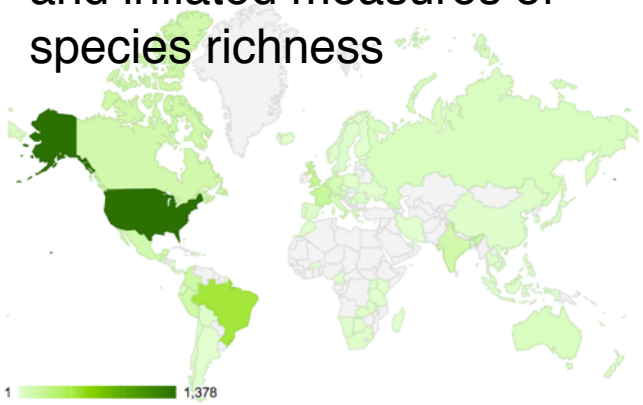
www.iplantcollaborative.org

iPlant Collaborative™

Taxonomic Name Resolution Service (TNRS)

The problem:

- Increase in biodiversity databases being used to study distribution, function and diversity for previously inaccessible data
- Lack of standardization of names results in mismatched observations and inflated measures of species richness



Data resolution and error reduction:

- **BIEN:** Reduced number of unique species names in database by 30% enabling accurate count of plant species in western hemisphere and computation of range maps for each.
- **TNRS Example:** Corrected 27,317 names in NCBI and ITIS, increasing name intersection by 15%.
- **Global Crop Diversity Trust:** Validated records for wild-types needing conservation.
- **iDigBio:** Comparing multiple sources (including Morphbank) to resolve species trees.

Usage Patterns:

- 57% returning visits to application
- Averaging nearly 1000 hits per month since release in June 2012
- Over 80 countries

Integrated Breeding Platform

Project Highlights

- Provides data, tools, services and network for plant breeders
  - Targeting smallholders in marginal environments
- Accelerate variety development for developing countries using marker technologies:
  - Major gene or transgene introgression
  - Gene pyramiding
  - Complex marker-assisted recurrent selection
  - Near future, genome-wide marker-assisted selection (GWMAS)
- 14 initial use cases (8 crops across 32 countries)
- Gates Foundation, UK Department for International Development and European Commission

https://www.integratedbreeding.net/

iPlant's Role

- Co-developed & hosting IBP's web portal
- Hosting trait dictionaries, tools
- Prototype Configurable Workflow System in Atmosphere
- Liaison between IBP & U.S. breeders
- Future, access to scalable analyses

The iPlant Collaborative - Acknowledgments

Postdocs:
Barbara Banbury
Jamie Estill
Bindu Joseph
Christos Noutsos
Brad Ruhfel
Stephen A. Smith
Chunlao Tang
Lin Wang
Liya Wang
Norman Wickett

Executive Team:
Steve Goff
Dan Stanzione

Students:
Peter Bailey
Jeremy Beaulieu
Devi Bhattacharya
Storme Briscoe
Ya-Di Chen
John Donoghue
Steven Gregory
Yekatarina Khartianova
Monica Lent
Amgad Madkour
Aniruddha Marathe
Kurt Michaels
Dhanesh Prasad
Andrew Predoehl
Jose Salcedo
Shalini Sasidharan
Gregory Striemer
Jason Vandeventer
Kuan Yang

Metadata → Data → Tools → Workflows → Viz

Faculty Advisors & Collaborators:
Ali Akoglu
Greg Andrews
Kobus Barnard
Sue Brown
Thomas Brutnell
Michael Donoghue
Casey Dunn
Brian Enquist
Damian Gessler
Ruth Grene
John Hartman
Matthew Hudson
Dan Kliebenstein
Jim Leebens-Mack
David Lowenthal
Robert Martienssen
B.S. Manjunath
Nirav Merchant
David Neale
Brian O'Meara
Sudha Ram
David Salt
Mark Schildhauer
Doug Soltis
Pam Soltis
Edgar Spalding
Alexis Stamatakis
Ann Stapleton
Lincoln Stein
Val Tannen
Todd Vision
Doreen Ware
Steve Welch
Mark Westneat

Staff:
Greg Abram
Sonali Aditya
Roger Barthelson
Brad Boyle
Todd Bryan
Gordon Burleigh
John Cazes
Mike Conway
Karen Cranston
Rion Doodey
Andy Edmonds
Dmitry Fedorov
Michael Gatto
Utkarsh Gaur
Cornel Ghiban
Michael Gonzales
Hariolf Häfele
Matthew Hanlon
Anthony Heath
Barbara Heath
Matthew Helmke
Natalie Henriques
Uwe Hilgert
Nicole Hopkins
Eun-Sook Jeong
Logan Johnson
Chris Jordan
B.D. Kim
Kathleen Kennedy
Mohammed Khalfan
Seung-jin Kim
Lars Koersterk
Sangeeta Kuchimanchi
Kristian Kvilekval
Aruna Lakshmanan
Sue Lauter
Tina Lee
Andrew Lenards
Zhenyuan Lu
Eric Lyons
Naim Matasci
Sheldon McKay
Robert McLay
Angel Mercer
Dave Micklos
Nathan Miller
Steve Mock
Martha Narro
Praveen Nuthulapati
Shannon Oliver
Shiran Pasternak
William Peil
Titus Purdin
J.A. Raygoza Garay
Dennis Roberts
Jerry Schneider
Bruce Schumaker
Sriramu Singaram
Edwin Skidmore
Brandon Smith
Mary Margaret Sprinkle
Sriram Srinivasan
Josh Stein
Lisa Stillwell
Kris Urie
Peter Van Buren
Hans Vasquez-Gross
Matthew Vaughn
Fusheng Wei
Jason Williams
John Wregglesworth
Weijia Xu
Jill Yarmchuk

www.iplantcollaborative.org