



Genomes to Fields 2014 Workshop

Maize Phenotypic Information Platform

Carolyn Lawrence

Chicago, IL

December 10



Genomes to Fields

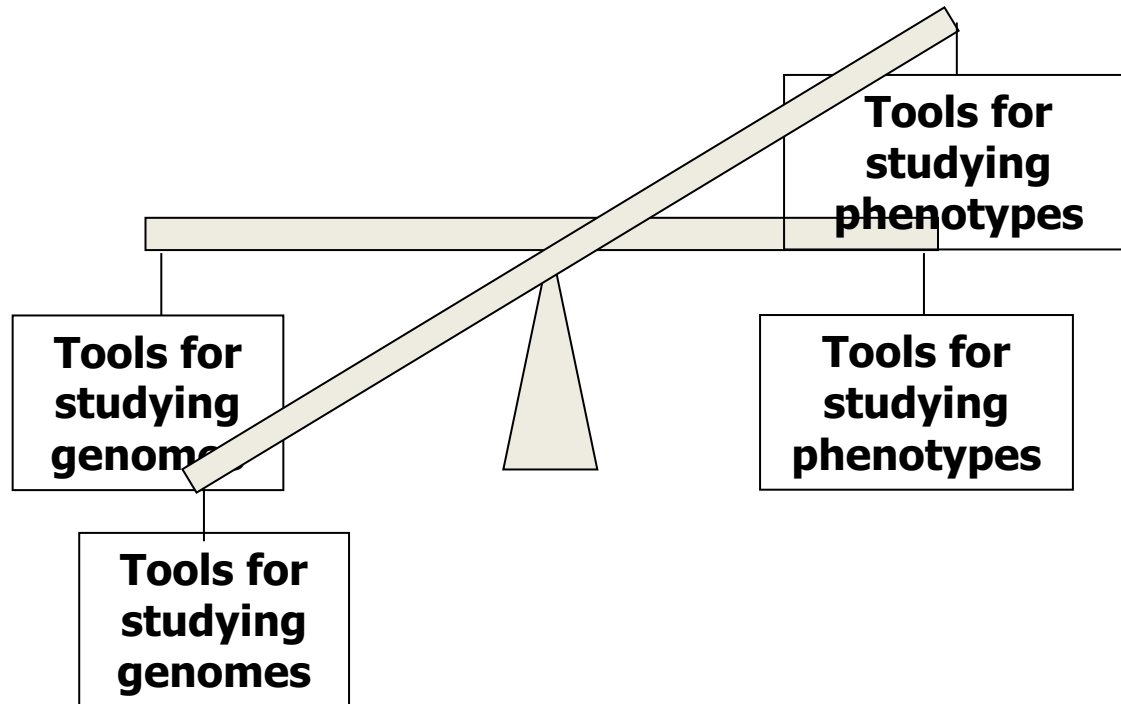
Phenotype

- ✧ A **phenotype** (from Greek *phainein*, 'to show' + *typos*, 'type') is the composite of an organism's **observable characteristics** or traits, such as its morphology, development, biochemical or physiological properties, phenology, behavior, and products of behavior.
- ✧ A phenotype results from the expression of an organism's genes as well as the influence of environmental factors and the interactions between the two.
- ✧ Phenotype is EVERYTHING



Tools for genotype and phenotype

~imbalanced~



Slide credit: Edgar Spalding

Why is managing 'phenotype' hard?

- ✧ Extremely diverse data type
- ✧ Associated to individuals, populations, or species
- ✧ Data documented at different levels (summary -vs- measurement)
- ✧ Comparative (mutant –vs- wild type) or absolute (plant height)
- ✧ Different terms between disciplines (stacking a trait)

- ✧ Data integration - needs extensive connections to other types of data
(seed stocks, genes, experimental methods, environment data)
- ✧ Data representation - how to represent the data in a consistent way
across experiments, research groups, and organizations
- ✧ Data accessibility - we must get data into others' hands



Phenotypes are Big Data

Big Data is characterized as having **extreme or variable values** of one or more of the following characteristics:

- ✧ Volume¹ (size) Images, sequence, expression data
- ✧ Velocity¹ (acquisition rate) Images, sequence
- ✧ Variety¹ (structure) Data formats, alternative standards
- ✧ Variability² (in meaning) Nomenclature, ontologies
- ✧ Complexity³ (in relationships) Mutation to genotype to phenotype...
- ✧ Veracity (quality or provenance) Gold standard datasets, low quality ones
- ✧ Volatility (changes over time) Versions of data

1 Doug Laney, "3-D Data Management: Controlling Data Volume, Velocity, and Variety," 2001.

2 Brian Hopkins, "Blogging From the IBM Big Data Symposium - Big Is More Than Just Big," 2011.

3 Valentin T Srihar, et al., "'Big Data' Is Only the Beginning of Extreme Information Management," 2011.



Information management food chains

Knowledge



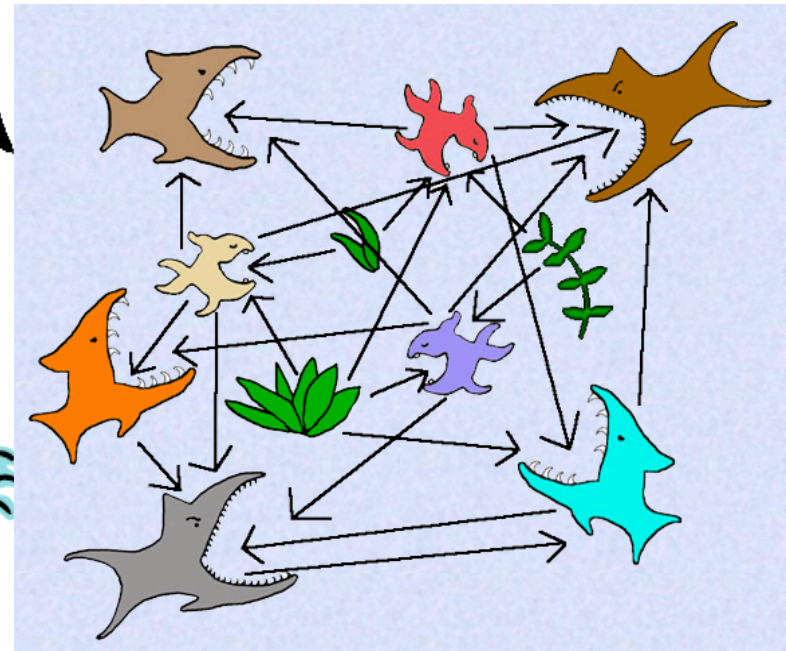
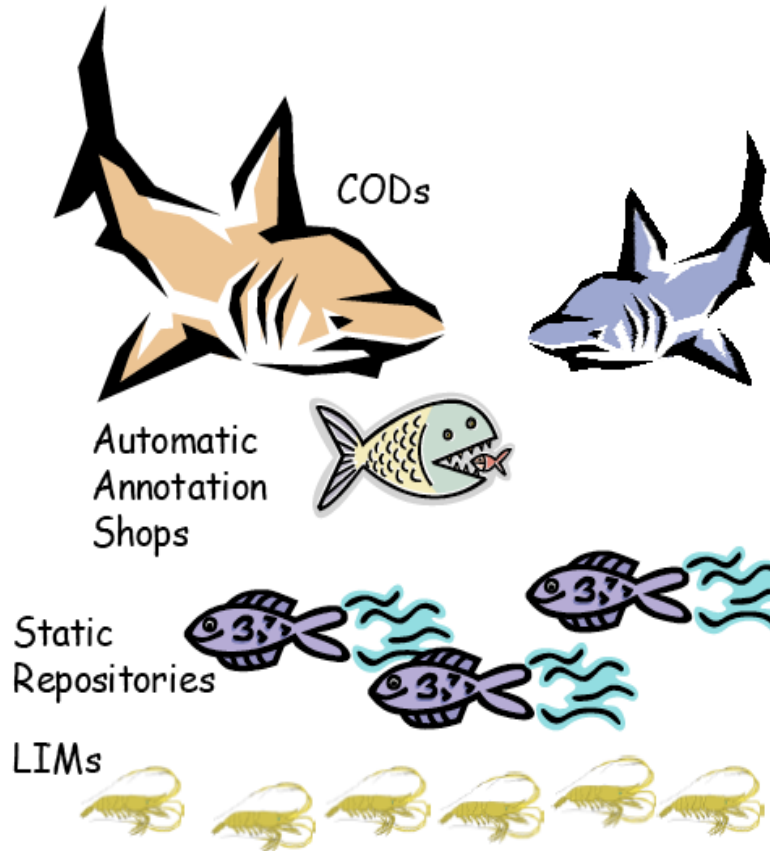
Presentation



Information



Data



Plant Biology Databases: A Needs Assessment (2005)

<http://www.cod.edu/people/faculty/fancher/FoodChain.htm>



Our approach to information management

✧ Information must be **communicated**

...or it is effectively lost

✧ Results should be **reproducible**

...or we're not doing science



Balancing act: enforcing standards AND allowing flexibility

✧ For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights

Data scientists spend from 50 percent to 80 percent of their time mired in the mundane labor of collecting and preparing unruly digital data before it can be explored for useful nuggets.

-New York Times 17 August 2014

- ✧ For well-understood datatypes **standards** and **adapters** are required to compute across multiple datasets effectively
- ✧ For emerging datatypes like high-throughput phenotyping, imposing **standards could impede** the development of novel and revolutionary data documentation and analysis techniques



How we build information platforms

Explore the Landscape



Engage with Providers

- ✧ iPlant
- ✧ MaizeGDB
- ✧ IBP/BMS
- ✧ Private Industry
- ✧ Others...

Agile Development & Spiral Design



Two simultaneous activities



Building the ship as we sail

Assembling the platform



Building the ship as we sail



✧ Genotype

- ✧ GBS, etc.

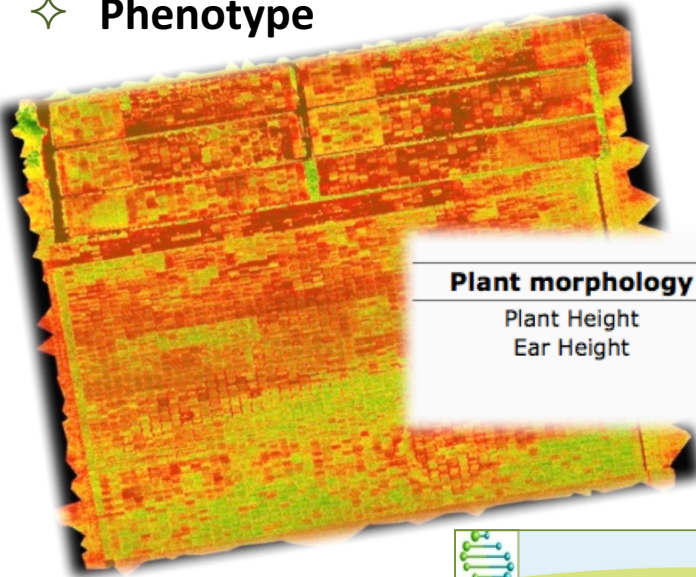


✧ Environment

- ✧ Weather stations
- ✧ GIS (soil, interpolated weather, model)



✧ Phenotype



Core Traits

Plant morphology	Ear morphology	Agronomic	Productivity
Plant Height	Ear width	Stand count	Grain moisture
Ear Height	Ear length	Root lodging	Test weight
	Cob width	Stalk lodging	Grain yield
	Kernel row number	Days to silking/anthesis	



Assembling the platform

- ✧ July 25 SOUGHT COMMUNITY INPUT
 - ✦ Convened more than 20 researchers, data providers, and industry colleagues
 - ✦ Described the needs and asked for input
 - ✦ Followed up with many, identified IBP/BMS, CGBackOffice and MaizeGDB, and iPlant as leading groups to investigate for partnerships

- ✧ August BEGAN DEVELOPER CALLS WITH IBP

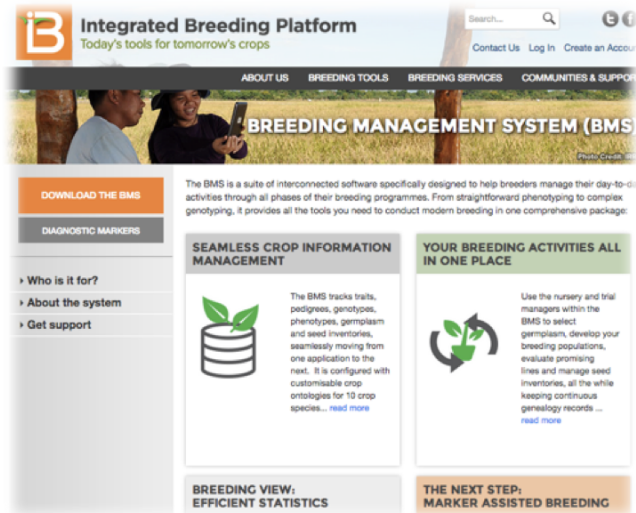
- ✧ September REQUESTED IPLANT INVOLVEMENT

- ✧ November 9 – 13 COORDINATED MEETING AT CIMMYT
 - ✦ Visited CIMMYT to meet with IBP/BMS curators (Kate Dreher, Julian Pietragalla, and Clarissa Pimentel)
 - ✦ Invited iPlant personnel (Ramona Walls)
 - ✦ Invited BMS developers from Indiana (Jan Erik Backlund) and New Zealand (Rebecca Berringer)

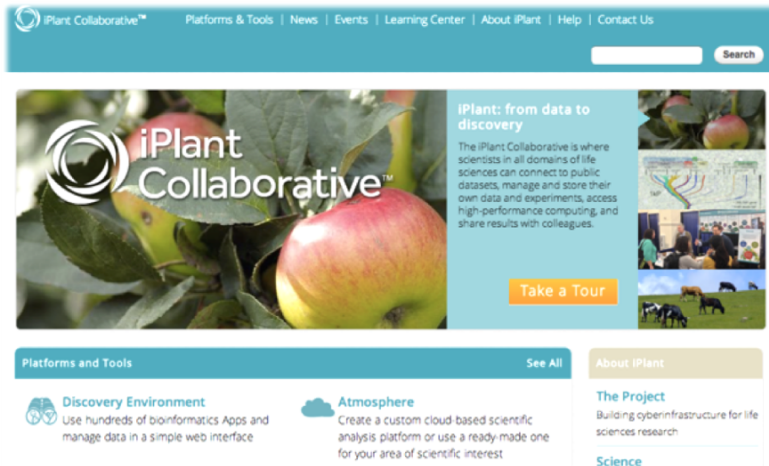
- ✧ December 1 and 2 COORDINATED MEETING AT ISU
 - ✦ Invited iPlant personnel to visit and describe the platform (Nicole Hopkins and Jeremy DeBarry)
 - ✦ Brought in G2F leaders (Pat Schnable, David Ertl) and GxE data coordinator (Jode Edwards)
 - ✦ Brought in ISU BMS outreach personnel (Walter Suza, Assibi Mahama)
 - ✦ Brought in CGBackOffice team (Ed Buckler and Cinta Romay)



Assembling the platform



CG Back
Office
Buckler Lab



Genomes to Fields



✧ Assets:

- ✧ Funded into the future
- ✧ Manage all well-described datatypes now
- ✧ Community investment worldwide

✧ Initial Concerns:

- ✧ Deployed on local machine
- ✧ Pay model
- ✧ Modules specific to PC's
- ✧ Service and collaborative development



✧ Assets:

- ✧ Funded into the future
- ✧ Documentation
- ✧ HT image handling
- ✧ Desire to make G2F a success story of their own

✧ Initial Concerns:

- ✧ Disappointment of many researchers early on
- ✧ Will they invest time/effort in G2F?
- ✧ Can their systems be adapted to work together?



PRODUCTS



FOUNDATIONAL SERVICES



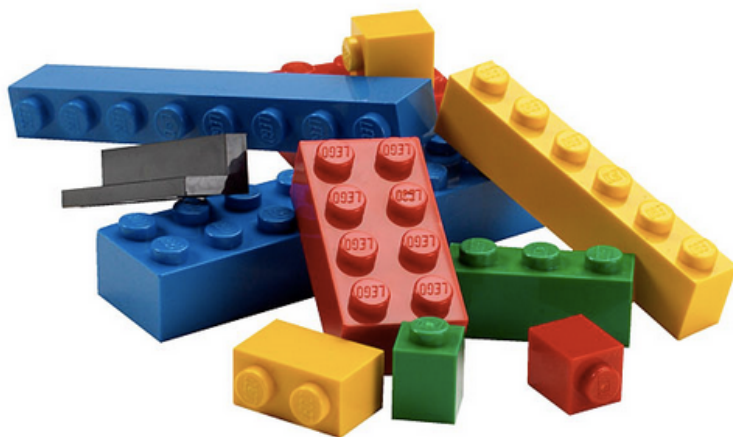
LOW-LEVEL SERVICES, SECURITY, ASSETS, etc

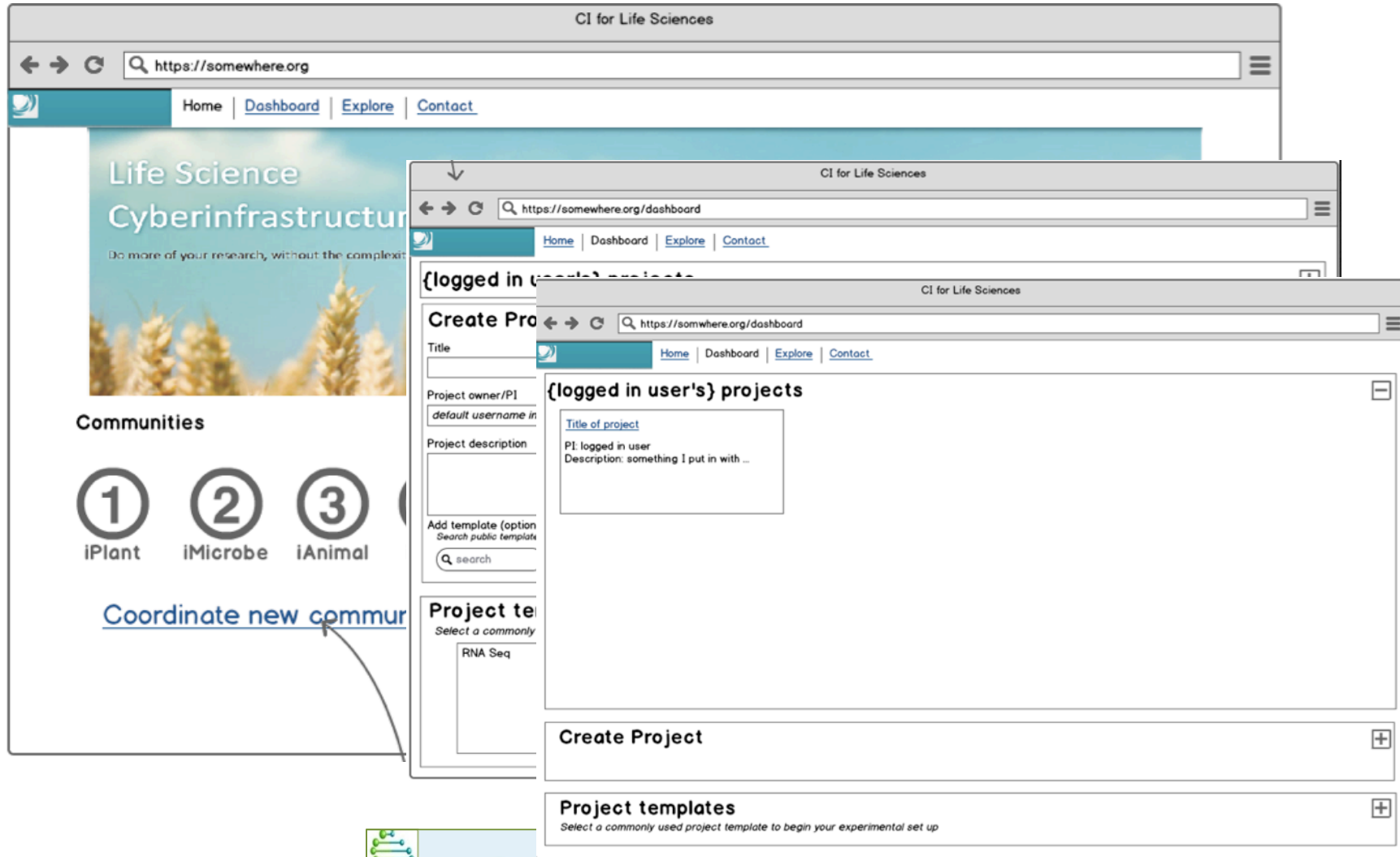


HARDWARE RESOURCES



Need	iPlant Service
data transfer and management	DE, iDrop, iCommands
images management and analysis	Bisque
working with BMS	Atmosphere, permanent VM, volumes
manage G2F project	Projects interface





CI for Life Sciences

← → ↻

Home | [Dashboard](#) | [Explore](#) | [Contact](#)

Life Science Cyberinfrastructure
Do more of your research, without the complexity

Communities

1 iPlant 2 iMicrobe 3 iAnimal

[Coordinate new communities](#)

CI for Life Sciences

← → ↻

Home | [Dashboard](#) | [Explore](#) | [Contact](#)

{logged in user's} projects

Create Project

Title

Project owner/PI

Project description

Add template (optional)
Search public template

Project templates
Select a commonly used project template to begin your experimental set up

RNA Seq

Create Project

Project templates
Select a commonly used project template to begin your experimental set up



iPlant Collaborative™
Learning Center



Discovery Environment

DE 002: Characterizing Differential Expression with RNA-Seq (Tuxedo Method)

Pre-requisites and Data
Important Notes

A. Align RNA-Seq Reads to Reference Genome

B. Assemble Transcripts

C. Merge Transcripts

D. Compare Gene Expression

Final Results and Next Steps

DE 002: Characterizing Differential Expression with RNA-Seq (Tuxedo Method)



Goal: Identify all changes in gene expression levels between at least two sequenced transcriptome samples.

Approximate tutorial completion time: 2 hrs. when using the pre-computed iPlant sample data from a study in *Arabidopsis thaliana* (H. Zhang *et al.*, 2010).

Introduction and Tutorial Overview

RNA-Seq refers to whole transcriptome sequencing of cDNA, generally using an ultra-high-throughput ("next-generation") sequencing technology. The value of the RNA-Seq experiment is the ability to generate deep-coverage information about a sample's RNA. This can be used for a variety of purposes, such as: transcriptome assembly, gene discovery and annotation. RNA-Seq is also used to detect differential transcript abundance between tissues, developmental stages, genetic backgrounds, and environmental conditions.

Method: Tuxedo Protocol v2

Original Workflow Documentation: "Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks" Nature Protocols 7, 562–578 (2012) doi:10.1038/nprot.2012.016

iPlant Workflow

Platform	Objective	Apps
	A. Align sequence reads to a reference genome	<i>TopHat2-PE, TopHat2-SE</i>
	B. Assemble read data into transcripts	<i>Cufflinks2</i>



Genomes to Fields



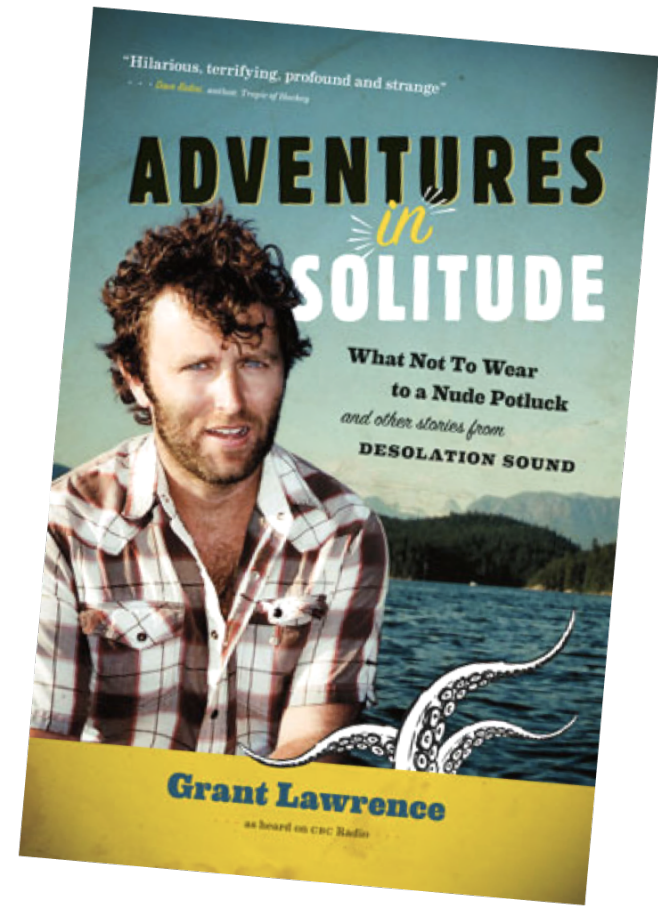
What will make this successful?

- ✧ Making data logically accessible

- ✧ Working together:

Beware of solitude!

- ✧ Discuss the problems in diverse group settings – *iteratively*
- ✧ Create manuals, example usage cases, and outreach materials and opportunities



Timeline

✧ Training/Outreach/Feedback:

Using BMS & iPlant for data access & analysis

✧ January 26 (?) in Ames

✧ March 12 at Maize Genetics Conference

✧ Anticipated milestones:

✧ December: BMS deployed on iPlant

✧ January:

✧ Project management and methodologies listed at iPlant via wiki

✧ BMS, iDrop, Datastore, and Bisque for current datasets

✧ May: Project Coordination functionality new at iPlant

✧ June (?): CGBackOffice coming online

Agile
Development &
Spiral Design

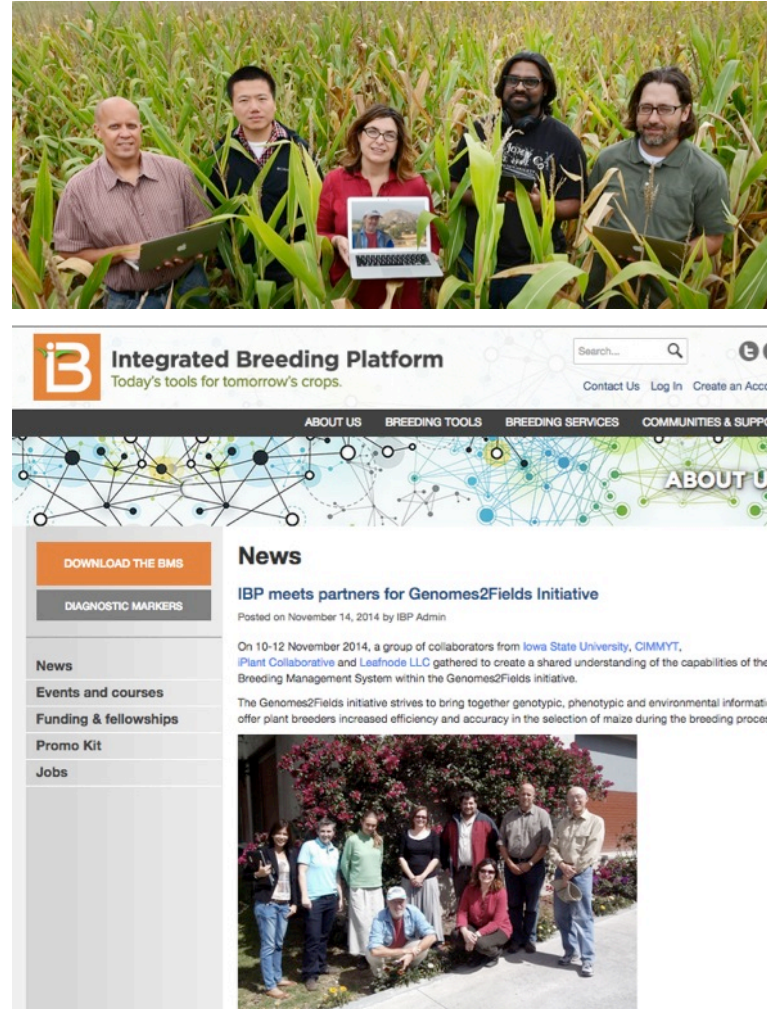


**Our goal is to enable
the process, not to
deploy a specific
system**



Acknowledgements

- ✧ Lawrence Lab
 - Jack Gardiner
 - Darwin Campbell
- ✧ GxE subgroup of G2F
 - Jode Edwards
 - Martin Bohn
 - Natalia de Leon
- ✧ CIMMYT, BMS, Leafnode,
- ✧ MaizeGDB
 - Carson Andorf
- ✧ iPlant
 - Nicole Hopkins
 - Jeremy DeBarry
 - Ramona Walls



WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SNOCHOSTEXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE UNKIDDE PARRIES
WHY ARE OLD KINGDOMS DIFFERENT

WHY ARE THERE SQUIRRELS

WHY IS THERE A 0 ON THE ROSSBOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON OHN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL

WHY AREN'T MY ARMS GROWING

WHY ARE THERE WEEDS IN MY DOG I FEEL DIZZY

WHY ARE THERE SLURPS OF DARTS

WHY IS THERE PHLEOM

WHY ARE THERE SO MANY CROWS IN ROCHESTER, NY

WHY IS PSYCHIC WEAK TO BUG

WHY DO CHILDREN GET CANCER

WHY IS POSEIDON ANGRY WITH ODYSSEUS

WHY IS THERE ICE IN SPACE

WHY ARE THERE ANTS IN MY LAPTOP

WHY IS THERE AN OWL IN MY BACKYARD

WHY IS THERE AN OWL OUTSIDE MY WINDOW

WHY IS THERE AN OWL ON THE DOLLAR BILL

WHY DO OWLS ATTACK PEOPLE

WHY ARE AK 47s SO EXPENSIVE

WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE

WHY ARE THERE GODS

WHY ARE THERE TWO SPOOKS

WHY IS MT VESUVIUS THERE

WHY DO THEY SAY T MINUS

WHY ARE THERE OBELISKS

WHY ARE WRESTLERS ALWAYS WET

WHY ARE OCEANS BECOMING MORE ACIDIC

WHY IS ARWEN DYING

WHY AREN'T MY QUAIL LAYING EGGS

WHY AREN'T MY QUAIL EGGS HATCHING

WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SNOCHOSTEXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE UNKIDDE PARRIES
WHY ARE OLD KINGDOMS DIFFERENT

WHY ARE THERE SQUIRRELS

WHY IS THERE A 0 ON THE ROSSBOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON OHN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL

WHY ARE THERE TINY SPIDERS IN MY HOUSE

WHY DO SPIDERS COME INSIDE

WHY ARE THERE HUGE SPIDERS IN MY HOUSE

WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE

WHY ARE THERE SPIDERS IN MY ROOM

WHY ARE THERE SO MANY SPIDERS IN MY ROOM

WHY DO SPIDER BITES ITCH

WHY IS DYING SO SCARY

WHY IS THERE NO GPS IN LAPTOPS

WHY DO KNEES CLICK

WHY AREN'T THERE E GRADES

WHY IS ISOLATION BAD

WHY DO BOYS LIKE ME

WHY DON'T BOYS LIKE ME

WHY IS THERE ALWAYS A JAVA UPDATE

WHY ARE THERE RED DOTS ON MY THIGHS

WHY IS LYING GOOD

WHY IS SEX SO IMPORTANT

WHY ARE THERE SLAVES IN THE BIBLE

WHY IS HTTPS CROSSED OUT IN RED

WHY IS THERE A LINE THROUGH HTTPS

WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK

WHY IS HTTPS IMPORTANT

WHY AREN'T MY ARMS GROWING

WHY ARE THERE WEEDS IN MY DOG I FEEL DIZZY

WHY ARE THERE SLURPS OF DARTS

WHY IS THERE PHLEOM

WHY ARE THERE SO MANY CROWS IN ROCHESTER, NY

WHY IS PSYCHIC WEAK TO BUG

WHY DO CHILDREN GET CANCER

WHY IS POSEIDON ANGRY WITH ODYSSEUS

WHY IS THERE ICE IN SPACE

WHY ARE THERE ANTS IN MY LAPTOP

WHY IS THERE AN OWL IN MY BACKYARD

WHY IS THERE AN OWL OUTSIDE MY WINDOW

WHY IS THERE AN OWL ON THE DOLLAR BILL

WHY DO OWLS ATTACK PEOPLE

WHY ARE AK 47s SO EXPENSIVE

WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE

WHY ARE THERE GODS

WHY ARE THERE TWO SPOOKS

WHY IS MT VESUVIUS THERE

WHY DO THEY SAY T MINUS

WHY ARE THERE OBELISKS

WHY ARE WRESTLERS ALWAYS WET

WHY ARE OCEANS BECOMING MORE ACIDIC

WHY IS ARWEN DYING

WHY AREN'T MY QUAIL LAYING EGGS

WHY AREN'T MY QUAIL EGGS HATCHING

WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY ARE THERE SLAVES IN THE BIBLE

WHY IS HTTPS CROSSED OUT IN RED

WHY IS THERE A LINE THROUGH HTTPS

WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK

WHY IS HTTPS IMPORTANT

WHY AREN'T MY ARMS GROWING

WHY ARE THERE WEEDS IN MY DOG I FEEL DIZZY

WHY ARE THERE SLURPS OF DARTS

WHY IS THERE PHLEOM

WHY ARE THERE SO MANY CROWS IN ROCHESTER, NY

WHY IS PSYCHIC WEAK TO BUG

WHY DO CHILDREN GET CANCER

WHY IS POSEIDON ANGRY WITH ODYSSEUS

WHY IS THERE ICE IN SPACE

WHY ARE THERE ANTS IN MY LAPTOP

WHY IS THERE AN OWL IN MY BACKYARD

WHY IS THERE AN OWL OUTSIDE MY WINDOW

WHY IS THERE AN OWL ON THE DOLLAR BILL

WHY DO OWLS ATTACK PEOPLE

WHY ARE AK 47s SO EXPENSIVE

WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE

WHY ARE THERE GODS

WHY ARE THERE TWO SPOOKS

WHY IS MT VESUVIUS THERE

WHY DO THEY SAY T MINUS

WHY ARE THERE OBELISKS

WHY ARE WRESTLERS ALWAYS WET

WHY ARE OCEANS BECOMING MORE ACIDIC

WHY IS ARWEN DYING

WHY AREN'T MY QUAIL LAYING EGGS

WHY AREN'T MY QUAIL EGGS HATCHING

WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY ARE ULTRASOUNDS IMPORTANT

WHY ARE ULTRASOUND MACHINES EXPENSIVE

WHY IS STEALING WRONG

WHY AREN'T THERE GUNS IN HARRY POTTER

WHY ARE THERE WEEDS IN MY DOG I FEEL DIZZY

WHY ARE THERE SLURPS OF DARTS

WHY IS THERE PHLEOM

WHY ARE THERE SO MANY CROWS IN ROCHESTER, NY

WHY IS PSYCHIC WEAK TO BUG

WHY DO CHILDREN GET CANCER

WHY IS POSEIDON ANGRY WITH ODYSSEUS

WHY IS THERE ICE IN SPACE

WHY ARE THERE ANTS IN MY LAPTOP

WHY IS THERE AN OWL IN MY BACKYARD

WHY IS THERE AN OWL OUTSIDE MY WINDOW

WHY IS THERE AN OWL ON THE DOLLAR BILL

WHY DO OWLS ATTACK PEOPLE

WHY ARE AK 47s SO EXPENSIVE

WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE

WHY ARE THERE GODS

WHY ARE THERE TWO SPOOKS

WHY IS MT VESUVIUS THERE

WHY DO THEY SAY T MINUS

WHY ARE THERE OBELISKS

WHY ARE WRESTLERS ALWAYS WET

WHY ARE OCEANS BECOMING MORE ACIDIC

WHY IS ARWEN DYING

WHY AREN'T MY QUAIL LAYING EGGS

WHY AREN'T MY QUAIL EGGS HATCHING

WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

QUESTIONS

FOUND IN GOOGLE AUTOCOMPLETE

