PAG31

**Friday, 20240112**

**Kbase for plant science**

***Bob Cottingham***

**Key words**: Kbase, network, transcriptome

* Introduction about the capacity of Kbase and improvement recently
* Three national lab
* Also **a education tool**; materials for teaching

1. Handon data science research and analysis without programming skills
2. Workshop materials

* Monday night, workshop on more case studies

Here is a reference about Kbase: https://www.nature.com/articles/nbt.4163

***Priya Ranjan***

Oak Ridge National Laboratory

* Introduction + use case transcriptomics and metabolism analysis and gene network
* Empowering researchers by integrating data and tools to accelerate analysis and support collaboration.

1. Integrate data
2. Access tools
3. Customize analysiss
4. Share with collaborator
5. Publish data

* Upload sequencing reads, assemblies isolate genomes MAGS metabolic models, expression data and more

Support Globus for large dataset

* With bacterial archeal gennomes

Plant genomes – Phytozome v13

Fungal genomes

* 200 apps

1. App catalog
2. Software development kit

* Website page have the components

1. Markdown cell
2. Code cell
3. App cell
4. Data object view

* Facilitate collaboration and sharing data and analysis
* Also can create an organization for a big project

Also cources, worshops

* Case study

e.g. a KBase case study on genomepwide transcriptomics and plant metabolism in response to drought tress in Sorghum.

* More than 30,000 users, more than 50TB user data
* Transcriptome and metabolism workflow

1. Genome
2. RNAseq reads

* [WWW.kbase.us/pag-2024/](http://WWW.kbase.us/pag-2024/) demo page
* Example of transcriptome analysis

1. Genome
2. Sequence read
3. Analysis

QC

Alignment

Assembly

Function enrichment

Metabolism interpretation

* System based on Docker. ---- what is docker?

**Network** biology Kbase workflows

How to prioritize annotations of genes in a geneset

Find the most relavant genes to a bioogial process of interest

Gene to a traits

* Multiplex networks embrace the complexity of interations and relationship between layer of information

1. Based on literature curated data and generated layer using hihigh performance computing from Arabidopsis and Populus trichocarpa.

------ might also work on real crps, not only models.

1. Use ***random walk with restart algorithms***, determine the connectivity of genes in a geneses to those in the multiplex enabling the prioritization fo genes and revealing the mechanistic insights.

* User can do

1. Visualize network connections and see what lines of evidence connect genes

e.g., protein protein interaction, coexpression,

multiple evidence about the interaction

1. Prioritized the most highly connected genes to pathway
2. Identify mechanistic connections between genes and create mechanistic models.

GO terns, function, pathway

**MENTOR**: multiplex embedding of network topology for omics resources

* RWRtools LOE pipeline genes in the geneset are clustered based on ***their functional connectivity as determined by genes in multiplex***
* User
* A) visualize genes in the gene set that are highly functionally connected to toehr genes
* Work in teams to discover mechanistic connects between genes and create mechanistic models
* Coming soon
* GWAS is coming also probably

**Action:**

* Create an account, receiving news about webinars
* Q&A,

**Parking lot:**

* ? free?
* ? efficiency? New tools?
* ? balance programming and using Kbase
* ? further development
* ? breeding or genetics study? Mainly for discovery?
* ? challenges the user might encounter?
* ? trouble shooting?
* Long-term plan of the develpment

This is a good area and tools to explore the gene interaction, not the single gene

Even single genes, also need to understand network involved of the single genes.

How I am going to use Kbase?

* What is the case studies? Switchgrass

NDSU

Agronomic good and mature fast in ND

* Faller x Jinqiang 5

1. 1B and 7B QTL, new
2. Greenhouse 1B QTL, not stable
3. 7B, early maturity from Chinese variety --- FT1?

* UAV bacterial leaf streak

Boost x ND830, RxS

Inoculation

=====

Shifeng Cheng

* NAM population
* HapMap

LD based

K-mer hapmap

**Imputed** the NAM population with HapMap for NAM GWAS

Identified many new QTL

* Flowering time Ppd, VRN, and FT
* Nitrogen use efficiency. NLP gene
* Combining QTL mapping, GWAS and NAM GWAS

NAM GWAS- combining QTL and GWAS

Providing many QTLs with the capacity to identify the candidate genes

[www.g2b.com](http://www.g2b.com)

HapMap

* Trait antagonism (trade off) --- pleiotropic effects

Negative correlation between traits

Grain yield, harvest index, plant height, biomass

* Develop NIL using marker assisted backcrossing
* Delivery Watkins alleles into commercial breeding – collaboration and benefits
* Recurrent parent YM22
* Developed NIL and super NAM population
* Global Wheat project g2B – Genomics-design wheat breeding CIMMYT-China-UK

Think BIG and design BIG!

**Action**:

* Check the website, G2B
* Find and Read the paper published by Shifeng Cheng

https://www.biorxiv.org/content/10.1101/2023.10.04.560903v1

Parking lot:

* What is the recurrent parents of NIL?

**Harness Sequence-based haplotype diversity for improved yield and climate resilience in wheat**

CIMMYT global WHEAT PROGRAM AND THE international WHEAT IMPROVEMENT NETWORK

* focus

1. Yield
2. Disease
3. Quality

* activities

1. 650 PUBLIC AND PRIVATE TRAINING SITE
2. Disease resistance

* Haplotype heat tolerance without yield penalty

Comms Biology 2023, QTL on 2B and ??

* CIMMYT QTL catalog???
* **Synthetic hexaploidy wheat** showing better heat tolerance than checks, Baj, Borlaug100
* Review paper, bottleneck to crop productivity
* ***Night temperature*** explain more yield variation than day time temperature (plant activities at night)
* How stellate phenotyping use in breeding???
* Data-physiology, phenotype, environment data, deep learning and machine learning.
* Q&A

Mendel traits, but complex traits, it is difficult to control and have to use large data set to drive decision making

**Action:**

* Check the CIMMYT QTL catalog + KASP markers in EiB
* Read the paper about night temperature effect on wheat performance
* Read the comms biology, 2023 QTL on heat tolerance

====

Genotyping array. **BRISTOL.AC.UK**

**Gary Barker**

* WHY

1. Thermos fisher, axiom genotyping

45K SNP 384 format

$43/sample

Highly reproducible

1. Why a new one?

10 years ago

Gene-based SNP

No position, from genetic mapping

Uneven distribution

1. Haplotype optimization

Redundancy, markers on the same haplotype

1. New strategy

Skin sequencing from 217 breeding lines and 111 landraces

Chromosome bins 1.5MB, 400 /Chr

MINILAM MARKERS IN EACH BIN

Up to 7 markers per bin

**So, haplotype optimization within bins – he has good knowledge about the optimal set of markers for wheat QC**

1. 44K, TaNGv1.0, 20% not work

So v1.1 now

27 more SNPs and more evenly distributed

Good correlation between number of SNP and chr size

Less telomeric bias than 35K (what is it?)

Good number of SNP in the intragenic region

Good performance in GWAS, as good as GWS

1. Effectiveness on novel germplasm
2. 383 samples, call rate 98.5%
3. Paper in preprint

Action:

* Read the paper about the design of the array

Parking lot

* Varieties from which countries?
* Based on the RefSeq v2.1?
* How many SNP with >5% MAF in the US and EU population? Difference?

====

Anthony Hall

* 2020, Nature, multiple wheat Genome reveals global variation in modern breeding
* De novo annotation of the wheat pangenome reveals complexity and density within the tetraploid wheat pan transcriptome, under review
* Some varieties have up to34% exotic DNA
* Paragon – elite in UK, large deletion

20 Hifi cell – 33x

45x Oni-C

17x ONT

--- N50, 95.9 Mb

14.4 Gb assigned to chr

4A 5B switch?

* Watkins 824 lines, 7 sources
* 10+ wheat genome project

====

**Nutrient quality – wheat biofortification**

* Fe and Ze deficiency
* Three options

Flour fortification

Management- agronomic fortification

Genetic biofortification

* Require genetic variation for biofortification

Explore the **Watkins collection, 32 countries in 1920-1930s**

* Variation of mineral concentration in elite and landraces

Landrace had high nutrition value but poor agronomic performance

So, **not only concentration but also ug/grain**

* Three RILs, from landraces x **Paragon**

Six minerals, Fe Zn Cu…

**Minerals in straw and total biomass** – mineral harvest index about the partitioning of minerals in the developing grains

* 734 QTL, focus on 23 strong QTL in at least two sample set, high LOD

Clusters in 5A, 6A, 7A

5 Mb region for each QTL for candidate gene identification

Strongest QTL, 5A for Ca content in grains

CSSA02G543300

Parking lot

* Location in grains?
* Accessible?

====

**Septoria experiment, Zymoseptoria tritici, fungus**

* Savary 2019, paper about the damage caused by disease
* Zymoseptoria tritici, fungus

Fast evolve

Resistant tot major fungicide

Apoplastic latent necrotrophy ***Brennan et al, plant pathology***

* Stb gene cloning

Using Watkins LANDRACE COLLECTION

Association mapping

Greenhouse inoculation – Septoria experiment, using 2 races

* QTL, **Chr 3AS, Stb6**
* **Stb15, in Arina, in 60% of EU cultivars, 6AS**

Haplotype analysis for the candidate genes – diversity population

* STBWat1, new but large region,
* Using mutant to confirm the candidate genes of Stb15

Transformation to Fielder, not stable, some event still have necrosis

* Stb15 in diverse wheat lines
* Paper in preprint

Parking lot:

* Agronomic performance penalty?

====

Flowering locus T (FT) --- variation in Watkins population and expression in OE line

Leeds, UK, Laura Dixon lab

* VRN, PPD, earliness per se --- FT
* FT from leaves, transferred and form FT+FD+14-3-3

12 FT + FT15

ALL FT function as FT1???

* Variation of FT-B1,

Five variations, H1-5

---- Co-selection with other flowering-related genes and their variation

Key: E1, missense alleles

* FT-A1, key - promoter variation allele
* FT-D1earliness
* GWAS, FT-A1

Promoter deletion, higher expression

Faster apex and earlier flowering

The promoter might have a repressor binding site

* Other FT, amino acids essential for 14-3-3 binding

FT1-12

* **Overexpression Ppd1**

Only one line had a higher expression

Significant earlier flower

**Long day condition, produced larger seeds, start bigger seeds from day 1 after flowering**

Not different in short-day condition

* Overexpression Ppd1

Upregulate FT,

Rachis, FT1, FT2 FT4

Ppd regulating FT is tissue-specific???

=====

**High throughput field phenotyping**

* Sensors on different platform
* Focus on stem elongation phase – critical phase for formation of sink potential

***Lengthen SE phase?***

Earliness pe se gene for fine-turn

* Stuction from motion for digital elevation model plot height

***Time of stem elongation + temperature response***

Not sure how he did accumulation???

Validated the significant correlation between SE and yield.

* NILs

Paper in Roth Kroonenberg et al., 2023. JXB

<https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad481/7491230>

**Action:**

* Read the paper
* <https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad481/7491230>

====

**Stresses- soil compact**

* Soil compaction – root resistance
* Ethylene sensitivity as an indicator
* Lab experiment – root length response to Ethylene treatment

Three groups, insensitive, moderate, sensitive

validation in the field using two groups, sensitive and insensitivity

* GWAS, 7A, four candidate genes

Using mutants to identify the candidate gene, Catalase

* Single-cell RNA sequencing , gel-based

**Comments: drought and soil compact:** drought stress will cause the soil to compact. So, drought stress can be dissected into different factors, one of which is compact soil.

====

**Synthetic wheat**

Liu 2021

* Population, 21 synthetic, backcrossing with two recurrent parents and RIL
* Skim sequencing
* 36 traits in 17 environments
* Rht-B1 on tiller number
* Glu-B3 on dough strength.

Action:

* Find the paper, Liu et al., 2021

=====

G2B

* **What questions at Darwin and Mendel’s time have not been answered?**

Gene is one direction

How about other directions?

* How to use the knowledge in breeding.

20240113 Friday

IWGSC

Wheatxmaize hybridization

Shaobin Zhong

* Cross- transient zygote – elimination- DH
* Transgenic wheat (CRISPR) x wheat, edit wheat genome.
* **Tsn1** confers susceptibility to tan spot and sensitivity to the fungal toxin ToxA
* **TaMLO** makes wheat susceptible to wheat powdery mildew

Three genes, target the conserved sequence

When three gene mutated, showing resistance

* **TaHRC-s** FOR SUSCEPTIBILITY to FHB

Three target region

* TaPFT and TaHRC/His

***Evaluation of the mutations in the greenhouse***

Providing more evidence about the causal gene.

* ***Target gene insertion is challenging in wheat, this hybridization method has advantage***

Cre recombinase – insert two lox sites Lox66 Lox2272 to insert resistant genes

Construction: Lox – Cre – target gene - Lox

Fhb7

Tsn1

* Using T0 and T1 as the pollen donor, but T1 maize pollen gave higher efficiency
* Advantage

1. Efficient maize transgenesis
2. Can cross with any wheat varieties for editing
3. DH wheat and embryorescue is easy

Parking plot

* Efficiency

=====

* 53,000 genes (about 50% of the HC genes) are corresponding genes or homeologs -- triads
* 70% of triads display balanced expression, not homeolog expression bias (HEB).
* Paragon x charger, Paragon x watkins

6 generations of SSD,

* Bias distance – relative distance between f6 lines and parental HEB

From elite and landrace cross, observed more difference between F6 progeny and parent

By elite x elite one, less

* eQTL analysis
* HEB is partially linked with genotype.

Related with heterosis.

parking lot:

* what stage of the RNA samples?
* Specific examples?
* Related to heterosis of the progeny?

====

**Small RNAs AND bIG IMPACT**

miRNA on wheat spike development

Carpenter (Anna co-author)

* miRNA repress genes via translational repression and mRNA degradation
* binding sites have outsized phenotypic effects
* e.g., miR166, paired gene, HB-D2, Dixon et al., 2022

pair

G:U pair and

Mismatch

Mutant, ps1, and ps2 had mismatches

e.g., AP2-5A, Q gene, targeted by miRNA172

e.g., AP2-2 miR172

* AP2-5, AP2-2 have opposite expression patterns during wheat spike development

Why? The same miRNA172?

* There are three copies of miR172

From smallRNA-Seq data,

A, b, c, have 7 3, and 1 loci, respectively

C low expression

* miR172a com-pair to AP2-5, but miR172b pair to AP2-2

expression pattern confirmed this result

* three hypothesis:

1. both AP2 and miRNA expressed in the same location
2. different target
3. transcription different

parking lot:

* how we use these results in regulating spike development?

====

Junli Zhang

**Promoter capture sequencing of wheat TILLING population**

**New promoter capture probes**

* exome capture, PNAS 2017

119 Mb 4x

162 Mb 6x

Mapped into v1.1, obtaining more data

* promoter capture

e.g., qSH1 shattering

method from Anthony Hall, 2kp promoter

* 4.3 M promoter mutations, 4.7M coding region
* Problem, promoter region is not well annotated, solution, **conserved regions**

***Using 50 bp of the first exon --- tip***

ATAC-seq???

* Regulatory capture

Old: Gardiner promoter capture

Enrichment efficiency is low

Bias, repetitive region

Only 2kb,

* ATAC-seq revealed more regulatory regions

New: arbor (biosciences) capture

Reference genome + ATAC + tetraploid

How to design the capture probe???

When capturing, duplication regions

====

**Durum wheat genomic sources**

* Tetraploid wheat global collection (TGC)
* Global durum panel 1056 ACCESSIONS.
* **2023 platinum quality assembly**

Pacbio \_ Hi-c + BioNano

10.4 Gb N50, 100 Mb

* Annotation

Illumina + ISO seq --- developmental stage + stress

HC 68,154 genes

* Projection of know durum wheat QTL, QTLome

5,651 QTLs/MTAs

6A, for disease resistance, LR, YR SR FUSARIUM POWDERY MILDEW.

* Collinearity between Svevo and Chinese Spring

B, similar

A complicated reversion

* New assembly, case advantage

Free threshing Tg1-B, 2B

Abstent in v1, but in the platinum assembly, a reversion between Zavitan and Svevo at the region.

* Pangenome

40 accession, 12 cultivar, 10 landrace 7 domesticated, 8 wild, and timopheevi and Urartu 2+1

Observed a reversion in 1A in several accessions

* How to match the annotation between v1 and platinum

====

Kambona

**Memory of past drought stress exposure effects**

* Drought stress limits stomatal conductance, loss of assimilation, production of ROS, affecting photosynthesis
* 40% of crop loss
* Produce and transmit ***imprints*** to the offspring
* One generation,

Intergeneration stress memory

Transgenerational stress memory

* 200 cultivars 1948-2013 predominantly from germmany

Published in 2023

Three generations, 1, 2, 3 --- 1 grandparent, 3 the latest offspring

Parking lot:

* How can you remove the genotype effect? Since for different cohort, you have different genotypes – A: selfing, the same genotype
* Drought treatment- stage and severity
* The interaction between microbe and plants? Since the same shelter every year.
* Repeat and validate the observation in different genotypes and environments

=====

**Jorge**

* FT2 and bZIPC1 both together regulate the number of spikelet, need to select both together
* WAPO-A1 and LAFY at cooperateively to control SNS in wheat, need to select both together
* LFY and Squamosa

It is mainly about gene interaction and function, mutations on spike development.

It is not easy to follow the logic.

Deep learning curve to follow.

====

**Regulatory network in tetraploid wheat**

* Gene annotation – transcriptional profiling – developmental and stress

Co-expression analysis

ATAC-sequencing – determining chromatin accessibility across the genome, uncover how chromatin packaging and other factors affect gene expression

* Resources of genes response to the stress, heat (not sure drought)

Parking lot:

* Data available in public database? – generating a browser and database to be searchable
* Must can be done in a controlled environment
* Source, pool, and flow

**How to work on flow?**

Flow is energy cost?!

How about the study in Arabidopsis and rice?

https://www.sciencedirect.com/science/article/abs/pii/S1360138515001995

* **Professor** is the one who knows everything and all details and big picture, rather than only the big pictures. This is the only reason for the continuous success!

====

**Barley**

* Plant has little divergent unstable transcripts, typical for animal transcrition

Plant lack CTCF (CCCTC binding factor\_

Plant lack NELF (

LACK CPF islands

Why????

* Barley , embryo
* Using cap analysis of gene expression (CAGE) TO IDENTIFY BARLEY PROMOTERS
* Promoter-promoter interaction 65%, also some interaction between promoter and enhancer

====

**Cap-Binding complex in plants**

* Mutants tolerant to droughts in potato, Arabidopsis
* CBP20 mutant in barley and Arabidopsis, drought tolerance

Higher wax

Lower electrolyte leakage

Faster stomatal closure

Plasticity of epidermal pattern

* Materials, Sebastian, cbp20ab, cbp80b, and double mutant

Seed germination – ABA treatment

Double mutant germinates similar with wild type

* Transcriptome, ISO-seq and RNA-seq

====

**H. bulbosum, wild species of barley**

* Provide disease resistance from introgression
* Outcrossing
* Diploid and autotetraploid
* Introgression provides barley yellow dwarf virus

Parking lot:

* Educate about how to give a **presentation**
* How to talk about the topics with different audience
* How to adjust the language
* Give the **opportunity** to practice at all the discussion sections!

====

**Pangenome and transcriptome data analysis**

* 2020 paper showed that CS is not a typical wheat and different from other wheat varieties a lot
* Synthetic intorgressions, regions can not be mapped to CS

**Statistical genomics**

====

GnoBaits probeobased

Parking lot;’

* What is the difference from other genotyping platform

Paper published in Plant communication, Development of high-resolution multiple SNPO arrays for genetic analysis and molecular breeding through genotyping by target sequencing and liquid chip

====

Multi-variant GS

* Bivariate GS

BLUP-HAT for evaluate the GS accuracy

* Combine one trait with each of the other traits to predict the target trait (one trait)

The predictive ability is higher than the prediction of individual traits

* Might also help to improve the predictive ability of traits with low heritability

Parking lot:

* How about the correlation between traits? Using the combination with all the traits, so the correlation does not affect it.
* BLUP-HAT is required for the computation efficiency

====

**GxE interaction across large-scale field trials**

Dan Runcie

* Extreme example, highland and lowland maize, perform different in the other environments
* Different ranking,

Plasticity- performance difference in different env

* Two ways of measuring plasticity

1. Mutant
2. GWAS – difference across locations – lines replicated by alleles

* How about the untested environments?

Need to know how genetic factors respond to environmental factors!!!

For **predictable GXE**, rather than unpredictable GXE

Need lots of trials for the env. Factors

* 3733 lines, 32 trials, and 13 locations

Fitness related traits

* **Marker by environment effect** – is the key component,

**d**ifficult, false positive

1. within trials: genetic background effect
2. across trial: genetic background effect
3. covariance of genetic background
4. covariance of replicated line

* Solution MegaLMM

*Can be used to study drought, heat response loci*

Environment factors, elevation

Fit mgcv/gamm4 R package

* JointGWAS is coming.

Parking lot:

* How about the other factors?

====

**Genetic correlation**

Shizhong Xu

* Pearson’s correlation
* Genetic correlation

1. Pleiotropy - permanent
2. Linkage- temporary

* What is Delta method to estimate standard error of the estimation or prediction?

Jim Holland has the classic paper about genetic correlation in SAS

=====

International weed genome consortium (IWGC) PanOat

**Oat**

=====

**Genetic diversity of Nordic oats**

Lund

* Nordic oat, Avena sativa

Breeding for >100 years

Good quality

Varieties used by USA breeders

* 764 accessions

80 landrace

5 wild

350 cultivars

Others

----- 59 duplications, 99 duplicated with global GBS study

2022 in the field for phenotyping

------ Sang sequenced

Added 35 new cultivars

GBS for genotyping

* Mapped markers to chromosomes

PCA showing large chunk clustered together

Keep 425 diversity accessions

Landrace widely scattered, similarly old cultivars, and new cultivars lost the diversity and clustered together with most of the accessions

* Four groups:

69 landraces

1940 121

1940-1979, 102

New cultivar, 90, after 1980

But, cluster analysis showing 7 clusters

* We need to be cautious about the PCA results

**Parking lot:**

* Any specific traits driving the diversity loss?
* Oat is outcrossing? Why rate of heterozygosity?

No, oat is self-pollinated.

* Chromosomes contribute differently to the population diversity!

====

**Global (Genomic) Oat Diversity (G.O. D)**

Wubishet.Bekele

* 4.5 K accessions

There is a spring block, spring type from different breeding programs

* Increased accessions to 9,000 accessions
* Again, Sang is the reference genome
* 20,000 bi allelic SNP
* //wubi.shinyapps.io/2024-01-11-Avena/
* //wubi.shinyapps.io/2024-01-111-Sativa/
* Divided into 21 groups based on PCA analysis

Explained by the origin source

* Translocation, 1A-1C, and two insertion

Question

* If you remove 20% of the US accession, the same results?
* If we remove LD markers, the same results
* the clustering is stable?
* GBS data is available; We need the raw data to call SNP if also using GBS markers

=====

**Hexaploid Oat Pangenome**

IPK

Martin Mascher’s group

* Resources
* 31 accessions

**24 SATIVA**

LONGIGLUMIS

INSULARIS

STERILIS

BYZANTINE

FATUA

SYNTHETIC

* Sequencing

PacbIO + HiC

12 SMRT cells

10.8 Gb

N50, 23-273 MB

* Chip-seq to localize the centromere
* Gene annotation

RNASeq 6 tissues, 3 reps, of all 24 accessions

125,000 genes

* Some translocations, inversions

e.g., 2A-2C translocation from radiation mutation, probably, in a dwarfing gene

* K-mer GWAS, reference free

7A, 7D for heading time

There is a break point on 7D,

Also saw methylation nearby the break point

Checked the expression of genes in the region

There is one flower gene, but need further investigation

* ***Challenging for the graph-based Pangenome***

Difficult to use, still use one reference genome

Is there pan-genome marker-based GWAS?

WheatCAP – 20240114

Jorge – start

* Shuyu will move to the breeder position
* Jim is looking for a PhD student
* Report submitted before the end of 2023
* Variety release + students trained – two key indicators

Comments: happy with what we are doing

huge needs of breeders

Connect T3 with UAS hub

* 17,800 samples genotyped
* 249 trials, 75,000 plots
* 41 students currently
* 2023

Four workshops or short course:

WheatCAP student presentation at PAG + GS + T3 (hours of training and hands-on) + UAS

Including some students not involved in WheatCAP + discussions

* Integrate ourselves in a large team showing the advantages

Parking lot:

* Actually use T3 in breeding practice, store the data and analyze the data
* Training given by a breeding program with hand-on experience.

Trainees are students and field technician

Good documentation about T3 for new users

* ?? any training required??
* GS training

Use real data and R script, which can be used in practice

NCSU support the workshop without space charge

Jim Holland is impressed about the network and collaboration.

UAS

Juan lANDIVAR – Center director

* Problem: filed trial layout, breeding program will do it.
* Field size, 5 acre or/and more
* Data flow is critical for the implementation
* Cloud based data portal, with Oracle

1. Data storage
2. Data processing
3. Communication

150TB data

* Process

1. Data upload
2. Rgb,
3. Curation – quality of the data, trial layout, grids quality
4. Features from the images
5. What to do use the data

* Steps

1. Plot boundary and grid structure – quality
2. Canopy height
3. Canopy cover
4. Canopy volume
5. Time-course developmental of all the feature

Results: Good correlation for plant height

* Growth curve monitoring

Convert the growth curve into a bell shape plot to extract the information to monitor the growth rate

Experiment:

Early flowering, ***long filling stage before mature*** – longer reproductive period

Based on the monitoring of plant height to link the growth rate with yield, and monitor the grain filling period

Duration + the slope of the bell curve

***The problem in California*** is the frost which impact the spike development.

***Early planting might help.*** but the frost might be an issue whether the early planting fit in the cropping cycle

* Convert the data into yield prediction –models

**Jose Landivar – go to person on UAS**

Reporting the survey of the workshop

* May 2023, 25 graduate student
* Datashape (USDA can not use it) alternative Pix4DMapper, software
* Share the data ahead and training materials
* Discussion to share experiences, share knowledge
* Food: Gluten free option, vegetarian
* ***QGIS plugin*** development- connect T3 to get the trial layout information, plot boundary.

Detect the plot boundary automatically.

* Upload the data right away after collection; if there is any issue when analyzing the data, feedback will be given to improve the data collection.
* Data requires **precise georeferencing**
* Ensure giving credits to the UAS hub team

Parking lot:

* Satellite imaging, needs calibration
* When should we stop using the ground data? How about ground data to validate the prediction??
* Acreage cost? Discuss with Jose about the cost.
* ***Different pipelines get the same results? Any cross checking?***
* Outsource the UAS service and get the data back late and upload the data together late,

What data is saved in T3?

* ***Still a question***, how to make the most use of the data from images using deep learning and machine learning. Follow up with the new development in image analysis and feature capturing using AI.

====

**Genomics information**

* Sequencing

10 PacBio

50, 10xcoverage

200 lines 0.5-1 coverage

* PHG data mapped to the pangenome version
* Wild species

H. villosa

Ae. Ventricose

Ae. Geniculate

Et al.

* Correct reference bias in RNA-seq studies using pangenome

Team with JIC to develop strategy to deal with the bias

Be cautious on the bias when interpreting the results

* Now working on the v3 PHG

V1, Jordan, 2020.

* PHG

V2

RefSeq 2.1

5 M SNP after MAF 0,01, 1.5 M markers

47% A, 24% B, and 19% D markers

In total, 500 accessions

* Imputation

Fastq or vcf file for imputation (including relevant germplasm; select the relevant founder for imputation)

Using haplotype rather than SNP for prediction and GWAS

**Imputation** – related to MAF

* There is also evolvement in the software in developing PHG

**Action:**

* **Pangenome**, know the term, but do not know the detail
* **PHG** – plant and human, how it can be used
* **Imputation** – how to evaluate the accuracy.

====

Genotyping

* Mid-density genotyping

***Agriseq ThermoFisher***, including QTL markers, 32-3400 SNP, more than 100 markers per Chr

$10 per sample

$2.4 before genotyping

Aplicon sequencing

More information about the marker panel

How it works? case study?

Some MAS markers do not converge well in the Agreseq

* Illumina, 3k $14 per sample. Wheat oat barley, 1+1 way to mix samples.

1,500 samples, from tissues, per week

* CIMMYT

DarTag, $12 per samples including DNA extraction

PHG helps to impute haplotype information

* Advantage

MAS data in the platform

***MAS to haplotype-assisted selection – determine the favorable haplotype.***

* Breeding program pays for the sequencing and USDA genotyping center support the sample preparation.

**Action:**

* What we need to develop in the breeding program?

Increase the breeding efficiency?

* Write down and extend the ideas or thoughts I have for future communication.

====

**Lunch discussion**

* How to use the shared genotypic and phenotypic data

Using the GS to predict the performance across breeding programs enhancing germplasm sharing

Prediction combining crop modeling

Develop tools for parent selection based on optimal

***Whatever tools we need*** in breeding should come up as a proposal for development or collaboration!

***Mate selection:*** predict crosses

* How to develop the next WheatCAP project

Regional?

* About starting a new position

Know the people who is efficient

Know the procedure

Find the student

====

Jean-Luc

* **WheatCAP**

52,500 accessions

27,493 accessions genotyped

* **Protocol for**

Sample name

Plate name

Accession name

Project name

Trial name

Location name

Stage name

----- double check with Josh and establish the system

* Known informative markers - ***MAS***

As genetic character feature, need to update and marker updated usually

Select lines based on their markers

Upload data of MAS information

Marker can be ontology of trait!!!! As the way we manage phenotypic data

Assay as the ontology

* GXE analysis based on multiple trials with unbalanced design, and predict the performance in untested environments.
* How to manage the genotypic and phenotypic data?

We need to manage ALL the data generated in database

* How the genotypic data is managed?

By trial or plot?

====

**Student presentation**

Rayasa with Jessica

* Harvest is a bottleneck

High-throughput phenotyping + machine learning

* Three location, two seasons
* 2415 entries

177 common entries across env

* Phenotypic data imputation by combining drone data and read data with missing data

Parking lot:

* Market
* Breeding scheme
* What book are you reading now?
* How the trials were designed? P-rep design
* Data quality is heritability?
* Think about the reality and how to implement it?
* Explain the methods well or visualize it

====

Xiaoting Xu

H35 for Hessian fly in overland (a winter wheat)

* H7A, H36; 3B, H35
* Sequenced the resistant variety overland using PacBio
* 38Kb
* Compare overland with Ch,

Overland has insertion, with a R300 gene and B120, candidate gene

* H13 is more resistant than H35

Hassian fly overcame H13, the biotype might overcome it?

* Insertion is only in a French variety, but all other varieties does not have the insertion

Parking lot:

* How can we link QTL mapping project with breeding?
* Train students on QTL and also breeding

===

Kyle Parker

* Texas diverse landscape
* Focused on target environments, understand the environment, the critical factors
* Ask the correct question?!!

Parking lot:

* Factors affecting yield, not only climate data

Also, insect, soil type, disease pressure, and management differences, and other factors

====

* Dump the database or data into the GPT and figure out what can happen.
* How to use all the data stored in the database

Data curation is a big issue.

====

**Livestock and the food systems challenges in the global south: the role of genomics**

ILRI

University of Edinburgh/Roslin, sruc

* HungerMap, global insights and key trends
* Three options:

1. Produce more in North
2. Transform smallholder livestock systems
3. Improve livestock industrial production know how

------- This is a good way to present, providing the options and focus on one

* US dairy population getting low, but milk yield increased

Example of improvement

====

**Steven Xu**

Insect-resistant gene also increase yield 15% in winter wheat and 5 in spring wheat

Salt resistant gene in wheat, increase yield, called magec wheat.

He has connections with Sheng Luan (department chair in UC Berkley), Jianli Chen- wheat breeder in Idaho, Mingcheng Luo – UC Davis; Yongqiang Gu – USDA; Shuyu Liu, Guorong Zhang,

Wheat group

Academic group

====

Structure variation is longer in heterozygosity; affecting deleterious mutations?

Large structure variation are strongly deleterious

One reference paper:

<https://www.annualreviews.org/doi/full/10.1146/annurev-animal-080522-093311>

====

**Genomic selection of synthetic hexaploidy wheat**

Tetraploid durum x diploid Aegilops

Tan sport

Septoria nodorum blotch

spot blotch

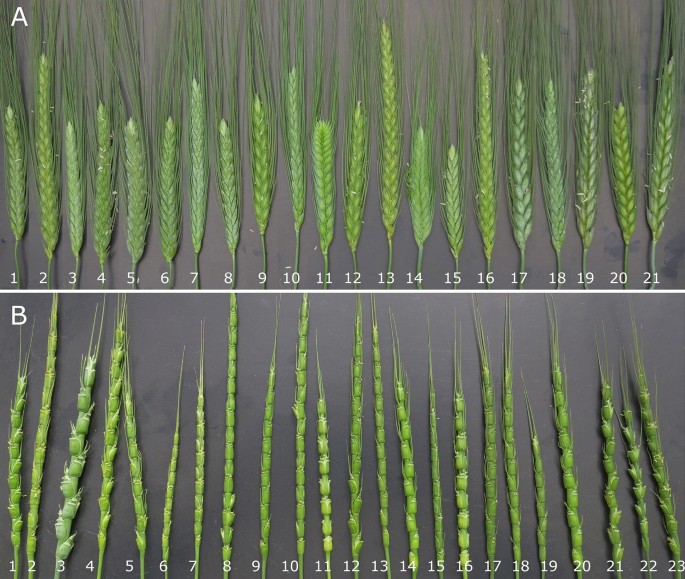
GS to predict these three diseases!

Controlled environments for disease screening.

GCA prediction! Major gene control it?

How about the yield? And grains?

**Mutation** to introduce Q gene favorable allele?



====

IRRI

Salinity stress Rice

* GS scheme, Send part of the breeding population to NARES and train the model to predict performance or GEBV of the remaining of the breeding population
* Sparse design for NARES yield or phonotypic data collection

==== go to person on bioinformatics ==

**Viswanathan Satheesh, Iowa State University**

Tools for variant calling

* GATK Haplotype caller
* Bioinformatics workbook

Organize folders following the data analysis steps – **good practice**

* Nextflow
* Data analysis should be a smooth flow, **step-by-step**!
* Software dependence is an issue in data analysis – well noted about the R packages and software used and their version.
* -nf-core/sarek, navar designed for human and widely used
* PI, Matthew Hufford

====

20240116 PAG

Katrien Devos

Ensuring food security in a changing climate

* Millets
* Tolerance to climate stress, less researched and high improvement potential
* Challenges: Commercialization, tools development, community establishement
* **Millets belong to small grains cereal!**
* High calcium in Millets
* “poor man’s crop” and poor management of farming practices, breeding method is old and inefficient
* Take advantage of genomics tools for finger millet
* Finger illet,
* Eleusine indica x BB – 1.3 Mya ----- AABB
* Ethiopia domesitacation

1. BC, to India 3000 BC—seed shattering , Sabs-B1b, A1c + A1b, so two step process
2. B1b, and 2) A1c and A1b in the southc at two locations

* PCA: India is different from African population

Population structure is driven by the wild segments

* Trai mapping using RIL populatins

GBS A and B colinear, >80% genes are duplicated, young polyploid

Ususlly need two copied for the function or phenotype showing up.

B genome is larger than A genome, due to the more duplications

* Trai mapping

Shattering, seed size, purple anthers, blast leaf blast plant height

Sabs1, shattering

* Bast, leaf, head, stem

Rice, wheat pearl millet also have blast

Different strain from rice, infect rice but not finger millet

* Blast, isolate, into toal 226 isolates in east Africa

Illunima and Pacbio

Comparethe blast isolate between crops

14,060 genes in blast isolates

Population structure, two groups: kenya and Uganda, Ethiopian and Tanzania

* Rear to see immune

Population 1 has higher infection level

* Mining the effecter genes in isoltate, population 1 and 2, also compare twheat and rice, unique isolate

Effector genes in blast can determine host specificity

* JA response gene – regulation about the blast resistance
* Interaction: Y2H, co-, transit expression
* NS, Bio-innovate Africa,

Parking lot:

* Shattering
* Yield gain, data to show the benefits or advantage?
* How to tell stories can be accepted by broad audience

=====

**iNTEGRATIVE MODELING** FOR THE DEVELOPMENT OF IN SILICO CROPS

Amy Marshall-Colon

* 2023 plant hardiness zones updated
* New strategy to prevent yield loss caused by the temp increase and variation
* Opensimroot model low root cortical aerenchyma, 2011 paper

High RCA increases yield.

* Light, temp, co2,

ROOT GROWHT, MOLECULAR TRANSPORT, GENE REGULATION NETWORKS

Integrate the modules to uncover the black box and understand yield

* Crops in cilico,
* Computational framework

1. Programming languages- challenge, can not communicate directly with one another

Solution – connect the input and output, rather than reprogramming the model

Develop Yggdrasil to integrate

* Virtual crop models

Proof-of-concept, model soybean response to CO2

FACE experiment – free air co2 enrichment

1. Pure model – **leaf-level photosynthesis**
2. **Metabolite model**----- physiology factors
3. **Transcriptome model**. -----protein

* Paper , 2019, in silico plants, Kannan et al.

Gene expression network and starch synthesis + photosynthesis +

* **Soybean BioCro**, Matthews, et a., In silico plants, 4(1), 2022.- crop growth model
* **3D canopy model** –
* Community science

Annual symposium hackathon

Journal, in silico plants – about plant modeling, Steve Long

Jonathan lynch

James Schnabel

Parking lot:

* What is RCA – ***Root Cortical Aerenchyma***
* How about interactions between env. Factors?
* Genotypic difference or variation? Or general trends of the cro?
* How less we know about the physiology of plants when they grow in the field!!!!
* Quantity and quality

Easy to quantify but lots of details on quality

We also can qualify the pa

====

**CassavaBase**

====

BreedBase

* BTI independent unit
* Genotypic data organized by plates?
* **SimSearch tool** – search the same genotypes based on genotypic data within the database (the same genotyping platform)

====

Inclusive digital ecosystem:

Juan David Arbelaez Velez

* 2M acre oat in US
* 5 programs active in developing varieties
* Oat breeding, pedigree selection selfing crop
* DiGGer – Coombes, to design trials, has a private version for larger dataset
* ***Seed inventory and management***

Create seed lots

Genotypic data, multi-species chip (Fiedler)

* ***Plant Breeding Coordinating Committee Germplasm Transitions initiative*** – the transition between the generation of breeders and breeding programs.

Parking lot:

* How to use seed inventory

====

* Link the data to a protocol

For metabolomic protocol

Protenomic

NIR

Parking lot:

* In wheat, the NIRS model is ready.

Should we upload the spectra into the database?

* Problem in handling the prediction from different years. We predict using new models and have the new prediction using the historical spectra, which means we have the updated prediction value for the old samples. How we can handle this issue?

===

Isaak

* Image analysis

Joyce

Parking lot:

* Always like a ***flow*** and showing the ***logistic*** of the activities
* **Case studies and demonstrations** are the best way to show how a tool works!
* **Easy access and understanding** are another important factors for tool adaptation

====

Robbin

Imagebreed Py

* GXE, multi-dimension
* Remote sensing data to monitor the crop throughout the season
* Challenges

1. Easy to collect data – but extracting feature and integrate with genomic data
2. Nicolas morales – soluson

* ImageBreed

Assign grid once and used in multiple times

Parking lot:

* ***How to make it sustainable in a team*** in University for production and continuous improvement
* Jupyter analysis environment

Action:

* Learn Jupyter and know about Docker
* Phone based image, how to analyze?
* UAV data in different programs to combine them together to build up a robust model for implementation!!!

====

Spike development and yield

Adam William Schoen

* Yield per spike, ***multi-grain*** and single grain, 1.9 x of yield per spikelet of the NIL at F3 families
* Background of the germplasm: from 1970s, introgression?

====

Interaction between suppressors and rust resistance genes

Jyoti Saini sHARMA

* RESISSTANCE GENES – DETECT THE PATHOGEN AND ACTIVATE THE DEFENSE RESPONSE
* SUPPRESSOR PREVENT LIMIT EXPRESSION OF RESISTANCE GENES
* CANTHATCH, mutant, removes the gene showing resistance and also removes the D genome, showing resistance
* Remove a suppressor, and activate other resistance genes.

====

James Milson

Hybrid wheat – quality

Philippa Borrill group

* Glu-1 and Glu-3, protein + grain characteristics
* Quality gene in heterozygous status? Do parents give high-quality hybrid?
* 1) balance effect, additive 2) parentally biased effect, dominant?
* 105 female and 61 male with 227 hybrids produced- genotype and phenotype (quality), 11 traits
* Falling number with amylase activity
* In heterozygous status, male parents did not affect protein content and moisture content

Parking lot:

* ***Not single factor analysis***, need to consider the effect of the background difference and uneven distributed of other factors in the background.

***His hypothesis*** is that the other factors evenly distributed in the four groups, homo, heter1, heter2 and homo.

====

XIanran Li,

***BRIDAGEcereal***. – indel in pangenome

* Half of the causal polymorphisms are large indels
* Eliminate the same sequence, the left are indel
* How to condense the information from multiple pan genome assembly

Too much compositions!

Using the slope between the two sequence to identify indel for certain genes

* B1, antless gene
* 4A, TaDL,
* Paper published in Molecular Plant
* How about new assemblies?

How about the annotation of the new assemblies?

* Limitation – intertion+ del; copy number,

====

Meriem Aoun, OKSTATE

* National association of wheat growers
* Stripe rust and leaf rust

Lr34/yr18, lr46/yr29, yr36, adult resistant genes

Yr5, 15, 37/yr17

* 459 us hard winter wheat

Parking lot:

* The key genes in the wheat breeding program, MAS survey!
* CO19D304R might have new resistant genes
* Many varieties have adult resistant genes, but few varieties have seedling resistant gene/all stage resistant geen, different from race-specific gene

====

VRN2, climate adaptation

Dominique

* Dixon et al., 2022, talk about vernalization
* Different from Arabidopsis, e.g., VRN2, repressor in the pathway
* ZCCCT1 and ZCCT2, three copies in ABD, and 6 genes
* Six copied function differentiation? Selection for specific allelic variations?

====

20240116

Cassava IITA-CIAT meeting

* **Germplasm sharing**

South, east, west, and central Africa breeders, global cassava breeding

**Sean**: 1) CBSD + CMD dual resistant germplasm from Vietnam, 2) workshops on flower-inducing and cooking quality.

**Peter**: Germplasm from Asia with a high starch, good plant type, and CMD + CBSD, good opportunity for introducing the traits to diversify the African breeding germplasm.

**Note**: the same clone names should be used (with Stephan Winter), so we can have the language.

**Sikiru**: we received the materials from Stephan and evaluated the CBSD resistance. The seeds we received from CIAT were evaluated, and I will send you the data and the pictures.

**Elizabeth**: how do we standardize the germplasm sharing?

**Ismail**: Willing to import germplasm to diversify the breeding population. We need to genotype the imported germplasm and understand the diversity.

**Form a germplasm-sharing committee**: enhance the germplasm sharing and information sharing. We are going to talk it later, in another meeting.

Shared location to run crossing nursery: Hawaii, the only option, but slow.

**Kayondo:** concerns on witches’ broom. The spores can attach to the seeds. We need to treat the seeds.

**Peter:** at IITA, Larva will be responsible for germplasm importation. It is possible to introduce germplasm through West Africa.

**Sean:** There is no CMD, witches’ broom in the crossing nursery in Vietnam.

**Peter:** share the pedigree.

**Ismail:** not overload IITA capacity.

**Peter:** Do not lose the opportunity to introduce new germplasm.

**Xiaofei:** if we share tissue culture materials from Vietnam, we must share them through Stephan Winter.

**Elizabeth:** We need high-quality materials, e.g., full-sib families. Sean will share the pedigree, so IITA can choose the families IITA needs.

**Peter:** We need to use the materials that were tested in the field with solid data to support the phenotype claims.

**Action: CBSD genotypic data**: Winnie will share the CBSD population data with CassavaBase once we have the VCF files of all samples from BGI

* **Workshop CtEH, budget 30,000 per workshop**

Gender-sensitive when organizing the workshop

**Action**: Elizabeth and Ismail will send the list of the trainees for the two workshops at CIAT before February 15.

**Flower-inducing workshop** – October

CIAT: on flower-inducing protocol

IITA: possible topics can be grafing,

NRCRI, NaCRRI, TARI,

**Quality workshop** – July

CIAT: on boiled cassava and starch

IITA: on processing produce

* **Monthly meeting of the CIAT and IITA meeting**

**Action**: Elizabeth will schedule a monthly meeting, including Gaby and Winnie.

* Global cassava breeding meeting in Nairobi

**Action**: Elizabeth will check with John about the date.

* **SOP process management harmonization**

June is a good time for Sandra to visit IITA

* **Genotyping platform**

**Kayondo**: shared the genotyping platform.

**Sean:** Validate the DarTag markers: CIAT timeline first quarter in March 2024.

**Action:** Kayondo will follow up on the timeline for validation with Sean and Gaby.

20240117

=====

Dirk Inze

VIB UGent center for plant systems biology

* DA module, proteolysis driven pathway

e.g., GW2=DA2

* KLU1/PLA1 encoding a cytochrome P450
* Angustifolia3/GRF module (AN3/GRF module) – larger leaves, dry weight
* Plant hormones

1. Ethylene, longer leaves, water deficit

* Translation fro lab to field is challenging!!!

Simmons et al., 2021,

Interaction of the growth regulator network

Combinational effects, many modules contribute to an aggregated phenotype

Lab and field, plant density, interaction with ever-changing environment

* Growth regulatory genes often work additive or synergistic

e.g., brassinosteroidids and gibberellins, on leaf size.

* highly interconnected, polygenic nature.
* Syper tansformation system with multiple gRNA to edit the connected pathways or networks
* CRISPR, INDELS ARE MOST COMON

Parking lot:

* Summary the literature and find the common rules and how GOD works
* Do not limit your potential and what you can do. As long as you are focused, you can surprise yourself very easilty
* Open an area and leave it to others and continue to open another new area! Rather than focus on the area and do not have others to come to the area.
* How he summarize the 8 module?
* Paper, BREEDIT, LORENZO ET AL., 2023

====

Anther

* Redox controls AR fate (rely on the air flow, and co-evolve with environment
* Dynamic cell origin, rather than linearity
* Hypoxia – environmental information
* -21nt phasiRNA and 24 phasiRNA
* 3000 small proteins and 300 have secretion signals in maize anther

Parking lot:

* Structure affect the function
* If you want to answer basic science question, need to be focused on controlled environment and focused on tiny tissue, rather than a plant
* How to frame the questions you are going to focus on in your career?
* Always have something else to focus on, not have to join the wagon of the other people. But, the question is how to find it???
* Scientific seminar is too boring! The presentor seems to talk with herself. It is hard for most audience there. Few people know about the topic and structure.
* How deep we can dig into the topics!

===

Seki, from Japan

Ethanol-mediated novel survival strategy against drought, high salinity and heat stresses and its application to cassava

RIKEN Center

* Flowering, CMD, Genome analysis, transformation, starch accumulation, and stress tolerance --- achievements
* Plant also need ethanol for stress tolerance – why?

Parking lot:

* Japan for project development
* Root rot produces ethanol that can help to enhance the tolerance.
* How to apply ethanol? Leaf or soil?
* Ethanol can go into the air; how can we control the concentration?
* Field trial is challenging! Not applicable at least in production
* However, it provided an angle for basic study to understand the stress.

===

Fola

Ground penetrating radar – early bulking

From Crop Phenomics

* Root – shovelomics, x-ray CT, minirhizotron, MRI, GPR, tools
* GPR, is a geophysical too

Map the dielectric properties

* Dielectric constants, cassava is 52, clay, 5-40
* Challenge is to extract the information; the solution is:
* GPRSTUDIO
* TIME DOMAIN APPROACH, FREQUNCY DOMAIN APPROACH
* Quality control, correcting noise
* How to extract the feature??? What data we collected? The type of data How we process the data?

Parking lot:

* Read English with Basil
* Affected by the soil type?
* Ready to use or not?

===

Lina

Whitefly transcriptome

* PER 335 (MR), 317, 415, 368 608, COL2246 (S), TME3? Resistant to whitefly
* Compare R and 1 S clones in
* Two different expression pattern:

1. 317, 335, 608,
2. 358, 415
3. ECU72

* Different expression genes are clone specific, few shared differenctial expressed genes
* In 317, upregulated genes

1. Phenylalanine biosynthesis
2. Flavonoid biosynthesis pathway
3. Tannins biosynthesis in 317, examples of core enzyme, transporters, and transcription factors
4. similar pattern on key genes 317 and 608

* also metabolomics data

correlation on tannin monomer and proanthocyanidins abundance

Parking lot:

* Reach out to search and provide collaboration opportunities

====

Kehan Zhao,

UC Davis

* Zhu et al., 2023, temperature and production, and cassava yield
* 334 accession, landraces collected from Colombia

~25x coverage

* PCA, African clones together (sample bias?)
* SA has higher LD decay in general.
* EnvGWAS, on mean annual temperature, significant loci
* Adaptive loss-of-function alleles

Contribute to tolerance

Focused on large effect loci

Using loss-of-function (how to define the protein structure variation) of status as genotypic data, in gene level. Found four loci and the candidate genes --- genes for adapting to cold environments

Then use CRISPR to validate the genes

Parking lot:

* Identify loci for cold tolerance, but rather heat tolerance
* EnvGWAS shared the your loss-of-function loci?

====

Sean

* The question is how polycrossing is efficient or not.
* How much percentage of selfing?
* How many male parents?

====

* Deleterious mutation on flower development
* Rate of selfing is still a questions

=====

Ismail

Metabolomics of yellow and wheat root cassava relationship with dry matter content

* Linkage or pleiotropy?
* 52 clones, 3 seasons, and 2 rep, freeze dry
* The whole pathway changed, and competition between carbon and carotene.

Tend to be pleiotropy.

* Further work is needed to have a sure answer!

Mutation, overexpression.

Starch function???