The Genomic Tools for Sweetpotato Improvement (GT4SP): Project Update

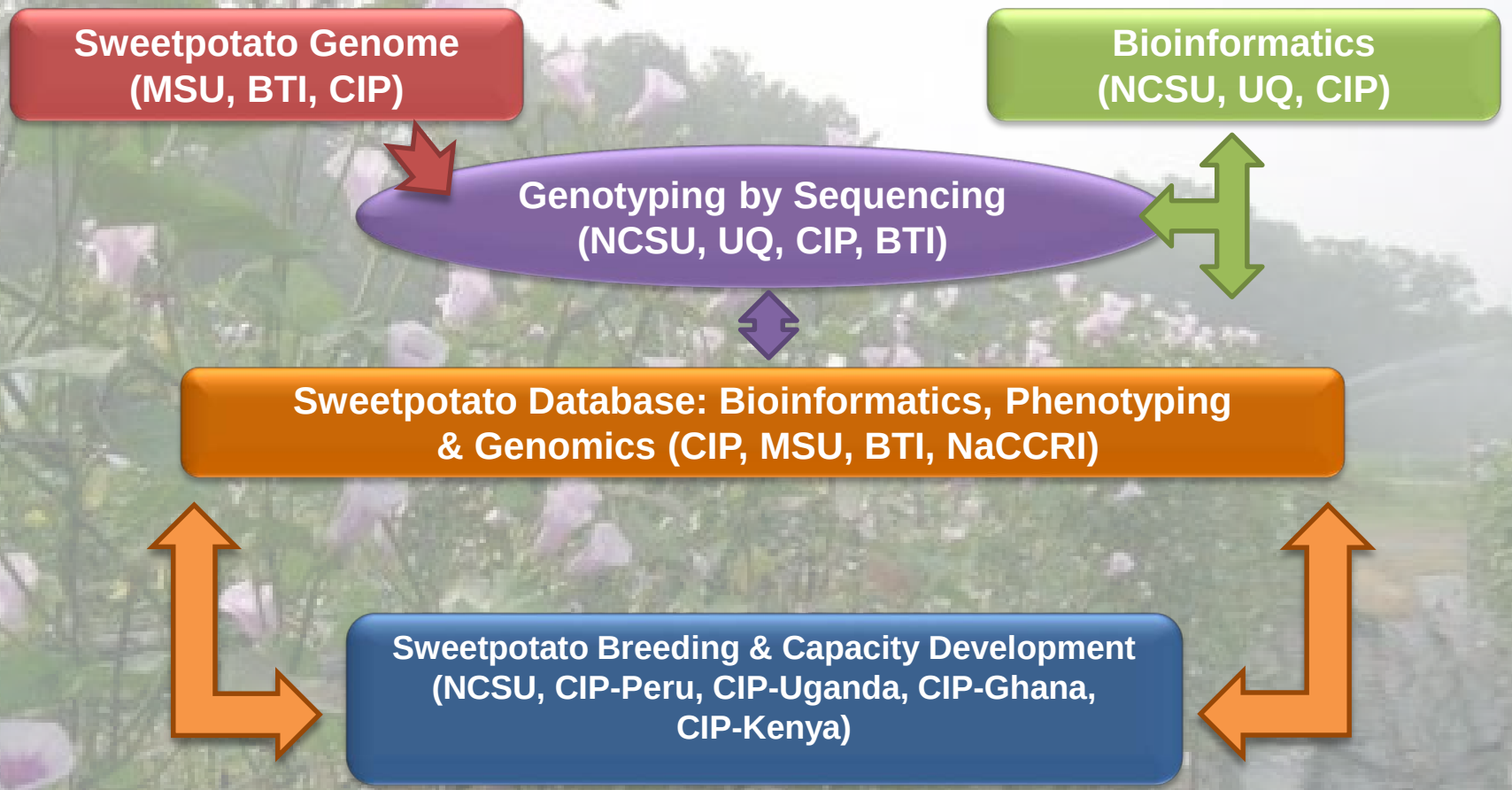
Developing next generation breeding tools for SSA sweetpotato breeders

Craig Yencho
Dept. of Horticultural Science
College of Agricultural and Life Sciences
North Carolina State University

SPHI Annual Meeting
Nairobi, Kenya
Sep. 24-27, 2018

GT4SP
GENOMIC TOOLS FOR SWEETPOTATO IMPROVEMENT

The Genomic Tools for Sweetpotato Improvement Project GT4SP



An ambitious project to sequence sweetpotato and develop modern breeding tools for a food crop that sustains million of people in SSA.

Collaborators: Boyce Thompson Institute at Cornell, Michigan State University, University of Queensland, Australia; The International Potato Center, Peru; BioSciences East and Central Africa, Kenya; National Crops Resources Research Institute, Uganda; Crops Research Institute, Ghana

NC STATE UNIVERSITY

Craig Yencho, Lead PI
Wendy Koch, Program Coordinator
Bode Olukolu, Molecular Breeding, GBS
Ken Pecota, Breeder
Jeremy Machacek, Breeder, Data capture
Xiaofei Zeng, Molecular Breeding, GS
Sharon Williamson, Research Specialist
Bonny Oloka, PhD Student, Uganda
Victor Amankwaah, PhD Student, Ghana

Zhao-Bang Zeng (CoPI) Statistics/QTL/GS
Guilherme Da Silva Pereira (Postdoc)
Marcelo Mollinari (Postdoc)
Dahlia Nielson (Bioinformatics)
Lina Quesada (coPI) Pathology

MICHIGAN STATE UNIVERSITY

Robin Buell (coPI) Genome Browser
John Hamilton, Bioinformatics Engineer
Jeongwoon Kim, Postdoc
Grant Godden, Postdoc
Krystle Wiegert-Rininger, Research Associate

BTI BOYCE THOMPSON INSTITUTE

Zhangjun Fei (coPI) Sequencing
Lukas Mueller (coPI) Database Development
Alex Obgona, Postdoc
Shan Wu, Postdoc
Alex Ogbonna, Database Development
Bryan Ellerbrock, Database Development

GT4SP
GENOMIC TOOLS FOR SWEETPOTATO IMPROVEMENT

CENTRO INTERNACIONAL DE LA PAPA
CIP
A member of the CGIAR Consortium
International Potato Center

biosciences eastern and central africa

CIP Lima
Dorcas Gemenet, Geneticist, Molecular Breeding
Wolfgang Gruneberg, SP Breeding, Global Lead
Merideth Bonierbale, Progam Leader
Jan Kreuze, Virologist
Reinhard Simon, Database Development
Raul Ezaguiere, Statistician

CIP SSA
Marc Ghislain, (coPI) Global Biotechnology Lead
Ted Carey, SP Breeding, Ghana, West Africa
Jolien Swanckaert, Postdoc, Ghana
Robert Mwanga, SP Breeding Uganda, East Africa
Reuben Tendo Ssali, Postdoc, Uganda
Luka Wanjohi, Database Development, Website
Mercy Kitavi, Molecular Breeding, Capacity Development

NaCRRI
NATIONAL CROPS RESOURCES RESEARCH INSTITUTE

Benard Yada, (coPI) Breeder
Gorrettie Ssemakula, Breeder
Milton Otema Anyanga, Entomologist

THE UNIVERSITY OF QUEENSLAND AUSTRALIA | IMB Institute for Molecular Bioscience

Lachlan Coin (coPI)
Chenxi Zhou, PhD Student, China

CSIR COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH GHANA

Marian Quain, Biotechnology, TC

NC STATE UNIVERSITY

A "Vision" for MAB Breeding in SSA

Breeding pipeline investments should include:

- **Genomic Resources –** ✓
  - **A reference genome**
  - **Marker development – we are way behind the curve….**
  - **A robust set of SNP markers and a low-cost genotyping platform**
  - **Advanced laboratory sequencing linked with developing country phenotyping and breeding activities**
  - **2x and 6x mapping, training and test populations**

- **Phenotyping, analytics and database resources –** ✓
  - **Improved phenotyping options**
  - **Web-based bioinformatic resources**
  - **New database, data collection and analysis resources**

- **Human Resources and Capacity Development –** ✓
  - **Continue to develop a dynamic team of breeders and allied disciplines**
  - **Training in the use of traditional and genomic breeding methods**
  - **Effective communication and collaboration**
  - **Multi-institutional training and capacity development**

- **A common vision and continuity of effort.** ✓

GT4SP – Science Update

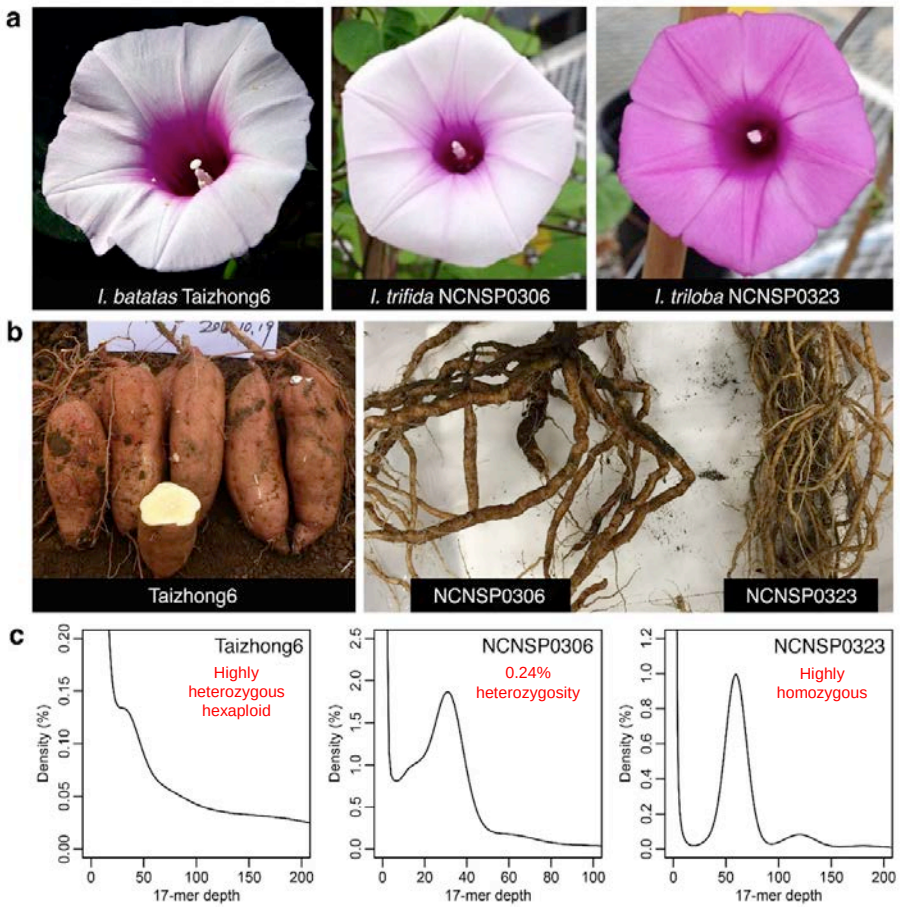
The cultivated sweetpotato genome

Cultivated sweetpotato is a highly heterozygous allo-auto-hexaploid (2n=6x=90) and has a large genome (1.6 Gb; Arumuganthan and Earle, 1991).

The origin of cultivated sweetpotato

Hypothesis I: Derived from the *I. trifida* autopolyploid complex (ranging from diploid to hexaploid) (Kobayashi, 1984).

Hypothesis II: Generated by natural hybridization between *I. trifida* and *I. triloba* (Austin, 1988).



"Taizhong6" was recently sequenced by Yang et al. 2017.

Fei Lab

BTI BOYCE THOMPSON INSTITUTE

The draft genomes of *I. trifida* and *I. triloba*

Filtered data generated from short reads (Illumina paired-end and mate-pair libraries) corresponded to an estimated coverage of **428×** and **291×** for the genomes of *I. trifida* and *I. triloba*, respectively.

Long read data (PacBio; 11× and 5× for *I. trifida* and *I. triloba*, respectively) were used to **fill and reduce gaps**.

De novo-assembled genome maps were produced via single-molecule technologies (BioNano) to **form longer scaffolds** and **detect mis-scaffolding**.

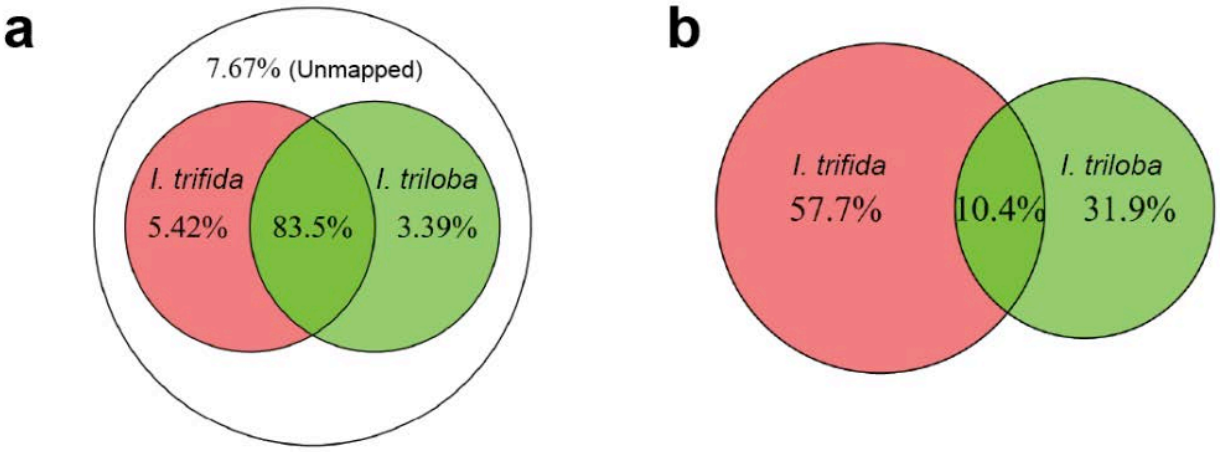
	I. trifida		*I. triloba*	
	Contigs	Scaffolds	Contigs	Scaffolds
Number	44,847	30,398	31,279	4,015
N50 (bp)	65,820	1,237,020	36,931	6,861,300
Longest (bp)	1,067,799	8,902,984	313,171	19,833,707
Size (bp)	433,252,193	462,000,517 **87.7%**	437,557,497	457,842,941 **92.3%**

of the estimated genome sizes
(526.5 Mb and 495.9 Mb for *I. trifida* and *I. triloba*, respectively.)

Fei Lab

BTI BOYCE THOMPSON INSTITUTE

Utility of *I. trifida* and *I. triloba* genomes as references for hexaploidy *I. batatas* sweetpotato



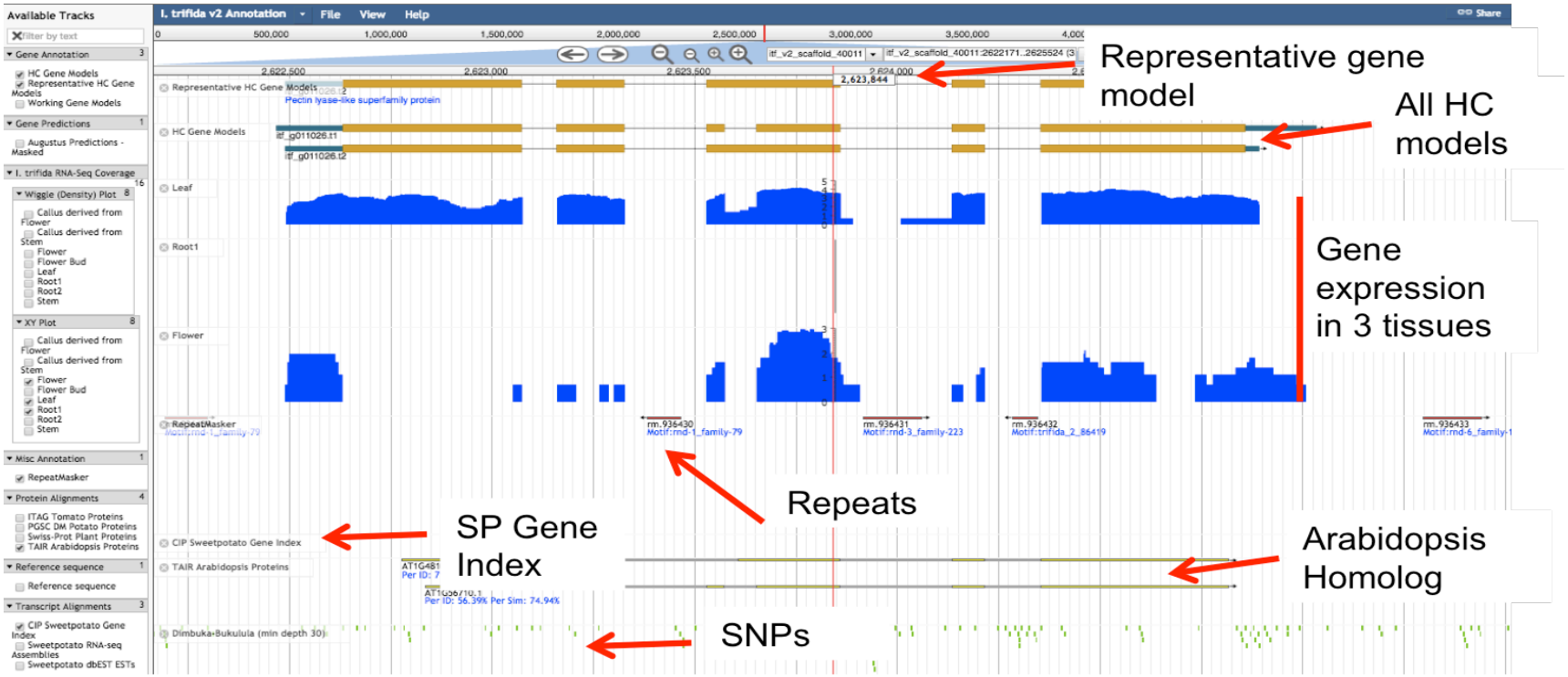
- ~ 60× **whole genome sequence data of** the 6x **African landrace 'Tanzania'** were generated using the 10x Genomics Chromium system and linked reads **aligned to the *I. trifida* and *I. triloba* genome assemblies**.
- ~ **83.5% of the reads could be aligned to both assemblies**, 5.4% and 3.4% of the reads aligned solely to the *I. trifida* and *I. triloba* genome assemblies, respectively, while ~7.7% of the reads did not align.

- Based on alignment scores, **~57.7% of the reads aligned better to the *I. trifida*** genome than the *I. triloba* genome, while **~31.9% of the reads aligned better to *I. triloba*** .

Coin Lab - UQ

The University Of Queensland AUSTRALIA | IMB Institute for Molecular Bioscience

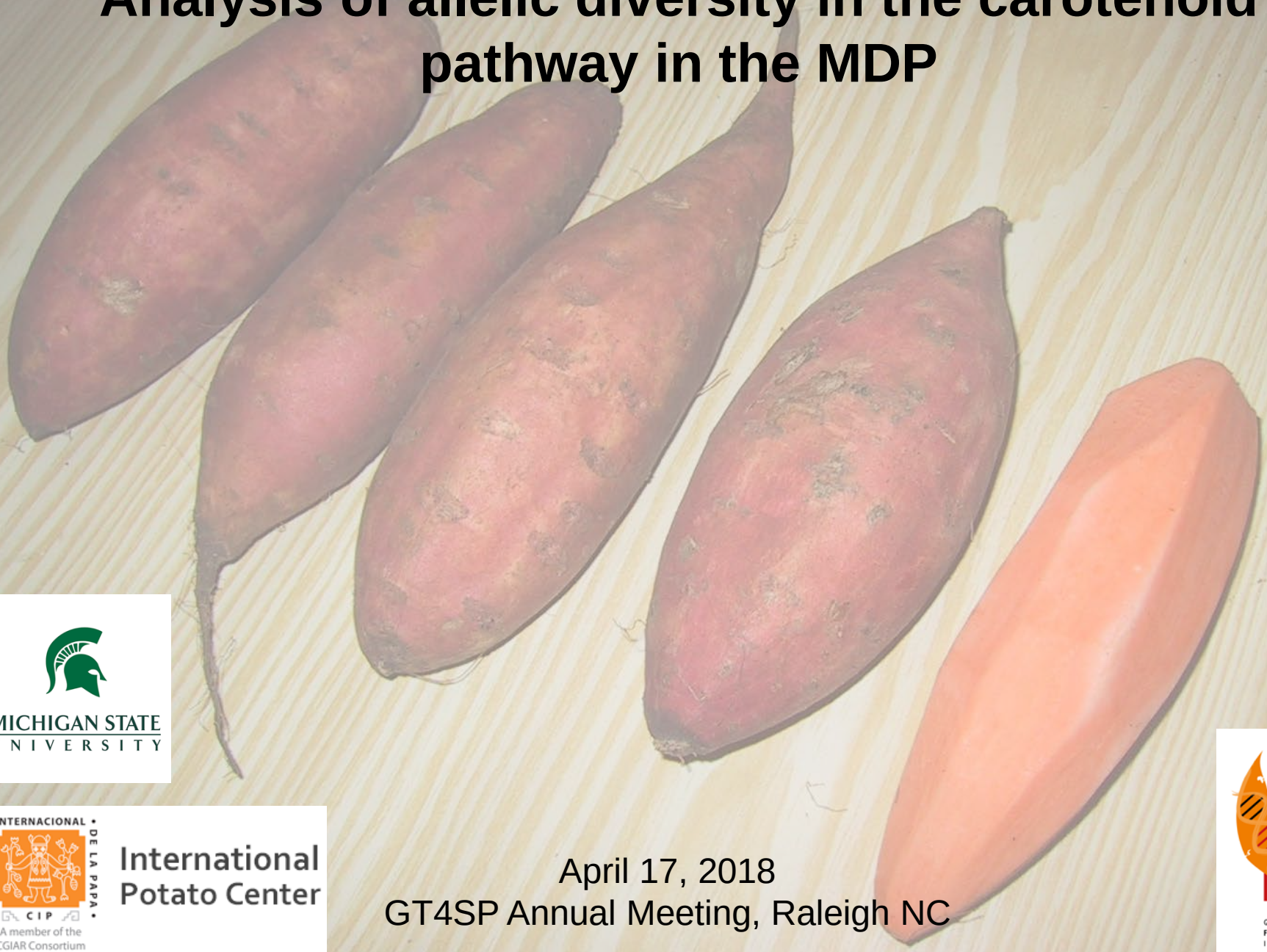
Jbrowse Genome Browser



Tracks available on the Jbrowse:

- Genome Annotation (Loci, Gene Models)
- Gene Predictions (Augustus, SNAP, FGENESH)
- MAKER Transcript Evidence
- MAKER Protein Evidence
- CIP Sweetpotato Gene Index Alignments
- RepeatMasker Repeats
- RNA-Seq Coverage – Wiggle
- RNA-Seq Coverage – XY
- I. trifida 0431-1 SNPs

Analysis of allelic diversity in the carotenoid pathway in the MDP



MICHIGAN STATE UNIVERSITY

CENTRO INTERNACIONAL DE LA PAPA
CIP
A member of the CGIAR Consortium
International Potato Center

April 17, 2018
GT4SP Annual Meeting, Raleigh NC

GT4SP
GENOMIC TOOLS FOR SWEETPOTATO IMPROVEMENT

The Mwanga Diversity Panel (MDP)

- 16 clones selected to make an 8x8 diallel population to exploit heterosis.

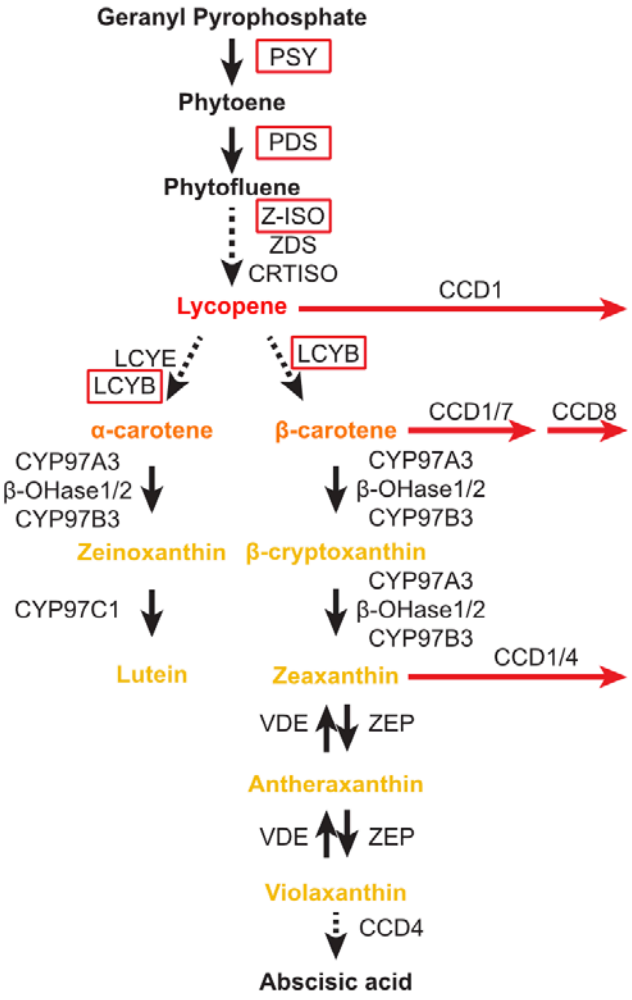
Population A

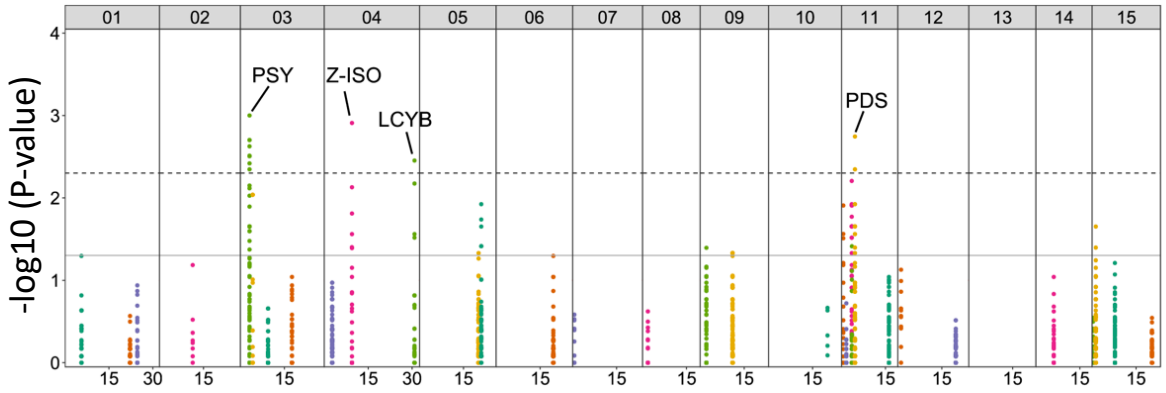
Name	Origin	Flesh color
Ejumula	Uganda	Orange
Kabode	Uganda	Orange
Kakamega	Kenya	Orange
NASPOT 5/58	Uganda	Orange
NASPOT 7	Uganda	Orange
Dimbuka-Bukulula	Uganda	Cream
NASPOT 1	Uganda	Cream
NK259L	Uganda	White

Population B

Name	Origin	Flesh color
Huarmeyano	CIP/Peru	Orange
NASPOT 5	Uganda	Orange
Resisto	USA(Kenya)	Orange
Magabali	Uganda	Cream
Mugande	Uganda	White
NASPOT 11	Uganda	Cream
New Kawogo	Uganda	Cream
Wagabolige	Uganda	Cream

Alleles associated with flesh color





Compared alleles in carotenoid biosynthesis genes between orange and white-flesh varieties.

Detected enriched alleles in homologs of:

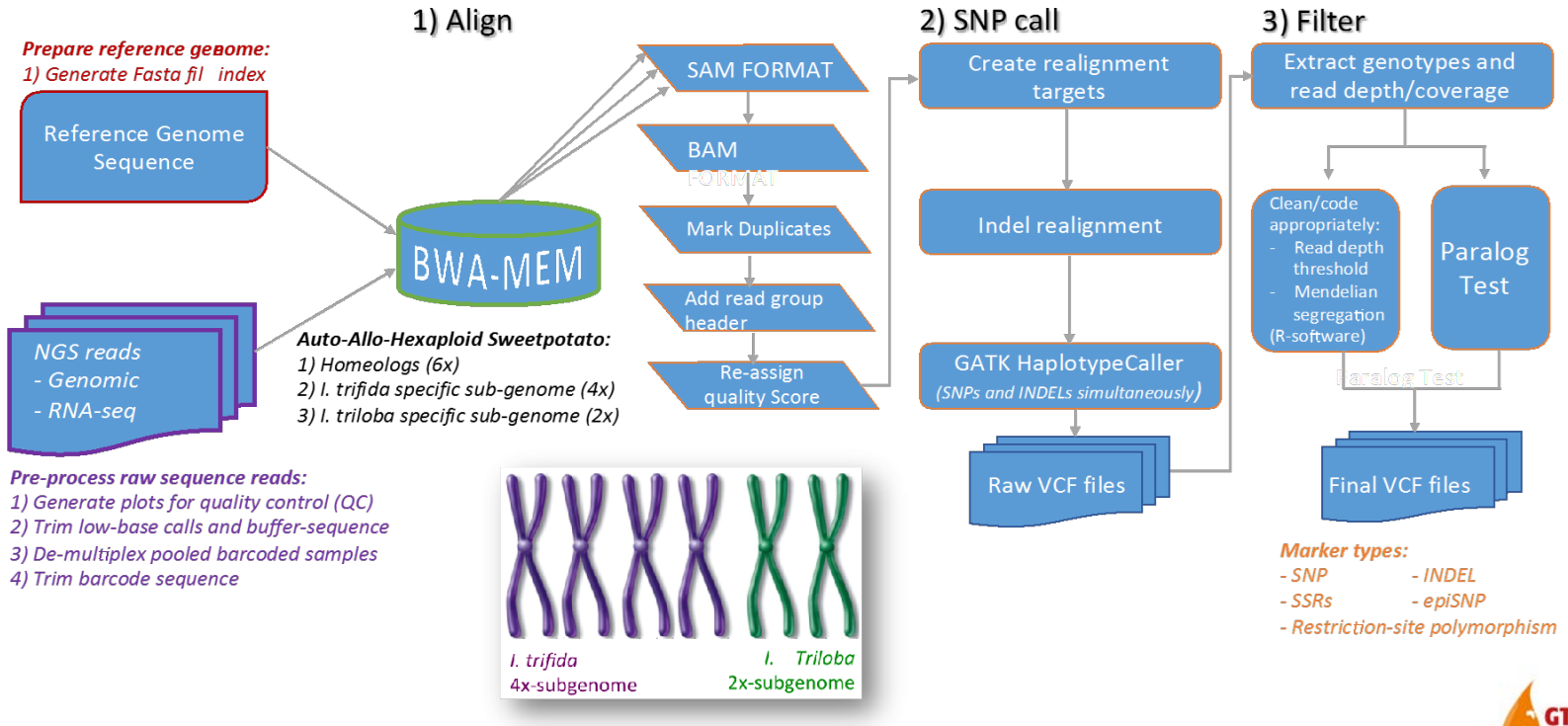
PSY (phytoene synthase),
PDS (phytoene desaturase),
Z-ISO (ζ-carotene isomerase) and
LCYB (lycopene β-cyclase).

Genotyping-by-Sequencing platform

Drs. Bode Olukolu and Yencho

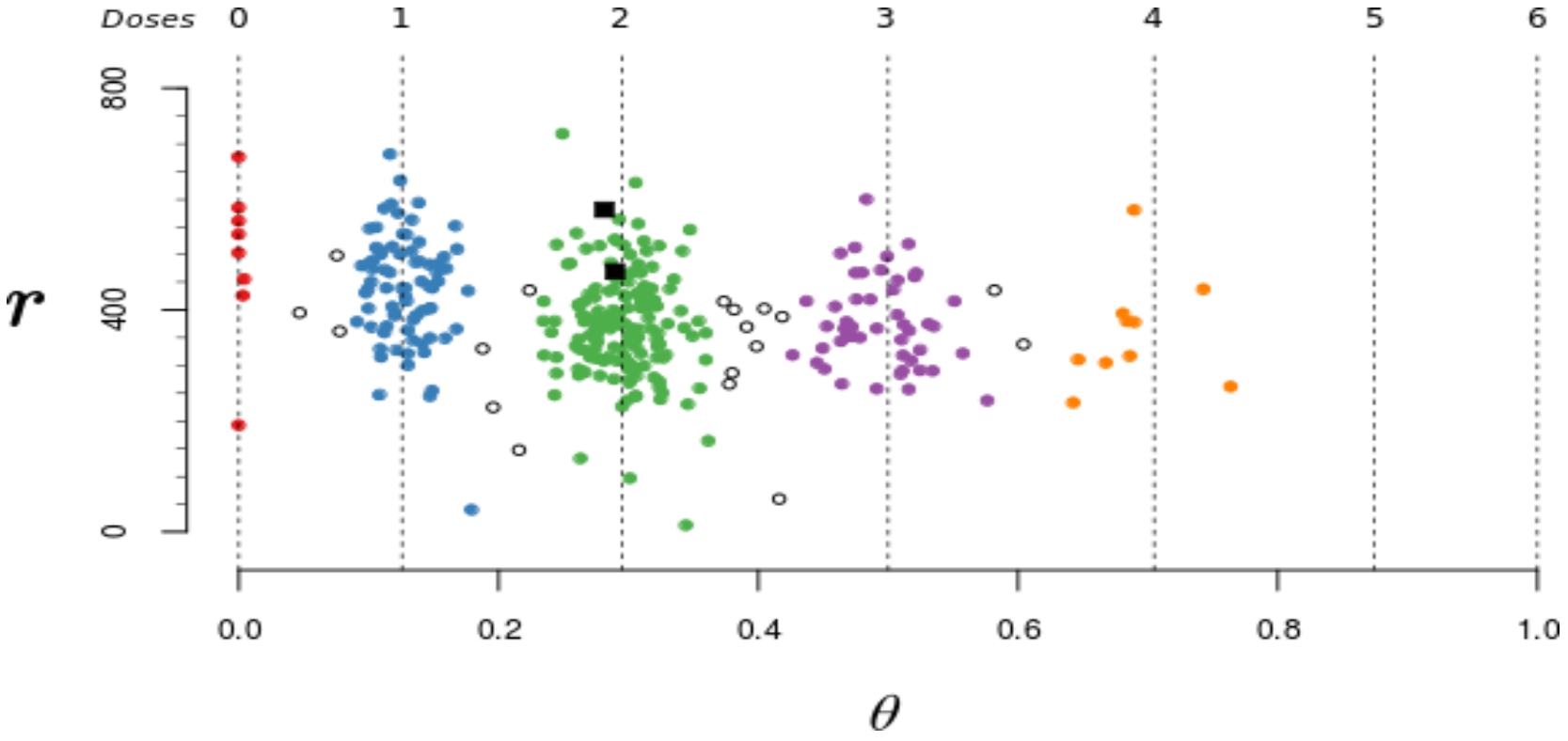
SNP calling pipeline: FastQ to Annotated SNPs

Pipeline Overview



SNP Dosage Calling

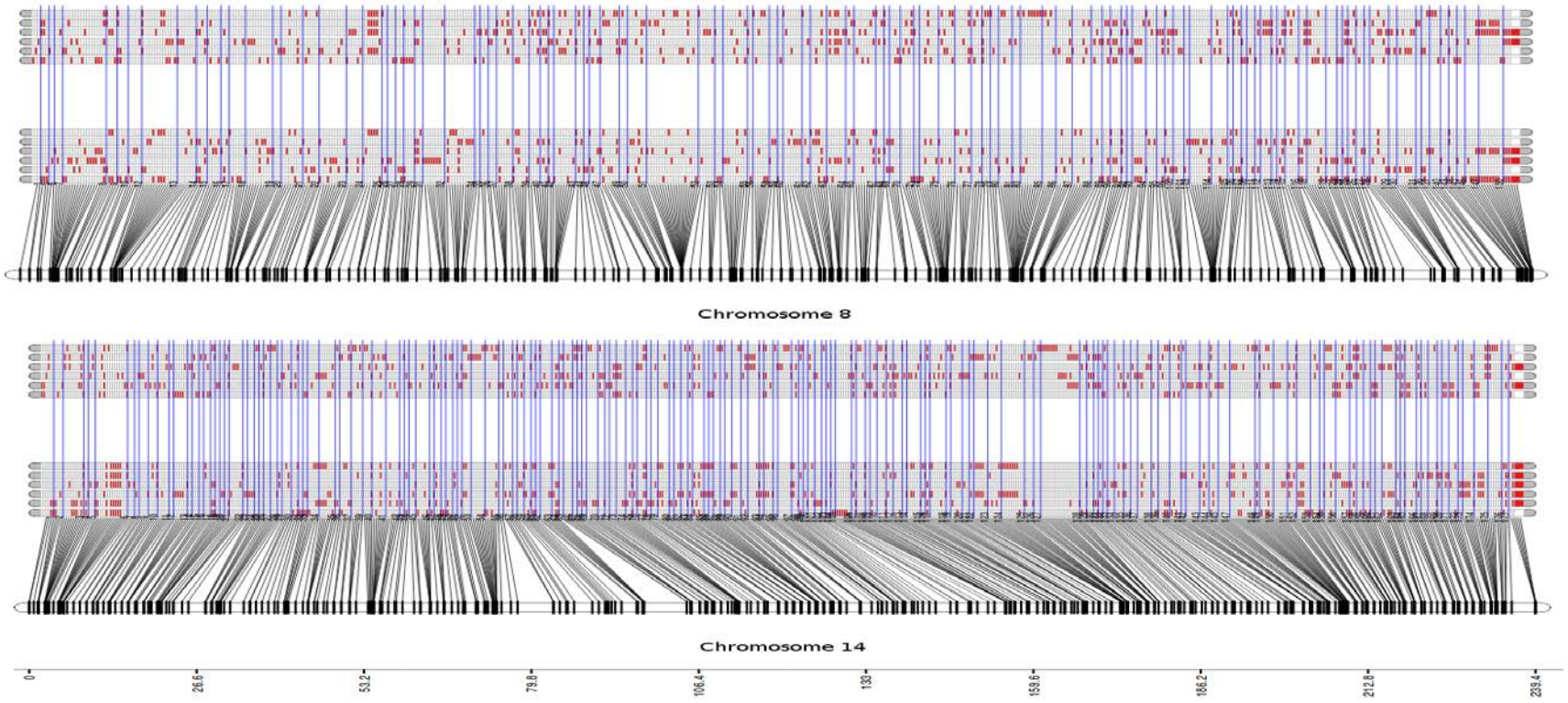
Zhao-Bang Zeng, Marcelo Molinari



Example of a high-quality dose markers located on chromosome 1. r and θ represent a polar transformation for the number of reads for both SNP alleles, i.e., if the number of reads is (x, y), $r = \sqrt{x^2 + y^2}$ and $\theta = atan2(y, x)$. The vertical lines indicate the expected positions for the dosages from 0 to 6. The genotypic classification of each individual is indicated with a different color. The white dots indicate individuals located in transition zones and are treated as missing data. The square dots indicated the parental genotypes. This SNP follow a polysomic segregation where the expected genotypic frequencies are 1:6:11:6:1:0:0.

Phasing and Linkage Mapping

Drs. Zhao-Bang Zeng, Marcelo Molinari, Ghillerme



Linkage maps based on marker blocks for chromosomes 8 and 14. For each map, two homology groups are shown, each one containing six homologous chromosomes. The allelic variants of the SNPs corresponding to the dosages are represented by red rectangles. The blue lines represent the limits between marker blocks.

GT4SP Mapping Populations

- M9 x M19 – Diploid *I. trifida*
  - 212 clones
- Beauregard x Tanzania (BT)
  - 316 clones
- Tanzania x Beauregard (TB)
  - 247 clones
- New Kawogo x Beauregard (NKB),
  - 287 clones
- Mwanga Diversity Population,
  - 8 x 8 mating design,
  - 1,920 clones
- Multi-location, multi-year phenotyping in Peru, USA, Ghana and Uganda

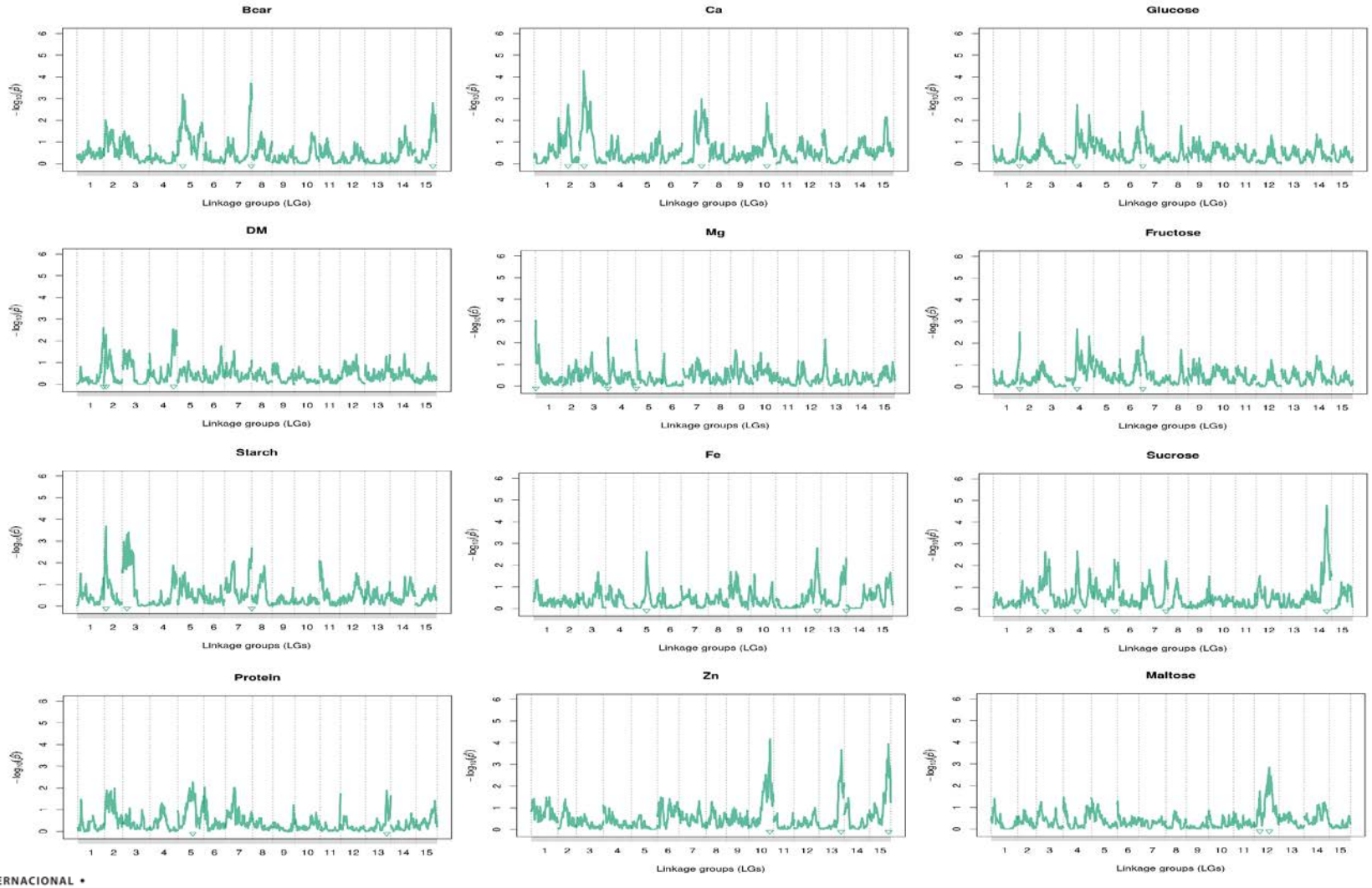
Beauregard

BT

x

Tanzania

BT – QTL Analyses of Nutritional Traits



International Potato Center

A member of the CGIAR Consortium

Multiplication of 8B x 8A Population

Parents (Population UG B)

- B1 - Resisto
- B2 - Magabali
- B3 - NASPOT 5
- B4 - Wagabolige
- B5 - Mugande
- B6 - NASPOT 11
- B7 - New Kawogo
- B8 - Huarmayano

Parents (Population UG A)

- A1 - Ejumula
- A2 - NASPOT 1
- A3 - Dimbuka-Bukulula
- A4 - NASPOT5/58
- A5 - NASPOT 7
- A6 - Kakamega (SPK004)
- A7 - Kabode (NASPOT 10 O)
- A8 - NK259L

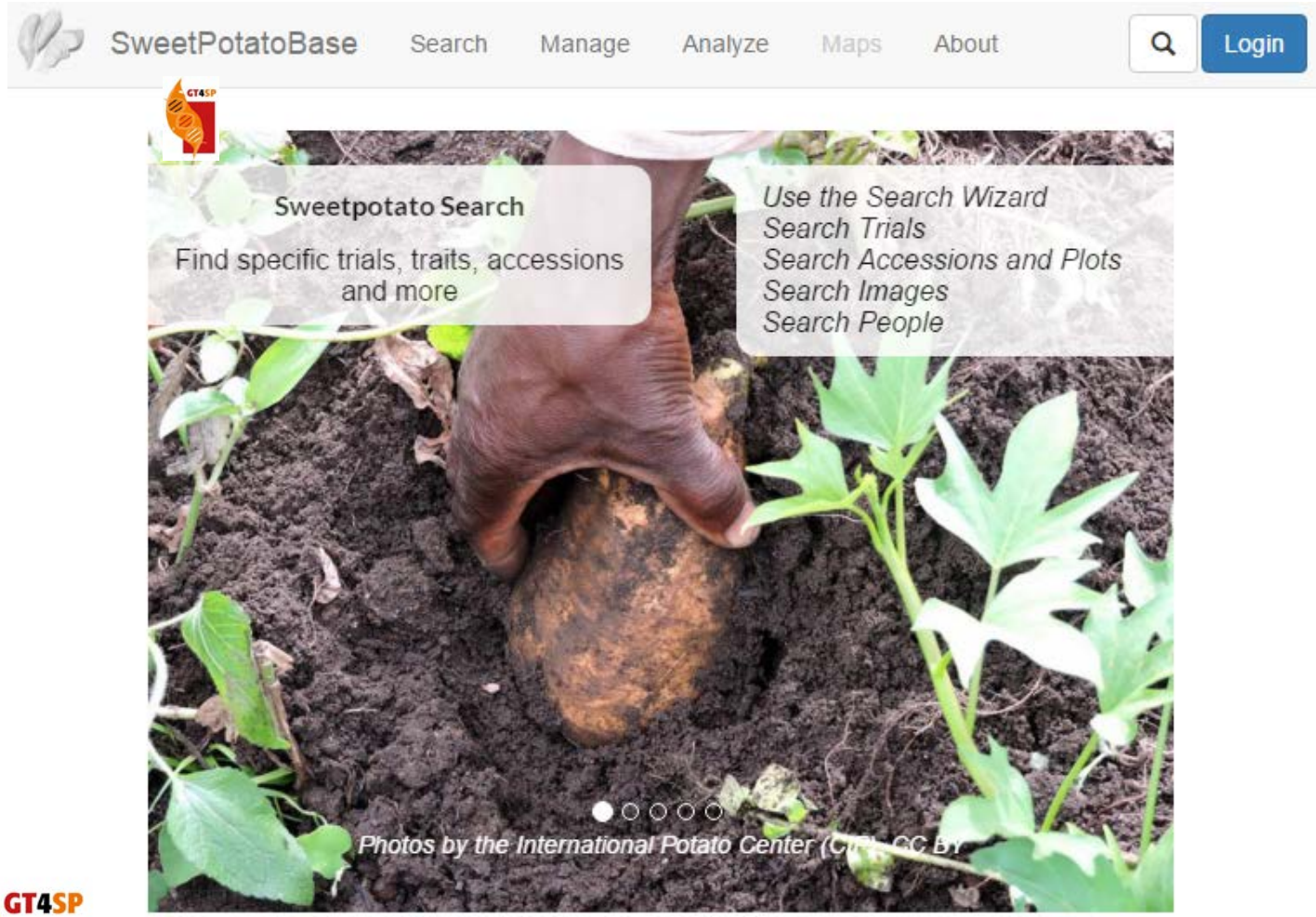
Multiplication at Namulonge in invitro lab-1, screenhouses-2, & net tunnels-3

In vitro plants from BecA/Nairobi

BOYCE THOMPSON INSTITUTE

Dr. Lukas Mueller

Managing Phenotypic Data SweetPotatoBase and FieldBook App.



GT4SP

SweetpotatoBase

Genomic Tools for Sweetp... | Sweetpotato Genomics Res... | http://hidap.sgn.cornell.edu/hd/

hidap.sgn.cornell.edu/hd/

Search

Most Visited | Apple | News | Amazon | SECU | NCSU | MyPack Portal | HS | GT4SP | CALS GC | SPBRD | ACCTS | SPARCS | CALS: eFAR | Expedia | WRAL | WCPSS

HIDAP4RTB

International Potato Center

A member of the CGIAR Consortium

- Phenotype
- Environment
- About new

Highly Interactive Data Analysis Platform for Clonal Plant Breeding

HIDAP v1.0 [07/06/2016]

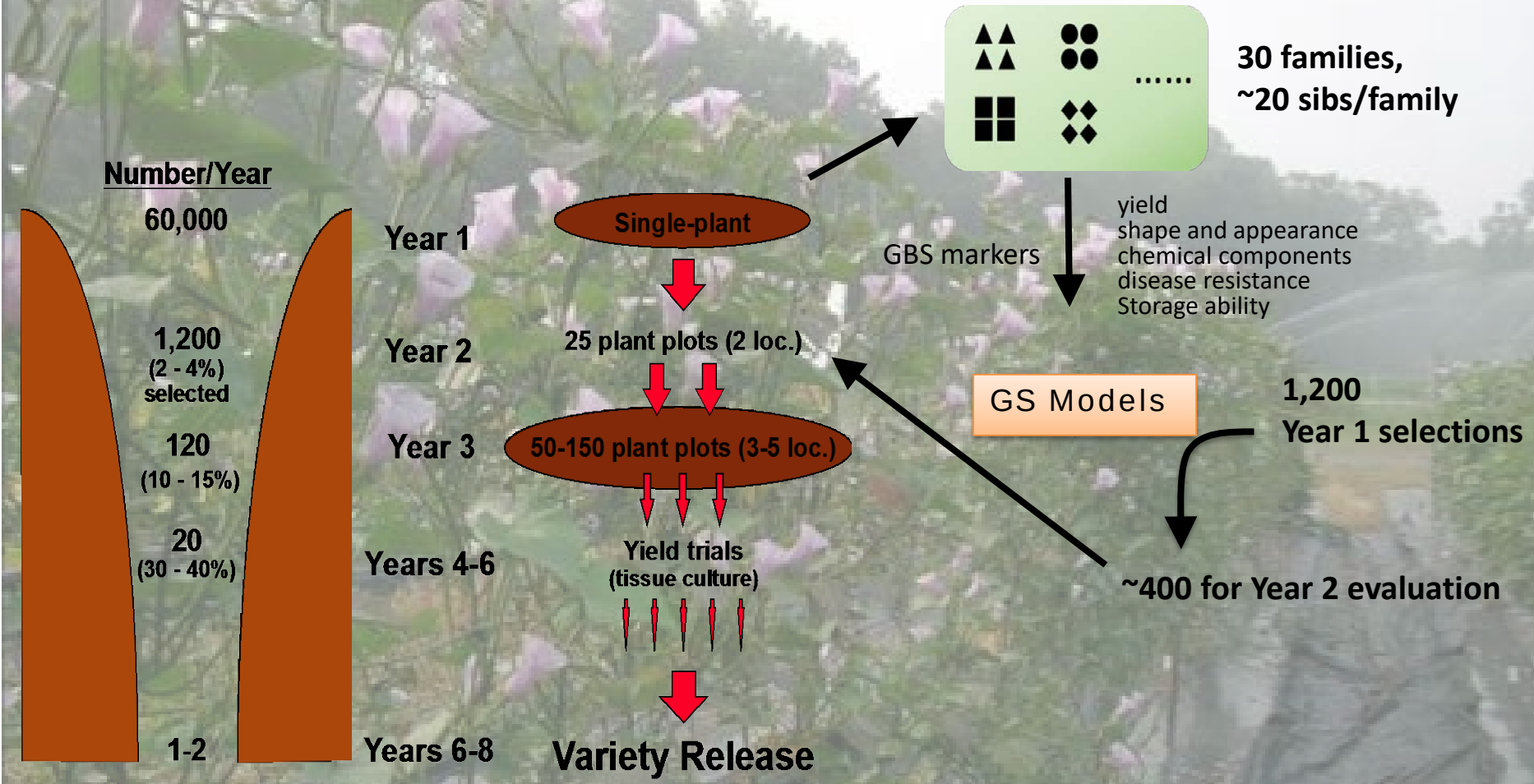
HIDAP is a tool designed to help breeders of clonal plants (likw potato and sweetpotato) carry out field trial planning, documentation, analysis and reporting

International Potato Center (CIP)

GT4SP

Dr. Mercy Kitavi – Lead, Capacity Development

Genomic Selection Proof-of-concept Study
Dr. Xiaofei Zhang



GT4SP Team in Uganda

Afwoyo!

Neyanziza!

Awadifo!

Webale!

Asante!

Thank you!

2nd Annual Mtg.
Kampala, Uganda, May 2017

Our Partners

International Potato Center

BILL & MELINDA GATES foundation

BTI BOYCE THOMPSON INSTITUTE

MICHIGAN STATE UNIVERSITY

The University Of Queensland AUSTRALIA | IMB Institute for Molecular Bioscience

Golden LEAF FOUNDATION

RESEARCH STATIONS DIVISION
Working Together for One Cause
NC DEPARTMENT OF AGRICULTURE
NC STATE UNIVERSITY

NC Certified Sweet Potato Seed Producers, Inc.

biosciences eastern and central africa

NORTH CAROLINA SWEETPOTATO COMMISSION

North Carolina Sweet Potatoes Nature's Health Food

UNIVERSITY of GREENWICH | Natural Resources Institute

NC STATE UNIVERSITY

Think and Do
ncsu.edu

http://www.blakeleypritchard.com/ncsu-chancellors-office-e-card/