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
The Genomic Tools for Sweetpotato Improvement Project
GT4SP

Sweetpotato Genome
(MSU, BTI, CIP)

Bioinformatics
(NCSU, UQ, CIP)

Genotyping by Sequencing
(NCSU, UQ, CIP, BTI)

Sweetpotato Database: Bioinformatics, Phenotyping & Genomics
(CIP, MSU, BTI, NaCCRI)

Sweetpotato Breeding & Capacity Development
(NCSU, CIP-Peru, CIP-Uganda, CIP-Ghana, CIP-Kenya)

An ambitious project to sequence sweetpotato and develop modern breeding tools for a food crop that sustains millions 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
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

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

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

CIP Lima
Dorcas Gemenet, Geneticist, Molecular Breeding
Wolfgang Gruneberg, SP Breeding, Global Lead
Merideth Bonierbale, Program 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

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

Lachlan Coin (coPI)
Chenxi Zhou, PhD Student, China
Marian Quain, Biotechnology, TC
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; Arumuganathan 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
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.

<table>
<thead>
<tr>
<th></th>
<th><em>I. trifida</em></th>
<th></th>
<th><em>I. triloba</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contigs</td>
<td>Scaffolds</td>
<td>Contigs</td>
</tr>
<tr>
<td>Number</td>
<td>44,847</td>
<td>30,398</td>
<td>31,279</td>
</tr>
<tr>
<td>N50 (bp)</td>
<td>65,820</td>
<td>1,237,020</td>
<td>36,931</td>
</tr>
<tr>
<td>Longest (bp)</td>
<td>1,067,799</td>
<td>8,902,984</td>
<td>313,171</td>
</tr>
<tr>
<td>Size (bp)</td>
<td>433,252,193</td>
<td>462,000,517</td>
<td>437,557,497</td>
</tr>
</tbody>
</table>

87.7% of the estimated genome sizes (526.5 Mb and 495.9 Mb for *I. trifida* and *I. triloba*, respectively.)
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
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

April 17, 2018
GT4SP Annual Meeting, Raleigh NC
The Mwanga Diversity Panel (MDP)

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Origin</th>
<th>Flesh color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ejumula</td>
<td>Uganda</td>
<td>Orange</td>
</tr>
<tr>
<td>Kabode</td>
<td>Uganda</td>
<td>Orange</td>
</tr>
<tr>
<td>Kakamega</td>
<td>Kenya</td>
<td>Orange</td>
</tr>
<tr>
<td>NASPOT 5/58</td>
<td>Uganda</td>
<td>Orange</td>
</tr>
<tr>
<td>NASPOT 7</td>
<td>Uganda</td>
<td>Orange</td>
</tr>
<tr>
<td>Dimbuka-Bukulula</td>
<td>Uganda</td>
<td>Cream</td>
</tr>
<tr>
<td>NASPOT 1</td>
<td>Uganda</td>
<td>Cream</td>
</tr>
<tr>
<td>NK259L</td>
<td>Uganda</td>
<td>White</td>
</tr>
<tr>
<td>Huarmeyano</td>
<td>CIP/Peru</td>
<td>Orange</td>
</tr>
<tr>
<td>NASPOT 5</td>
<td>Uganda</td>
<td>Orange</td>
</tr>
<tr>
<td>Resisto</td>
<td>USA(Kenya)</td>
<td>Orange</td>
</tr>
<tr>
<td>Magabali</td>
<td>Uganda</td>
<td>Cream</td>
</tr>
<tr>
<td>Mugande</td>
<td>Uganda</td>
<td>White</td>
</tr>
<tr>
<td>NASPOT 11</td>
<td>Uganda</td>
<td>Cream</td>
</tr>
<tr>
<td>New Kawogo</td>
<td>Uganda</td>
<td>Cream</td>
</tr>
<tr>
<td>Wagabolige</td>
<td>Uganda</td>
<td>Cream</td>
</tr>
</tbody>
</table>
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

1) Align
   - BWA-MEM
   - SAM FORMAT
   - BAM FORMAT
   - Mark Duplicates
   - Add read group header
   - Re-assign quality Score

2) SNP call
   - Create realignment targets
   - Indel realignment
   - GATK HaplotypeCaller (SNPs and INDELs simultaneously)
   - Raw VCF files
   - Extract genotypes and read depth/coverage

3) Filter
   - Extract genotypes and read depth/coverage
   - Paralog Test
   - Final VCF files

Pre-process raw sequence reads:
1) Generate plots for quality control (QC)
2) Trim low-base calls and buffer-sequence
3) De-multiplex pooled barcoded samples
4) Trim barcode sequence

Auto-Allo-Hexaploid Sweetpotato:
1) Homeologs (6x)
2) I. trifida specific sub-genome (4x)
3) I. triloba specific sub-genome (2x)
Example of a high-quality dose markers located on chromosome 1. \( r \) and \( \theta \) 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 \( \theta = \text{atan2}(y, x) \). 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
BT

Beauregard  x  Tanzania
BT – QTL Analyses of Nutritional Traits

![Graphs showing QTL analyses of nutritional traits for various elements and sugars.](Image)
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 - NASPOT 5/58
A5 - NASPOT 7
A6 - Kakamega (SPK004)
A7 - Kabode (NASPOT 10 O)
A8 - NK259L

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

In vitro plants from BecA/Nairobi
Managing Phenotypic Data
SweetPotatoBase and FieldBook App.

Sweetpotato Search
Find specific trials, traits, accessions and more

Use the Search Wizard
Search Trials
Search Accessions and Plots
Search Images
Search People

Photos by the International Potato Center (CIP). CC BY
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 (like potato and sweetpotato) carry out field trial planning, documentation, analysis and reporting.
Genomic Selection Proof-of-concept Study
Dr. Xiaofei Zhang

30 families, ~20 sibs/family

GS Models

GBS markers

yield
shape and appearance
chemical components
disease resistance
Storage ability

1,200
Year 1 selections

~400 for Year 2 evaluation

Number/Year
60,000
1,200 (2 - 4%) selected
120 (10 - 15%)
20 (30 - 40%)
1-2

Year 1
Single-plant

Year 2
25 plant plots (2 loc.)

Year 3
50-150 plant plots (3-5 loc.)

Years 4-6
Yield trials (tissue culture)

Years 6-8
Variety Release
GT4SP Team in Uganda

Afwoyo!
Neyanziza!
Awadifo!
Webale!
Asante!
Thank you!

2nd Annual Mtg.
Kampala, Uganda, May 2017
Our Partners

International Potato Center

NaCRRI

Bill & Melinda Gates Foundation

BTI Boyce Thompson Institute

NC Certified Sweet Potato Seed Producers, Inc.

MICHIGAN STATE UNIVERSITY

The University of Queensland Australia

IMB Institute for Molecular Bioscience

RESEARCH STATIONS DIVISION
Working Together for One Cause

NC DEPARTMENT OF AGRICULTURE
NC STATE UNIVERSITY

Golden LEAF Foundation

ConAgra Foods

McCain

Proven Winners

PEPSICO

North Carolina Sweet Potato Commission

NC State University

Think and Do ncsu.edu