

# Development of genetic and genomic resources for breeding improved sweetpotato varieties

Roland Schafleitner

International Potato Center - CIP



# Sweetpotato Genetic Resources

Sweetpotato  
importance

Sweetpotato  
Genomic  
Resources

22.000 ESTs  
1500 gene sequences  
1 Bac  
2 Low resolution maps  
~100 SSR markers  
LD cDNA microarray



# Aims - Accomplishments

## **Design and document a CGS**

accomplished  
and posted

Establish, upgrade health status, propagate distribute CGS subsets in vitro

accomplished

## **Develop and curate a virus-free diploid mapping population in vitro**

in progress

## **Produce a medium density reference map for sweetpotato.**

accomplished  
and posted

## **Establish and sequence EST libraries from leaves and roots of drought stressed plants**

Sweetpotato gene index

accomplished

## **Develop and optimise DArT array**

accomplished  
and posted

## **Design additional SSR markers**

accomplished  
and posted

## **Identify SNP marker (EST sequence screening)**

in progress

## **Apply cosII marker on sweetpotato**

in progress

## **Genotype the CGS with DArT markers and a subset of SSR marker**

Compare SSR and DArT data

in progress

## **Characterize a hexaploid QTL/trait mapping and the diploid reference population by DArT and the full SSR marker set**

# Sweetpotato CGS

## **480 accessions**

- **Biodiversity (geographical origin)**
- **High beta-carotene**
- **High starch**
- **High iron and zinc contents**
- **Nematode, virus and drought resistance**

**Pathogen-free in vitro plants ready for international distribution**

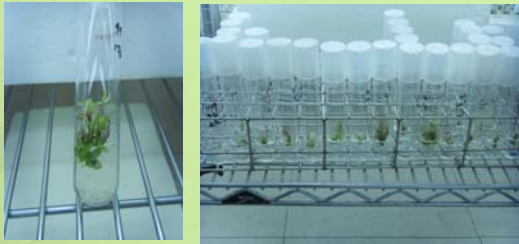
**Morphological information and molecular fingerprints (SSR genotypes) available**

**[http://gcpcr.grinfo.net/files/cr\\_files/gcpcr\\_file832.xls](http://gcpcr.grinfo.net/files/cr_files/gcpcr_file832.xls)**

# Virus-free diploid population

## ***Ipomoea trifida***

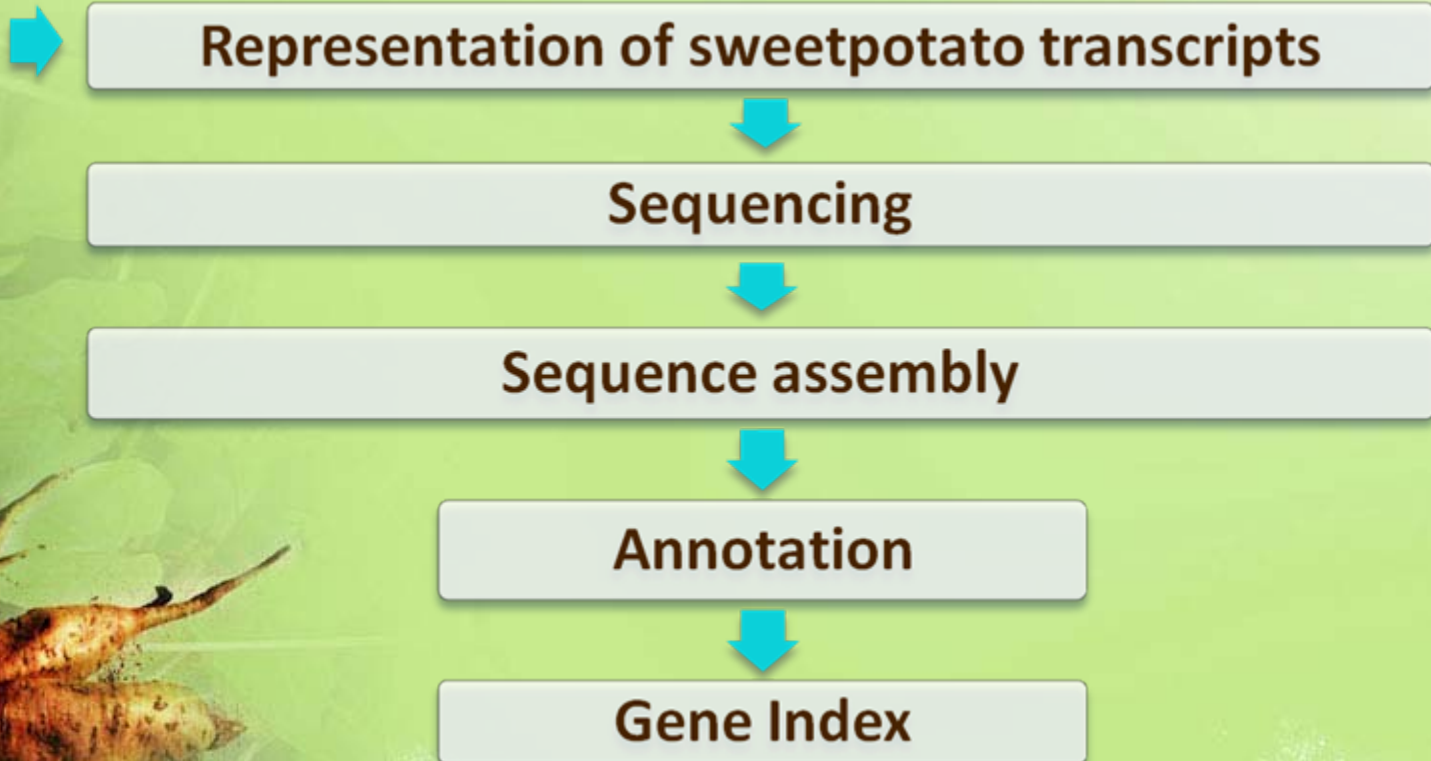
**Population M09 x M19**



- ~300 seeds
- 150 plants established in vitro
- Optimization of culturing conditions – increase growth rate

# EST library

**Production of normalized cDNA libraries from different tissues (and different clones)**



# 454 Sequencing

Raw Data:

523.914 sequence reads from 2 libraries (stems and leaves)

In total ~158.000 bases

~300 bp per read

22.000 Sanger ESTs  
from storage roots

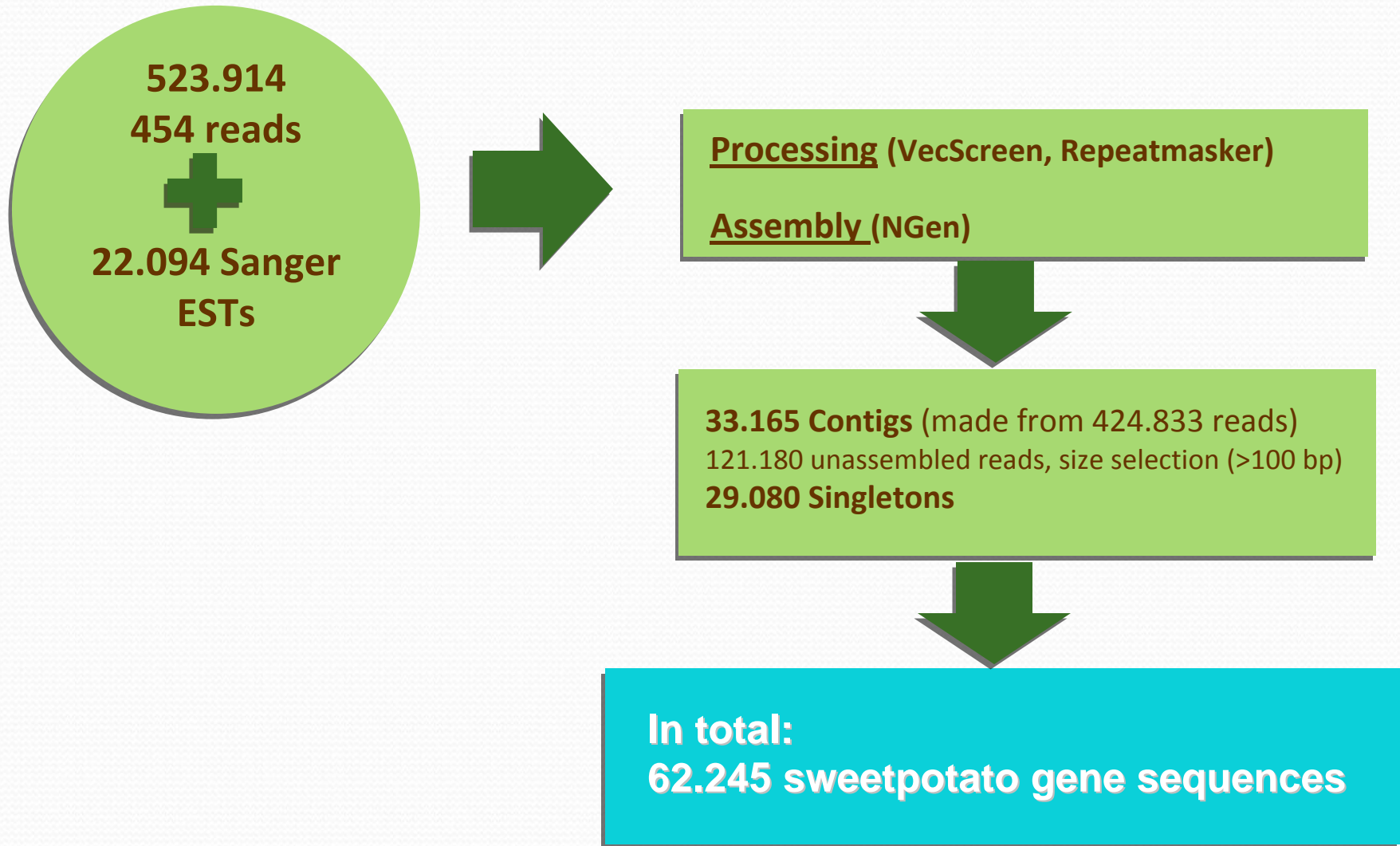
Sequences

Quality

The screenshot shows the BioEdit Sequence Alignment Editor interface. The main window displays a sequence alignment for file 'C:\Documents and Settings\schaffelner\Local Settings\Temp\RAWREADS.fas'. The alignment consists of multiple sequence reads, each with a unique identifier (e.g., 000014, 000021, 000033) and a length (e.g., 112, 112, 112). The sequences are aligned to a reference sequence, with gaps indicated by dashes. The interface includes a menu bar (File, Edit, Sequence, Alignment, View, Accessory Application, RNA, World Wide Web, Options, Window, Help), a toolbar with various editing tools, and a status bar at the bottom showing 'Selected 1 Res. 25.539B' and 'Total 20 Res. 244.319B'.

The screenshot shows the RAWREADS\_WordDat application window. The main window displays a list of sequence reads with their corresponding quality scores. Each entry includes a read ID (e.g., 000014, 000021, 000033), a length (e.g., 112, 112, 112), and a quality score (e.g., 112, 112, 112). The quality scores are represented by a series of numbers, likely indicating the confidence or accuracy of each base call. The interface includes a menu bar (File, Edit, View, Insert, Format, Help) and a status bar at the bottom showing 'For help, press F1'.

# Sequence Assembly



# Assembly Results ➡ 33.165 contigs

**Mean length: 788 bp**

**8011 contigs > 1000 bp**

**Average coverage: 14.1-fold**

**Average coverage of long contigs (>1000bp): 29.6-fold**

# Annotation

Sequence comparison to known genes on DNA and protein level

	NCBI	UNIPROT A. thaliana	GO- Annotation
Contigs and Singletons	37,836	24,994	23,247
Unique		14,145	NA
No hits	24,409	37,251	38,998

Chloroplast and mitochondrial genes:



**Redundance!**

*Blast against databases and selfblast suggests considerable redundancy in the sequence assembly*

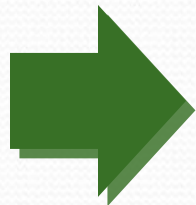
## Causes for redundancy

- **Heterozygosity and hexaploidy: allelic variants sometimes are kept separated in different contigs**
- **Paralogs**
- **Contigs represent gene fragments**
- **Sporamin A and B: Highly expressed genes, allelic variation plus sequencing errors lead to allocation of reads in different contigs**
- **Chloroplast and mitochondrial genes: up to 4500 454 reads derived from 1 gene!).**

# “No Hit” sequences

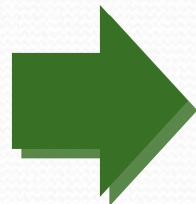
## Genes or artifacts?

- ✓ Most contigs with no hits have long open reading frames
- ✓ 50% of the contigs without hits have Poly-A Tails
- ✓ GC Content of contigs without and with hits are not significantly different



**Contigs without significant hits probably represent new genes**

- ✓ Singletons without hit have significantly different GC content than contigs and singletons with hits



**Singletons without hit may represent both, new genes or not-protein-coding RNAs**

# Gene Index

**The actual gene index is available at**

[http://gcpcr.grinfo.net/index.php?app=datasets&inc=dataset\\_details&dataset\\_id=712](http://gcpcr.grinfo.net/index.php?app=datasets&inc=dataset_details&dataset_id=712)

# SNP Detection in selected genes

200 Stress-related genes were selected for SNP detection

- DNA-binding, DNA remodeling, RNA-binding, Transcription
- Signaling
- Redox
- Hormone
- Mitochondria, Mitochondrial transporter
- Photosynthesis, Photorespiration, Chlororespiration
- Water
- Transport (Anion, Ca, N, peptide, sugar)
- Glucose-, Fructose-, Sucrose-, Starch, Raffinose, Trehalose metabolism
- Myo-inositol biosynthesis
- Amino acid metabolism (Proline)
- Polyamine synthesis
- Lipid metabolism (Lipid transfer proteins, Wax)
- Secondary metabolism (Phenylpropanoids, Terpenes)
- Phosphate metabolism
- Cell wall
- Protein destination
- Abiotic stress

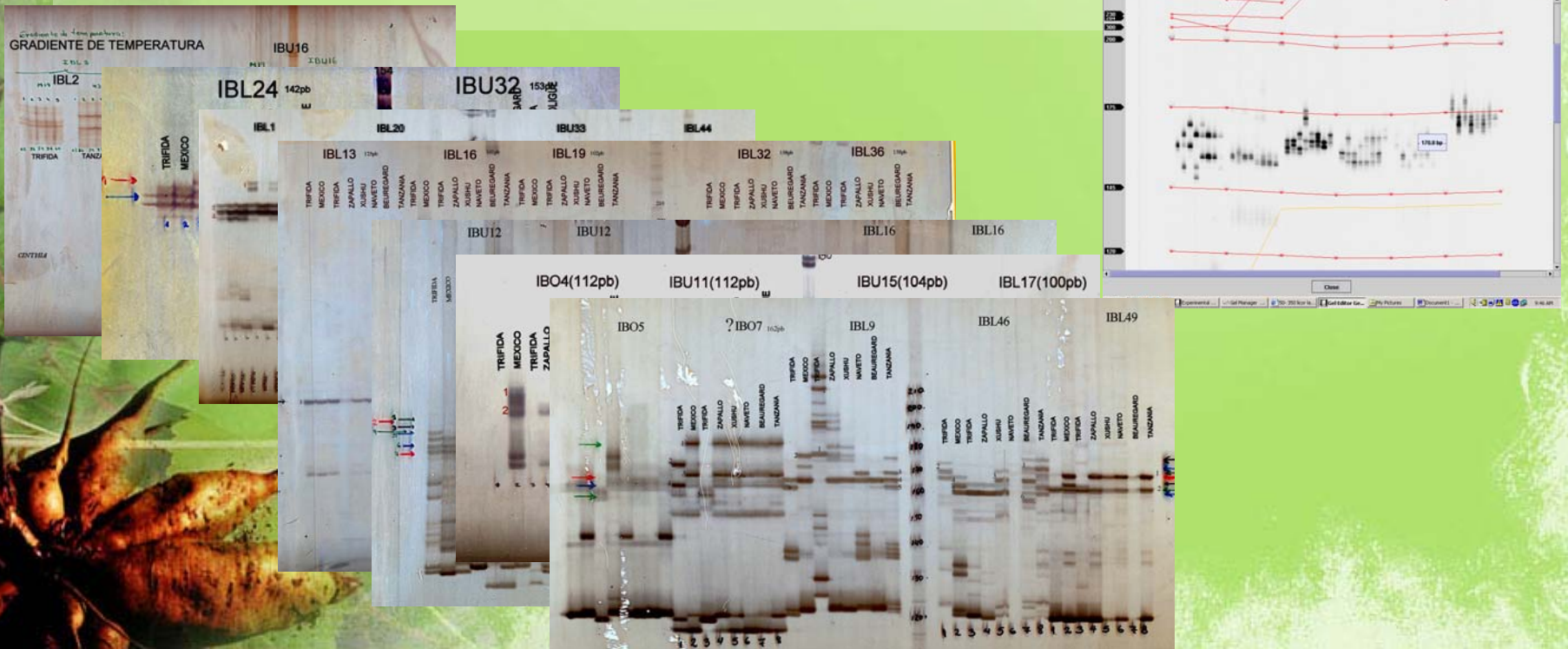


# SSR Markers

Around 300 new candidate SSRs identified and tested

>200 can be amplified and are polymorphic in 8 clone test panel and can be used for diversity studies

50 new SSRs are polymorphic in *I. trifida* cross



# DArT Markers

~ 2000 polymorphic fragments obtained

Useful for diversity analysis

~ 30% of the polymorphic fragments that can be used for genotyping populations

# ...in progress

- **COSII marker: 400 cosII sequenced. Testing in progress**
- **Diploid reference map**
- **CGS genotyping**

# Thanks to



**CIP:** Rosalina Tincopa, Genoveva Rossel, Omar Palomino, Jorge Espinoza, Julio Solis, Ji Young Kim, Carlos Rivera, Cynthia Quispe, Letitia Portal

**DArT:** Andrezej Kilian and staff

**E. Mondlane Univ.:** Yvone Martins Muchoa (Mozambique)

**NAARI:** Robert Mwanga (Uganda)

**Evrogen**  
**Lasergen**