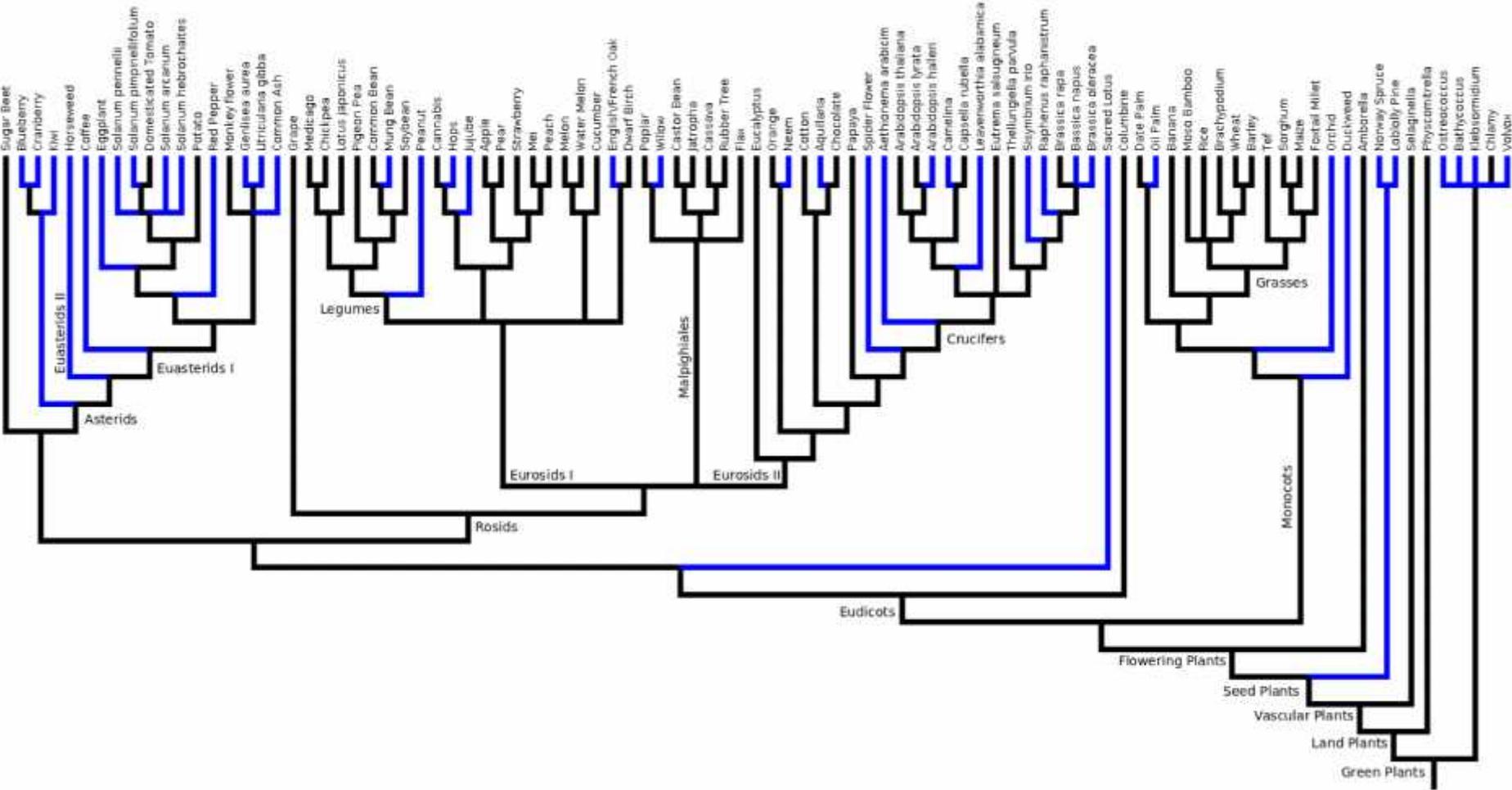


# Sweetpotato reference genome sequencing

Zhangjun Fei  
Boyce Thompson Institute for Plant Research  
Cornell University

# Plants with sequenced genomes

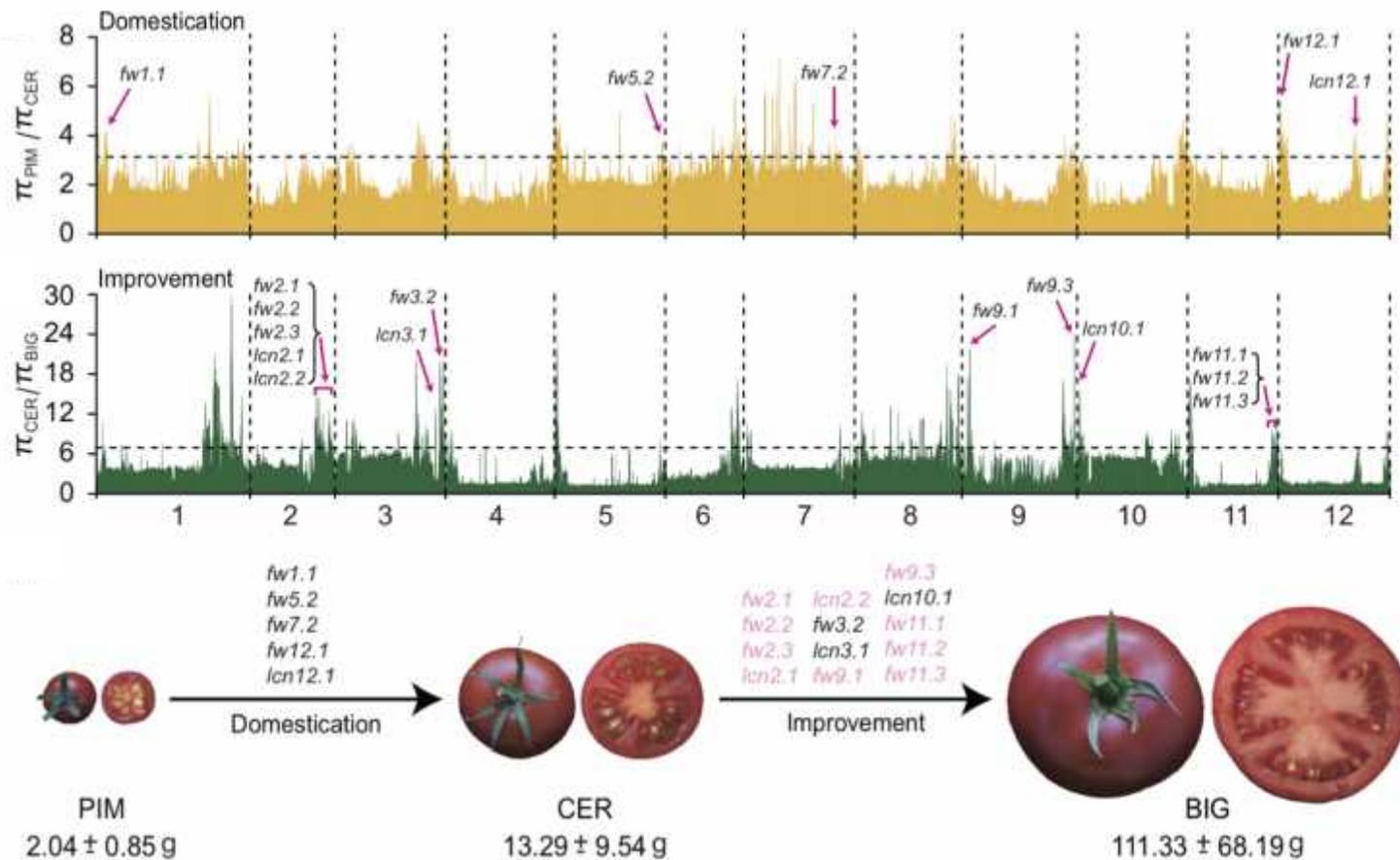


More than 70 plant species

# Why we need genome sequences?

- ❖ The genome contains all the genetic information of an organism, which determines its phenotype
- ❖ High-quality reference genome sequence provides a foundation that can facilitate basic researches, gene/QTL cloning, molecular marker discovery and marker assisted breeding
- ❖ Genome sequences can help understand genome evolution and domestication history
- ❖ The **industries** (growers, shippers, processors) depend on high quality, disease resistant cultivars
- ❖ The **breeding community** (seed companies and public breeders) must develop these cultivars.
- ❖ The **scientific community** develops knowledge to facilitate more effective plant breeding and train the next generation of plant scientists and breeders

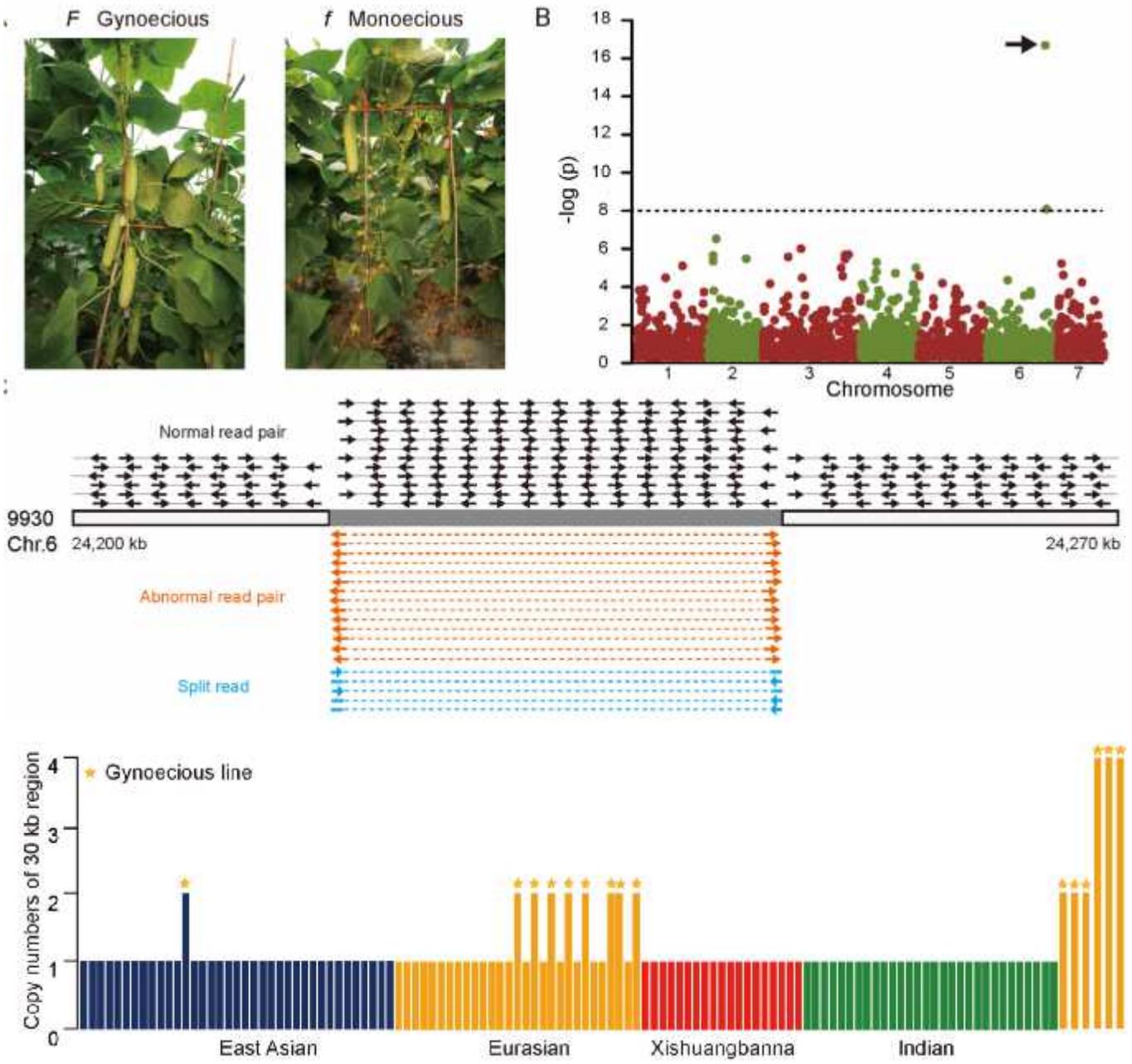
# Genomic analyses provide insights into the history of tomato breeding



This article is a *Plant Cell Advance* Online Publication. The date of its first appearance online is the official date of publication. The article has been edited and the authors have corrected proofs, but minor changes could be made before the final version is published. Posting this version online reduces the time to publication by several weeks.

## Genome-Wide Mapping of Structural Variations Reveals a Copy Number Variant That Determines Reproductive Morphology in Cucumber

- ❖ Cucumber is a model system for sex determination studies
- ❖ Cucumber has three types of flowers: male, female, and bisexual.
- ❖ Seven sex types of cucumber plants: androecious (only male flowers), **gynoecious (only female flowers)**, **monoecious (male flowers at the base and female flowers at the top of the main stem)**, hermaphroditic (only bisexual flowers), andromonoecious (male and bisexual flowers), gynomonoecious (female and bisexual flowers), and trimonoecious (male, female, and bisexual flowers).
- ❖ Gynoecious plants can set fruit at almost every node, improving the yield.



# Marker (SNP) discovery requires a high-quality reference genome sequence

OPEN ACCESS Freely available online



## A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species

Robert J. Elshire<sup>1</sup>, Jeffrey C. Glaubitz<sup>1</sup>, Qi Sun<sup>2</sup>, Jesse A. Poland<sup>3</sup>, Ken Kawamoto<sup>1</sup>, Edward S. Buckler<sup>1,4</sup>, Sharon E. Mitchell<sup>1\*</sup>

**1** Institute for Genomic Diversity, Cornell University, Ithaca, New York, United States of America, **2** Computational Biology Service Unit, Cornell University, Ithaca, New York, United States of America, **3** Hard Winter Wheat Genetics Research Unit, United States Department of Agriculture/Agricultural Research Service, Manhattan, Kansas, United States of America, **4** Plant, Soil and Nutrition Research Unit, United States Department of Agriculture/Agricultural Research Service, Ithaca, New York, United States of America

### Abstract

Advances in next generation technologies have driven the costs of DNA sequencing down to the point that genotyping-by-sequencing (GBS) is now feasible for high diversity, large genome species. Here, we report a procedure for constructing GBS libraries based on reducing genome complexity with restriction enzymes (REs). This approach is simple, quick, extremely specific, highly reproducible, and may reach important regions of the genome that are inaccessible to sequence capture approaches. By using methylation-sensitive REs, repetitive regions of genomes can be avoided and lower copy regions targeted with two to three fold higher efficiency. This tremendously simplifies computationally challenging alignment problems in species with high levels of genetic diversity. The GBS procedure is demonstrated with maize (IBM) and barley (Oregon Wolfe Barley) recombinant inbred populations where roughly 200,000 and 25,000 sequence tags were mapped, respectively. An advantage in species like barley that lack a complete genome sequence is that a reference map need only be developed around the restriction sites, and this can be done in the process of sample genotyping. In such cases, the consensus of the read clusters across the sequence tagged sites becomes the reference. Alternatively, for kinship analyses in the absence of a reference genome, the sequence tags can simply be treated as dominant markers. Future application of GBS to breeding, conservation, and global species and population surveys may allow plant breeders to conduct genomic selection on a novel germplasm or species without first having to develop any prior molecular tools, or conservation biologists to determine population structure without prior knowledge of the genome or diversity in the species.

# Objectives and team members

## **Objective 1. Development of the core genomic and genetic resources for sweetpotato improvement**

- ❖ Objective 1.A. Whole genome sequencing of NCNSP-0323, a homozygous diploid sweetpotato progenitor of cultivated sweetpotato I. batatas. Lead Scientists: [Fei \(BTI-CU\)](#) and [Buell \(MSU\)](#).
  - Objective 1.A.1. Genome DNA library preparation and sequencing.
  - Objective 1.A.2. Genome assembly and annotation.
  - Objective 1.A.3. Genome evolution and comparative genomic analysis of NCNSP-0323.
- ❖ Objective 1.B. Transcriptome profiling of multiple tissues of the sequenced NCNSP-0323 for genome annotation. Lead Scientists: [Fei \(BTI-CU\)](#) and [Buell \(MSU\)](#).
- ❖ Objective 1.C. Development of diploid mapping populations for high density SNP genome sequence anchoring and QTL mapping. Lead Scientists: [Gruneberg and Khan \(CIP\)](#); [Yencho and Quesada \(NCSU\)](#), [Fei \(BTI-CU\)](#) and [Buell \(MSU\)](#).
- ❖ Objective 1.D. Sweetpotato genome browser development. Lead Scientists: [Buell \(MSU\)](#) and [Fei \(BTI-CU\)](#).

# Sweetpotato genome sequencing

## Ideal system for whole genome sequencing

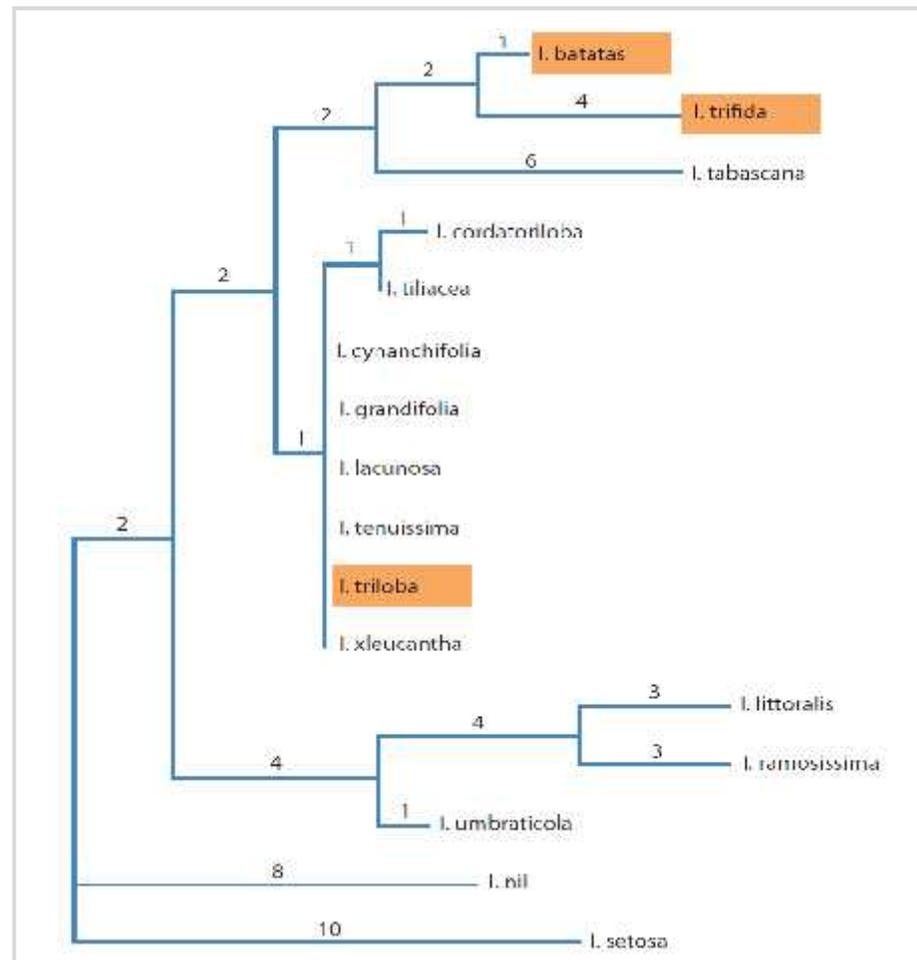
- ❖ Highly homozygous (such as inbred lines)
- ❖ Haploid/Diploid
- ❖ Relatively small genome size
- ❖ Containing small portion of repetitive sequences

Cultivated sweetpotato is an allo-auto-hexaploid ( $2n=6x=90$ ) with two non-homologous genomes ( $B_1B_1B_2B_2B_2B_2$ ).

**Polyploid and highly heterozygous**

# Sweetpotato genome sequencing

**Strategy:** Sequencing the closely related wild ancestors that are diploid and homozygous  
potato, wheat, cotton, strawberry



# Sweetpotato genome sequencing

**NCNSP-0323**, an inbred trifida line developed by Craig Yencho's group. It is derived from PI 618966 that is self-compatible.

The screenshot shows the GRIN Global Accession Detail page for PI 618966. The page is titled "Accession: PI 618966 - GRIN" and displays the following information:

**PI 618966**  
***Ipomoea trifida* (Kunth) G. Don**

Status: Available  
Amt Distributed: 10 count  
Type Distributed: Seed

Collected from: Michoacan Mexico  
Maintained by: Plant Genetic Resources Conservation Unit, Griffin, GA  
NPGS received: 07-Dec-1993  
PI assigned: 2001  
Inventory volume: 210  
Backup location: National Center for Genetic Resources Preservation  
Life form:  
Pedigree:  
Improvement status: Wild material  
Reproductive uniformity:  
Form received: Seed

Accession names and identifiers

DLP 4379
Type: Collector identifier
Grif 6177
Type: Site identifier
Comment: S-6
Mex-50
Type: Other or unclassified name

Annotations

Action	Date	By	Old Name	New Name
RE-IDENT	14 Nov 2000	Jamet, Robert L.	<i>Ipomoea triloba</i>	<i>Ipomoea trifida</i>

Source History

- Accession was collected: Nov-1993, Mexico  
Locality: Los Reyes D Salgado, Edo. Michoacan, Mexico; Elevation: 19 deg 21 min 0 sec North (19.35000000), Longitude: -102 deg 15 min 58 sec West (-102.26666667) GoogleMap & Elevation: 1250 meters
- Accession was donated.  
Donors:
  - Instituto del Paludo Center

# New sequencing technologies

## Next generation sequencing

- **Illumina (HiSeq)**
- Ion Torrent (Ion Proton)
- ABI/SOLiD
- **Roche/454**
- **Helicos**



- ❖ > 400 Gb per day
- ❖ Read length: 150 bp
- ❖ Error rate: <0.5%

## Third generation sequencing

- **Pacific Biosciences**
- Oxford Nanopore
- Complete Genomics



- ❖ 500 Mb in 10 hours
- ❖ Read length: up to 40 kb, average 10-15 kb
- ❖ Error rate: 15%

## Desktop sequencer

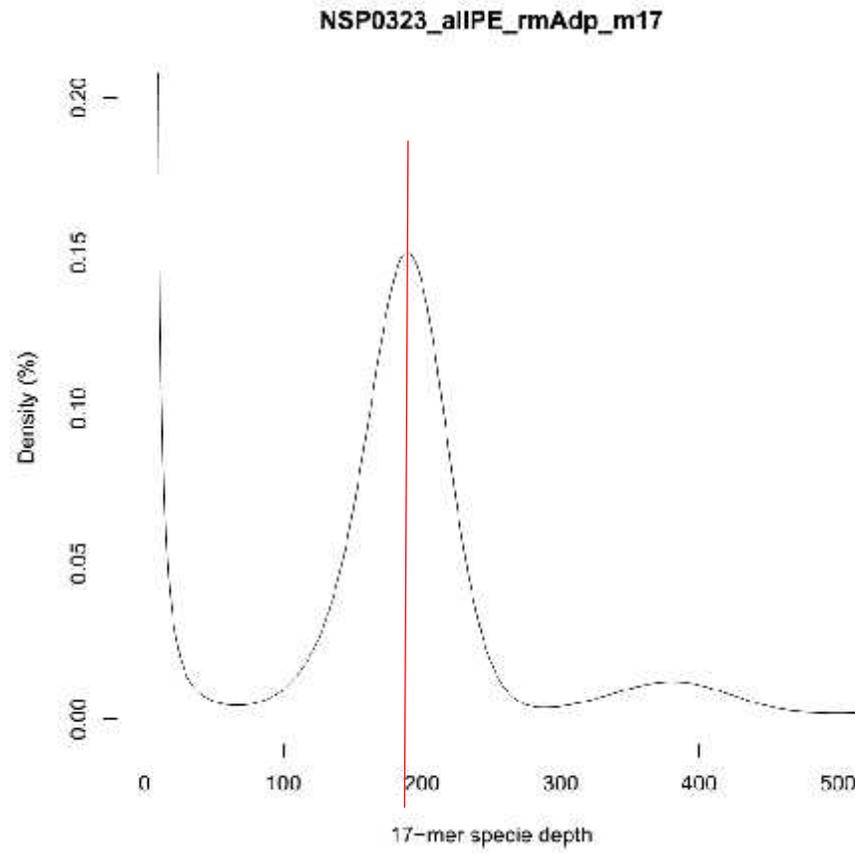
- Ion Torrent PGM
- Illumina MiSeq
- **454 GS Junior**

## Sequences generated for NCNSP-0323

Library	raw reads		high quality cleaned		
	read length	No. of bases	read length	No. of bases	depth
200 bp	101	94,061,140,875	96.80	45,198,018,055	91.14
500 bp	101	33,174,385,866	98.66	28,196,103,686	56.85
1 kb	150	29,385,863,700	142.48	26,281,109,493	52.99
2 kb	101	22,332,288,366	86.45	9,522,646,147	19.20
5 kb	101	21,778,758,068	86.18	8,867,701,122	17.88
10 kb	150	16,452,249,000	115.65	8,500,274,191	17.14
20 kb	150	20,168,044,500	111.43	8,783,475,429	17.71
PacBio	2,445	7,124,598,193			14.37
<b>Total</b>		<b>244,477,328,568</b>			<b>287.28</b>

In the processing to generating mate-pair reads from a 40 kb insert library

# k-mer distribution of NCNSP-0323 genome reads



## De novo assembly using Illumina reads

	Contig*		Scaffold*	
	Size (bp)	Number	Size (bp)	Number
N90	7,406	16,403	366,957	271
N80	12,394	11,938	737,508	188
N70	17,081	8,955	1,094,939	139
N60	21,859	6,711	1,468,492	105
N50	27,136	4,923	<b>1,818,409</b>	<b>77</b>
N25	44,938	1,759	3,070,329	27
N00 (Longest)	183,410	1	6,604,314	1
Total	434,914,015	36,644	447,272,757	5,007
*Only contigs and scaffolds >= 500 bp were included in the genome assembly				

**Estimated genome size: 494.1 Mb**

# Genome size of Ipomoea

Genetics Soc. Hort. Sci. 113 (1) 115-118, 1994.

## Nuclear DNA Content and Ploidy Levels in the Genus *Ipomoea*

Peggy Ocas-Akins

Department of Horticulture, University of Georgia Coastal Plain Experimental Station, Tifton, GA 31794

Robert L. Jarrett

U.S. Department of Agriculture, Agricultural Research Service, Regional Plant Introduction Station, Griffin, GA 30223

Table 1. Genotypes of *Ipomoea* species and mean DNA content determined by flow cytometry.

Genotype	<i>Ipomoea</i>	Origin	Source*	Ploidy	DNA content (pg/2C nucleus)
338264	<i>argillicola</i> R.W. Johnson	Australia	G	4x <sup>b</sup>	3.0
81.2	<i>batatas</i> (H.B.K.) G. Don	Ecuador	C	4x	3.5
518473	<i>batatas</i> (L.) Lam.	Mexico	G	4x	3.4
518474	<i>batatas</i> (L.) Lam.	Mexico	G	4x	3.0
518476	<i>batatas</i> (L.) Lam.	Mexico	G	4x	3.2
518478	<i>batatas</i> (L.) Lam.	Mexico	G	4x	3.2
DLP 5283	<i>batatas</i> (L.) Lam.	Ecuador	G	4x <sup>b</sup>	3.1
A7817	<i>batatas</i> (L.) Lam.	Ecuador	G	4x <sup>b</sup>	3.5
Red Jewel	<i>batatas</i> (L.) Lam.	United States	T	6x	4.8
Copper Resisto	<i>batatas</i> (L.) Lam.	United States	G	6x	5.2
Coastal Red	<i>batatas</i> (L.) Lam.	United States	T	6x	4.5
GA Red	<i>batatas</i> (L.) Lam.	United States	T	6x	4.6
62.39	<i>condata-eriloba</i> Demstedt	United States	C	2x	1.7
80.6	<i>condata-eriloba</i> Demstedt	United States	C	2x	1.7
540710	<i>condata-eriloba</i> Demstedt	Colombia	G	2x	1.7
72.2	<i>condata-eriloba</i> Demstedt	United States	C	2x	1.8
MC1310	<i>condata-eriloba</i> Demstedt	Mexico	G	2x	1.6
MC1961	<i>condata-eriloba</i> Demstedt	Mexico	G	4x	3.0
MC1984	<i>condata-eriloba</i> Demstedt	Mexico	G	4x	3.3
518495	<i>condata-eriloba</i> Demstedt	Mexico	G	4x	3.3
549093	<i>cyanocephala</i> Meisn.	Brazil	G	2x	1.7
63.36	<i>lacunosa</i> L.	United States	G	2x	1.7
85.27	<i>lacunosa</i> L.	United States	C	2x	1.5
62.82	<i>lacunosa</i> L.	United States	C	2x	1.6
78.19	<i>lacunosa</i> L.	United States	C	2x	1.8
536036	<i>oleocarpa</i> Jacq.	Mexico	G	2x	1.6
Q27809	<i>rotundifolia</i> Blume	Australia	G	2x	2.4
538274	<i>maifertii</i> Benth.	Australia	G	2x-4x <sup>b</sup>	2.7
540706	<i>nil</i> (L.) Roth	Colombia	G	2x	2.3
Violet	<i>nil</i> (L.) Roth	Japan	A	2x	2.3
530993	<i>obovata</i> (L.) Ker-Gawl.	Dominican Republic	G	2x-4x <sup>b</sup>	2.6
518482	<i>perfoliata</i> Mart.	Mexico	G	2x	1.6
549258	<i>per-nigrata</i> L.	Australia	G	2x	2.0
171664	<i>pupurea</i> (L.) Roth	Turkey	G	2x	2.0
87.3	<i>ramosissima</i> (Poir.) Choisy	Peru	C	2x	1.7
518483	sp.	Mexico	G	4x <sup>b</sup>	3.0
71.9	<i>signata</i> Poir.	United States	G	2x-4x <sup>b</sup>	2.7
17257	<i>nitrovariegata</i> R.W. Johnson	Australia	G	4x <sup>b</sup>	3.3
518479	<i>lobocarpa</i> McDonald & Austin	Mexico	G	4x	2.6
87.4	<i>tenaxissima</i> Choisy	United States	C	2x	1.8
518488	<i>tillyana</i> (Will.) Choisy	Mexico	G	4x	4.0
540723	<i>trifida</i> (H.B.K.) G. Don	Colombia	G	2x	1.7
74.1	<i>trifida</i> (H.B.K.) G. Don <sup>c</sup>	Unknown	C	2x	1.7
87.2	<i>trifida</i> (H.B.K.) G. Don <sup>c</sup>	Ecuador	C	4x	3.4
80.3	<i>trifida</i> (H.B.K.) G. Don <sup>c</sup>	Mexico	C	4x	3.4
79.5	<i>trifida</i> (H.B.K.) G. Don <sup>c</sup>	Mexico	C	4x	3.4
72.4	<i>trifida</i> (H.B.K.) G. Don <sup>c</sup>	Mexico	C	4x	3.6
530997	<i>tiloba</i> L.	Dominican Republic	G	2x	1.6
62.85	<i>tiloba</i> L.	Unknown	C	2x	1.5
67.66	<i>tiloba</i> L.	Caribbean	C	2x	1.6
530999	<i>tiloba</i> L.	Dominican Republic	G	2x	1.7
1989	<i>umbroscala</i> House	Mexico	G	2x	1.7
518496	<i>wrightii</i> A. Gray	Mexico	G	2x <sup>b</sup>	2.0

Plant Molecular Biology Reporter

Volume 9(3) 1991

pages 208-218

## Genetic Resources

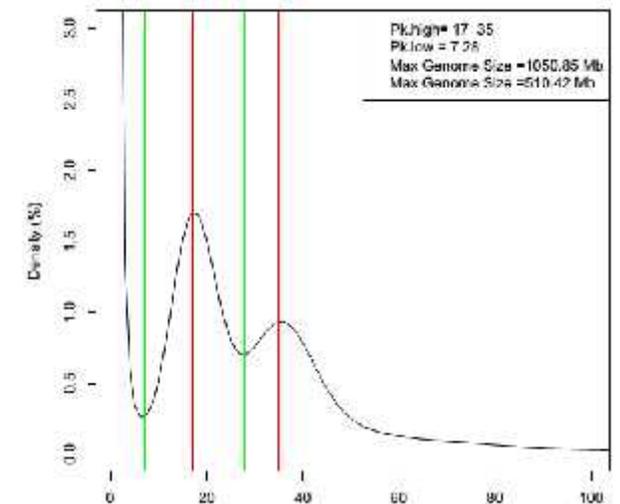
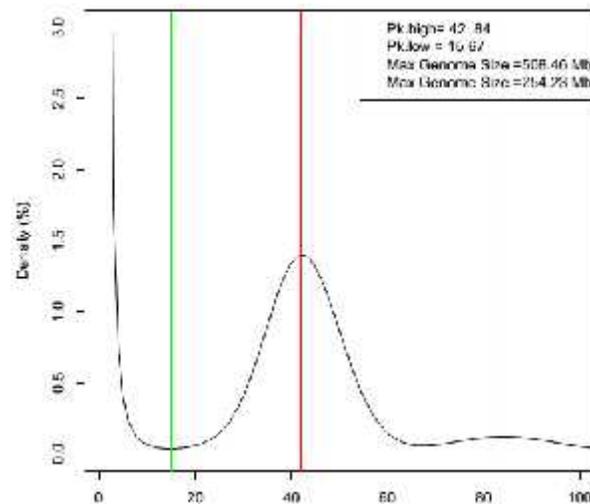
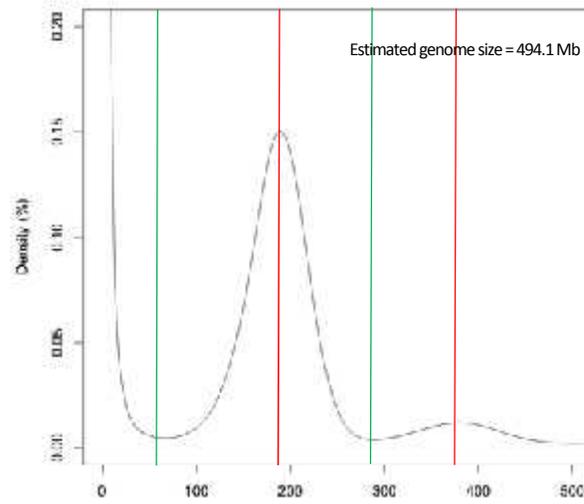
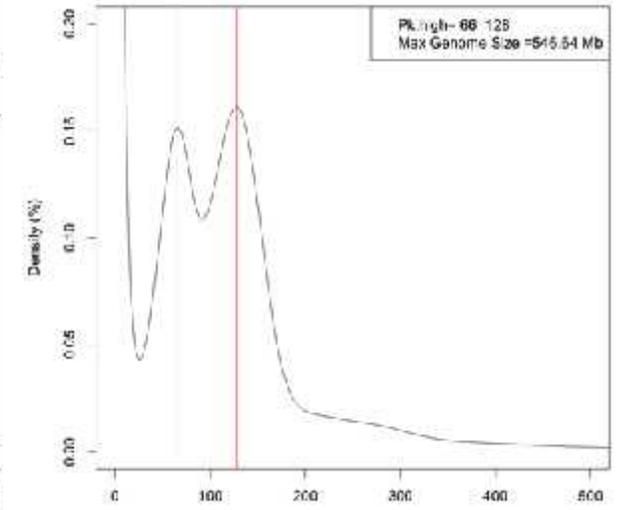
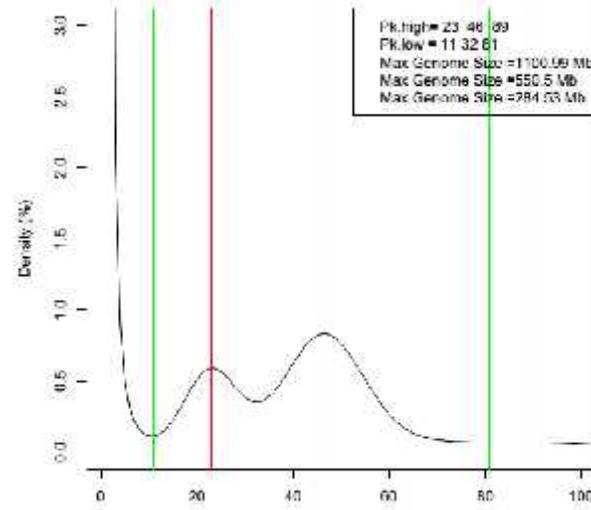
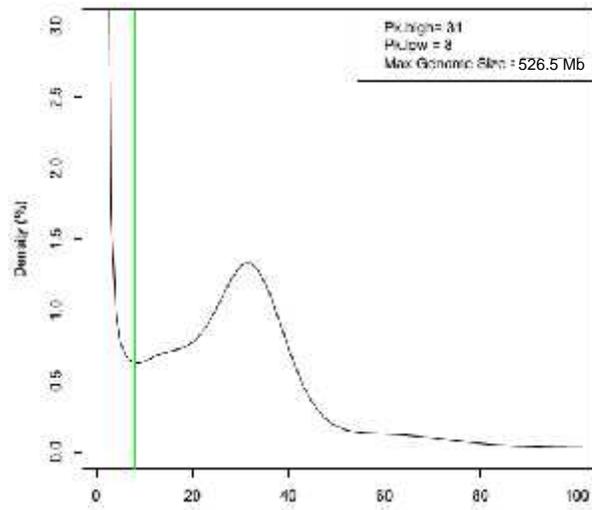
## Nuclear DNA Content of Some Important Plant Species

K. Arumuganathan and E. D. Earle\*

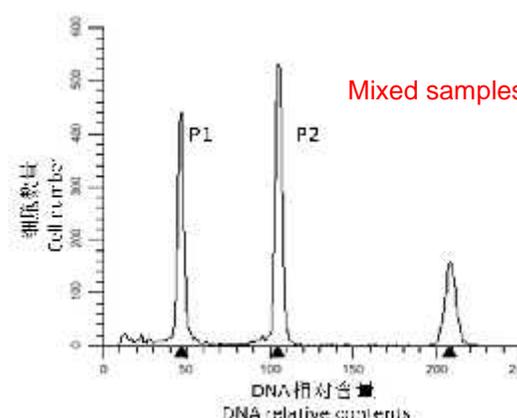
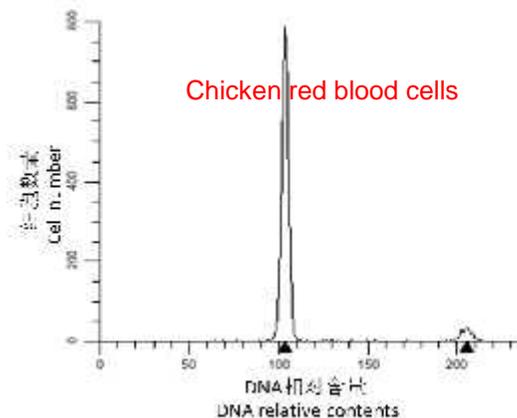
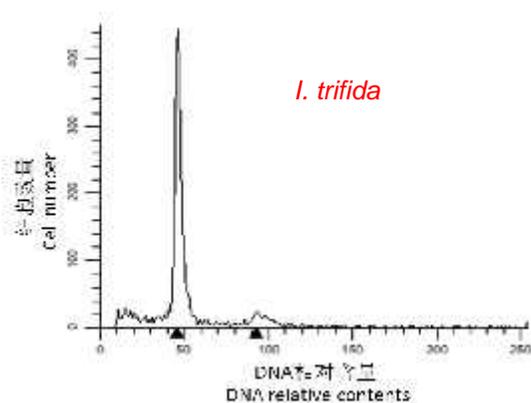
Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853-1902, USA.

Scientific name	Common name	Family	pg/2C <sup>a</sup> (N)	-Mbp <sup>b</sup> /1C
<i>Carica papaya</i>	Papaya	Caricaceae	0.77	372
<i>Cicer arietinum</i>	Chick pea	Leguminosae	1.53	738
<i>Citrullus vulgaris</i> (= <i>lanatus</i> )	Watermelon	Cucurbitaceae	0.88, 0.90 (2)	425, 434
<i>Citrus sinensis</i>	Orange	Rutaceae	0.76, 0.82 (2)	367, 396
<i>Crepis capillaris</i>	Crepis	Compositae	3.87	1867
<i>Cucumis melo</i>	Cantaloupe	Cucurbitaceae	0.94, 1.04 (2)	454, 502
<i>Cucumis sativus</i>	Cucumber	Cucurbitaceae	0.76	367
<i>Cucurbita pepo</i>	Zucchini	Cucurbitaceae	1.04, 1.08 (2)	502, 521
<i>Datura stramonium</i>	Jimson weed	Solanaceae	4.11	1983
<i>Daucus carota</i>	Carrot	Umbelliferae	0.98	473
<i>Dioscorea alata</i>	Yam	Dioscoreaceae	1.15	555
<i>Diplontaxis erucoides</i>		Cruciferae	1.31	632
<i>Eruca sativa</i>		Cruciferae	1.16	560
<i>Glycine max</i> (2n=4X)	Soybean	Leguminosae	2.31	1115
<i>Gossypium hirsutum</i> (2n=4X)	Cotton	Malvaceae	4.39, 4.92 (2)	2118, 2374
<i>Helianthus annuus</i>	Sunflower	Compositae	5.95-6.61 (3)	2871-3189
<i>Hordeum vulgare</i>	Barley	Gramineae	10.10	4873
<i>Ipomoea batatas</i> (2n=6X)	Sweet potato	Convolvulaceae	3.31	1597
<i>Lactuca sativa</i>	Lettuce	Compositae	3.47	2639
<i>Lens culinaris</i> (= <i>esculentia</i> )	Lentil	Leguminosae	8.42	4063
<i>Lycopersicon chersonesum</i>		Solanaceae	1.83	883
<i>Lycopersicon esculentum</i>	Tomato	Solanaceae	1.88-2.07 (6)	907-1000
<i>Lycopersicon pennellii</i>		Solanaceae	2.47-2.77 (3)	1192-1337
<i>Lycopersicon peruvianum</i>		Solanaceae	2.27	1095
<i>Malus x domestica</i> (2n=2X)	Apple	Rosaceae	11.54-1.65 (3)	743-796
<i>Mangifera indica</i>	Mango	Anacardiaceae	0.91	439

# Genome size of Ipomoea



# Genome size of Ipomoea



Sample number	Peak value of <i>I. trifida</i> (A)	Peak value of CEN (B)	A/B	pg/2C	Mb/1C
1	42.12	103.39	0.407389496	1.01847374	499.1
2	44.26	105.10	0.42112274	1.05280685	515.9
3	45.45	103.53	0.439003187	1.097507968	537.8
4	46.94	104.79	0.447943506	1.119858765	548.7
5	46.53	105.07	0.442847625	1.107119063	542.5
6	46.29	103.70	0.446383799	1.115959498	546.8
7	45.99	104.57	0.43980109	1.099502725	538.8
Average	45.368 6	104.307	0.434927349	1.087318373	532.8

Provided by Dr. Qinghe Cao at Xuzhou Sweetpotato Research Institute



# Hexaploid sweetpotato



- Huachano, a Peruvian landrace which is amenable to genetic transformation
- About 80G raw sequence data was generated using the Illumina HiSeq 2000 system

## Raw data

~246 M read pairs in 200 bp library.

~152 M read pairs in 500 bp library.

Total length 80,317 Mb

## Cleaned data

	# paired	# Single	Read size	Total
SP200	185 M	32 M	94 bp	37.68 Gb
SP500	127 M	17 M	90 bp	24.36 Gb
Total	312 M	49 M	92 bp	62.04 Gb

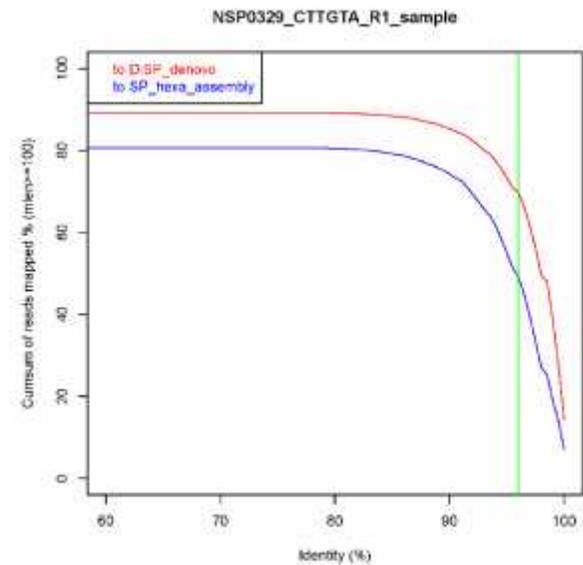
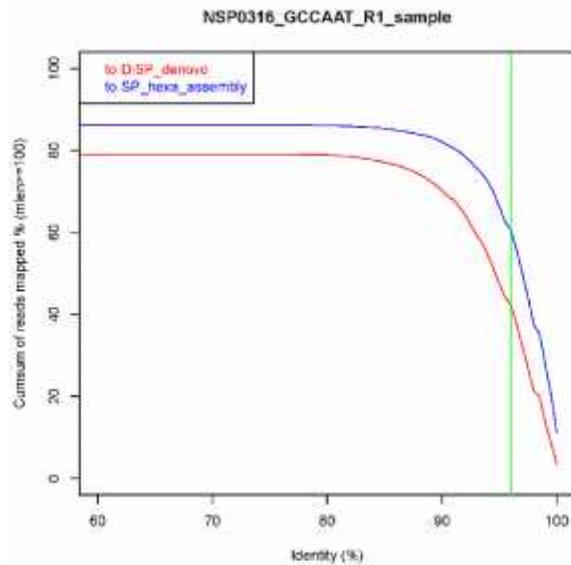
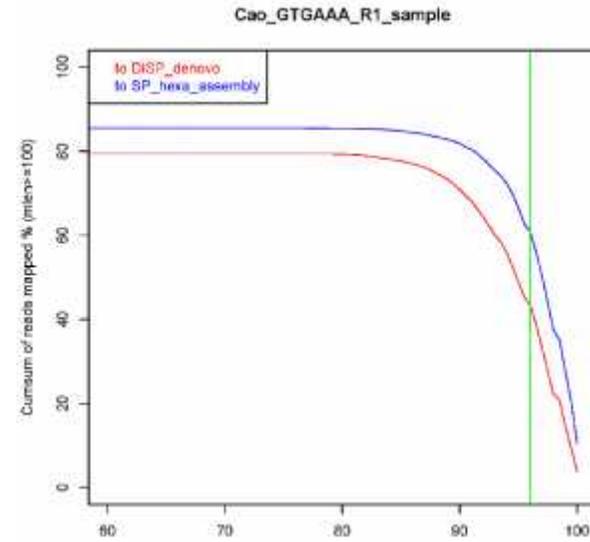
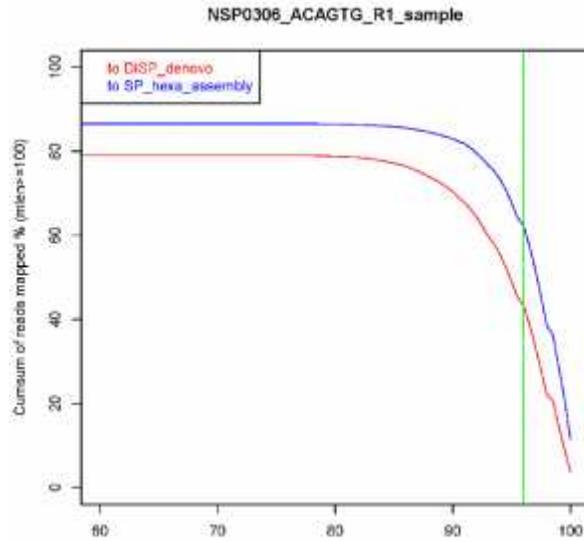
# Hexaploid sweetpotato

## De novo assembly

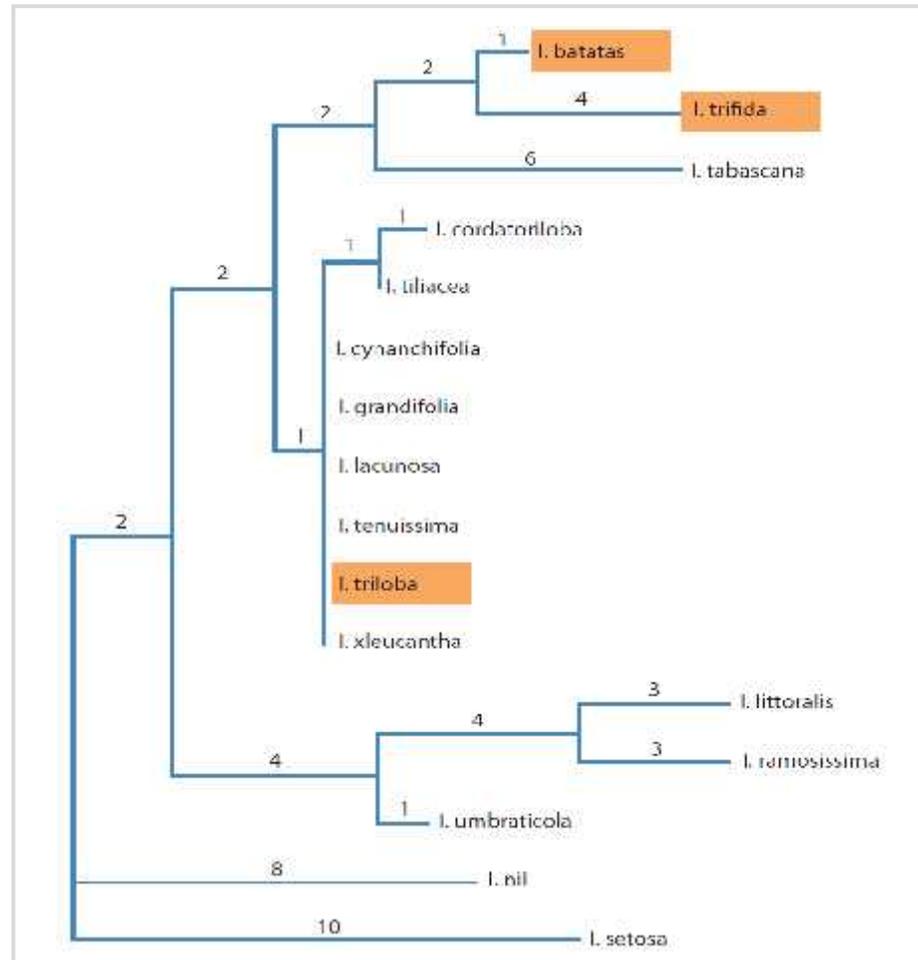
Scaffolds >= 200 bp (GC% = 38.35%)				
	Contig		Scaffold	
	Size (bp)	Index	Size (bp)	Index
<b>N90</b>	236	737,451	282	538,782
<b>N80</b>	302	552,738	421	392,362
<b>N70</b>	382	407,767	563	292,438
<b>N60</b>	480	292,372	701	212,612
<b>N50</b>	626	202,185	903	149,749
<b>N25</b>	1,267	59,915	1,695	46,336
<b>Largest</b>	19,628	1	21,622	1
<b>Total</b>	492,615,538	998,299	<b>498,123,765</b>	751,346

# Align resequencing reads to assembled genomes

**NCNSP-0323 is a triloba!**

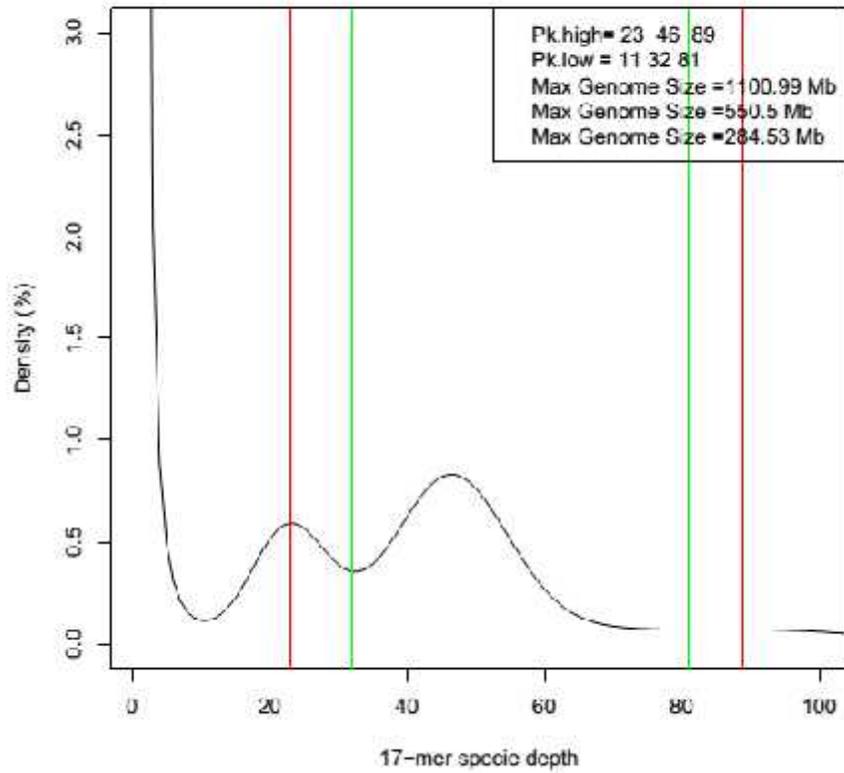


# Phylogeny of Ipomoea

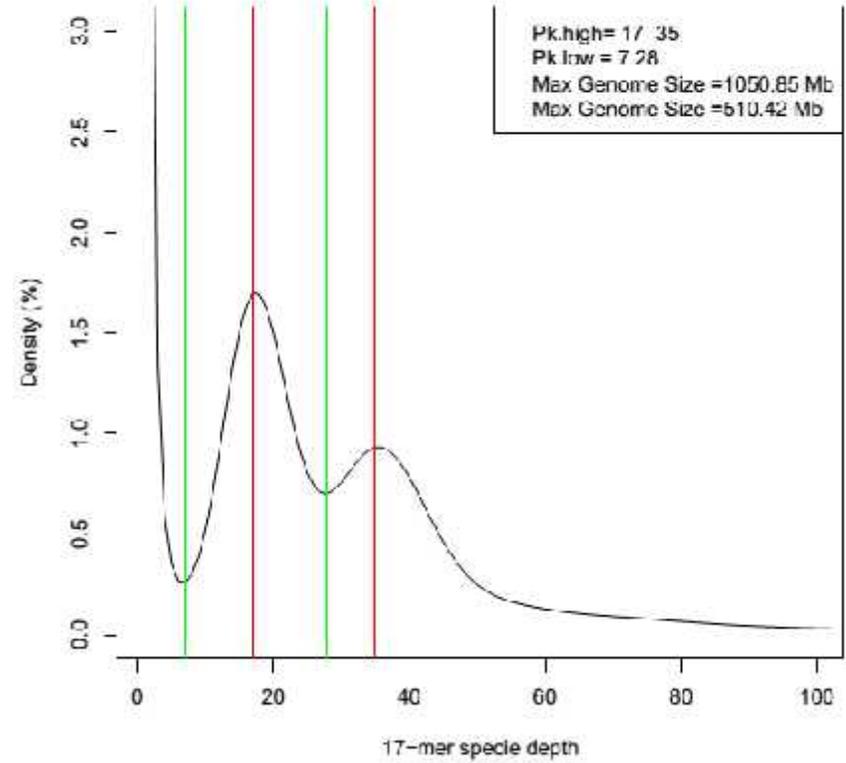


# K-mer distribution

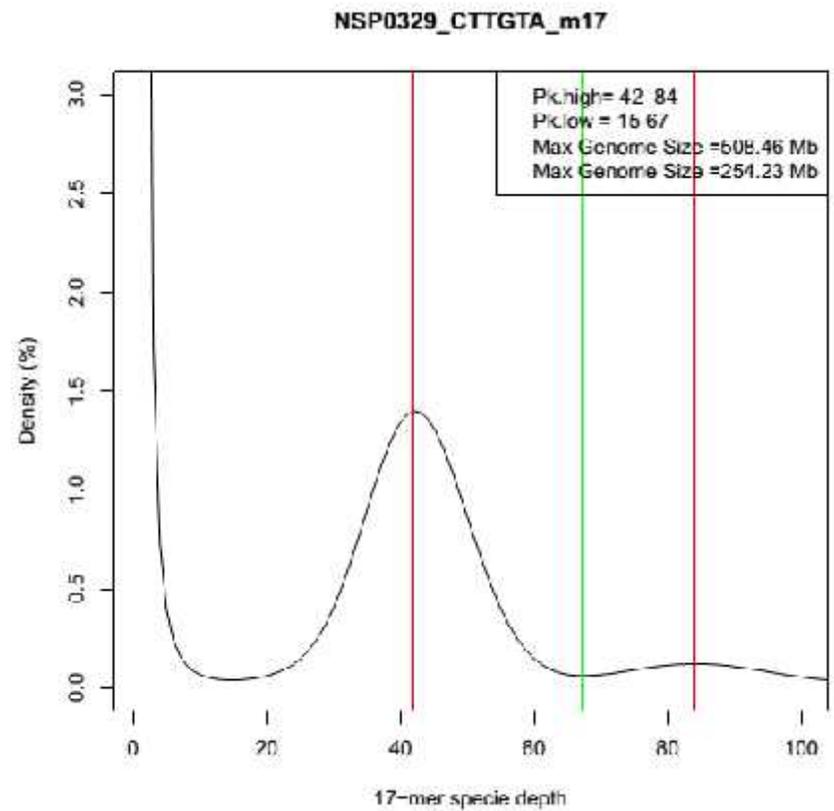
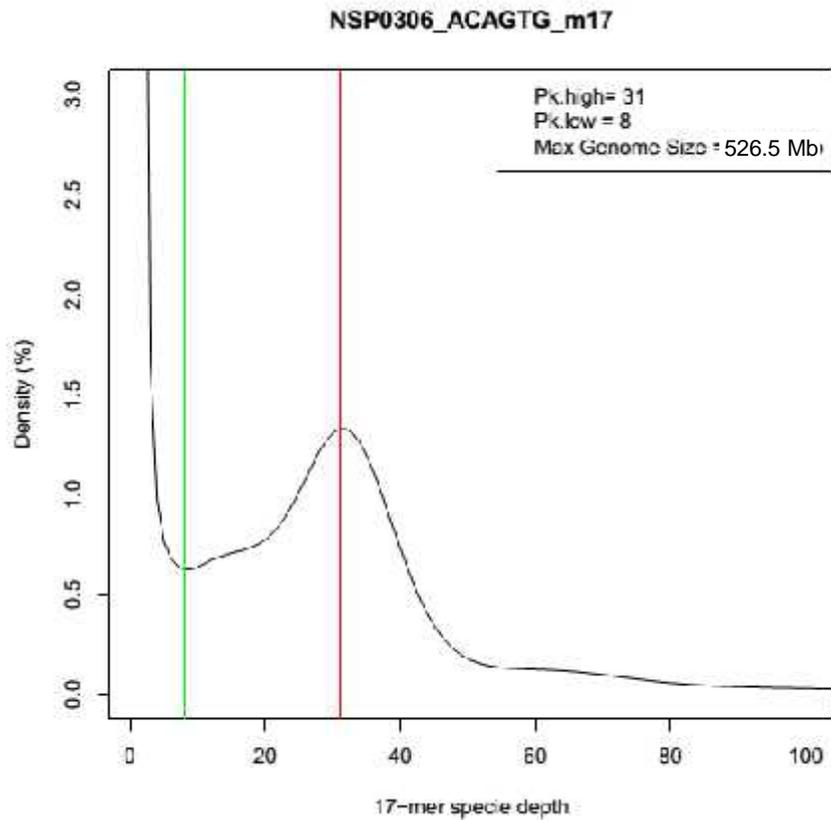
NSP0316\_GCCAAT\_m17



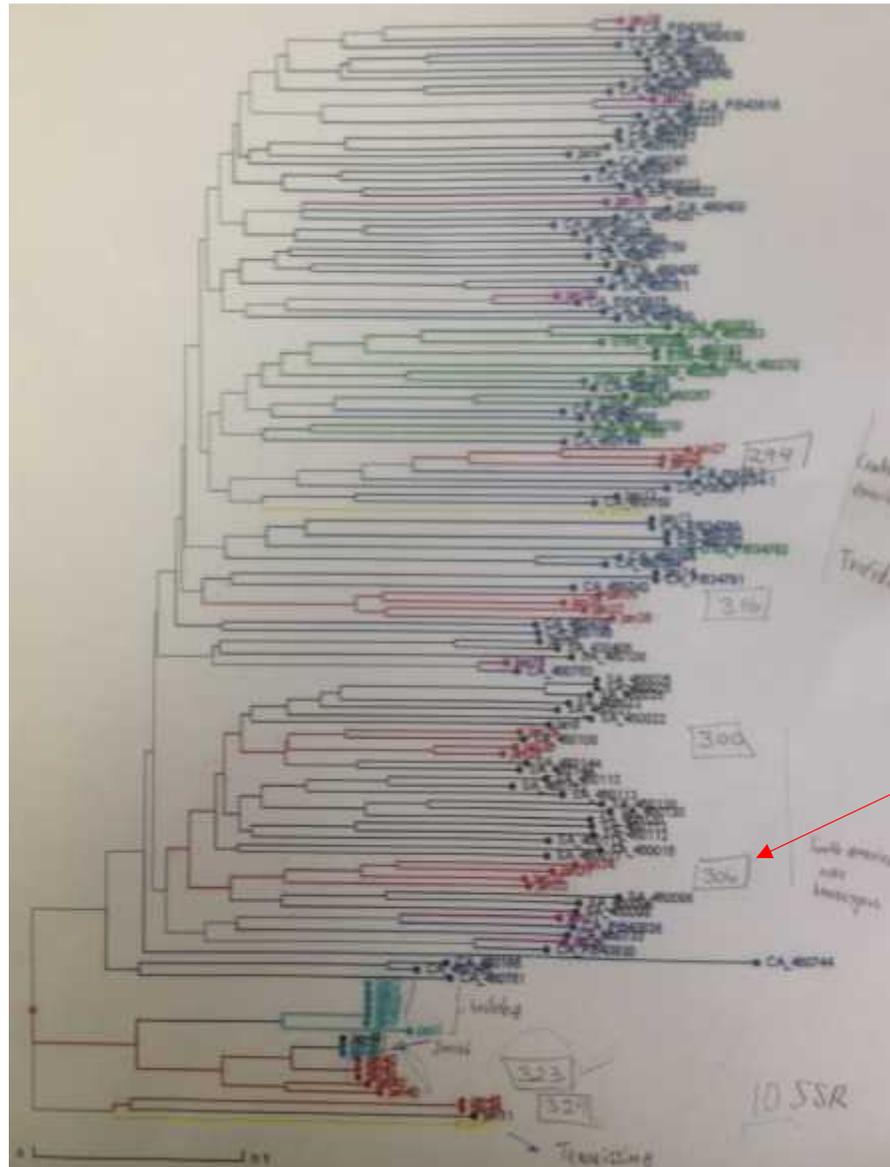
Cao\_GTGAAA\_m17



# K-mer distribution



NCNSP-0306 is a trifida



# NCNSP-0306

Accession: PI 540724 - GRI...  
npgsweb.ars-grin.gov/gringlobal/accessiondetail.aspx?id=1435660

PI 540724

**Ipomoea trifida (Kunth) G. Don**

Status: Available  
Amt Distributed: 10 count  
Type Distributed: Seed

[Add to My Favorites](#)  
[Add to Order](#)

Collected from:	Magdalena Colombia
Maintained by:	<a href="#">Plant Genetic Resources Conservation Unit, Griffin, GA</a>
NPGS received:	01-Dec-1987
PI assigned:	1990
Inventory volume:	199
Backup location:	<a href="#">National Center for Genetic Resources Preservation</a>
Life form:	
Pedigree:	
Improvement status:	Wild material
Reproductive uniformity:	
Form received:	Seed

**Accession names and identifiers**

	CIP 460137
Type:	Donor identifier

	DLP 2998
Type:	Collector identifier

**Source History**

- Accession was collected, Colombia  
Locality: 29km S Fundacion-Copey, Fundacion Province. Latitude: 10 deg. 15 min. 0 sec. North (10.25000000), Longitude: -74 deg. -4 min. -59 sec. West (-74.08333333) [GoogleMap](#) it. Elevation: 160 meters.  
Collectors:
  1. [Jarrel, Robert L., USDA, ARS](#)
  2. [de La Puente, F., International Potato Center](#)
- Accession was donated, Dec-1987, Peru  
Donors:
  1. [International Potato Center](#)

# Sequences generated for NCNSP-0306

## Illumina

Library	Raw read			Final cleaned data			coverage
	length	No. read pair	Total bases	No. read pair	Length	Total bases	
300bp	160	159,099,305	50,911,777,600	141,356,152	156.86	44,347,028,370	84.3
300bp	160	89,626,008	28,680,322,560	81,356,804	156.69	25,494,814,725	48.5
500bp	160	61,736,032	19,755,530,240	58,619,296	156.04	18,294,336,035	34.8
500bp	160	67,694,044	21,662,094,080	61,862,235	154.42	19,105,508,952	36.3
500bp	160	200,763,012	64,244,163,840	345,242,652	147.72	50,998,566,404	97.0
1kb	160	110,274,832	35,287,946,240	100,356,563	153.76	30,861,049,402	58.7
5kb	160	74,240,204	23,756,865,280	54,558,279	122.92	13,412,601,755	25.5
8-10kb	160	97,459,912	31,187,171,840	60,177,282	118.99	14,320,897,658	27.2
15-20kb	160	148,892,802	47,645,696,640	42,590,255	122.24	10,412,280,197	19.8
30-40kb	30-100	120,890,886	22,433,505,116	31,162,162	89.92	5,603,954,875	10.7

## PacBio

Read number	Bases	Average read length	Read length range
2,665,738	11,909,256,312	4,467.53	35-37430

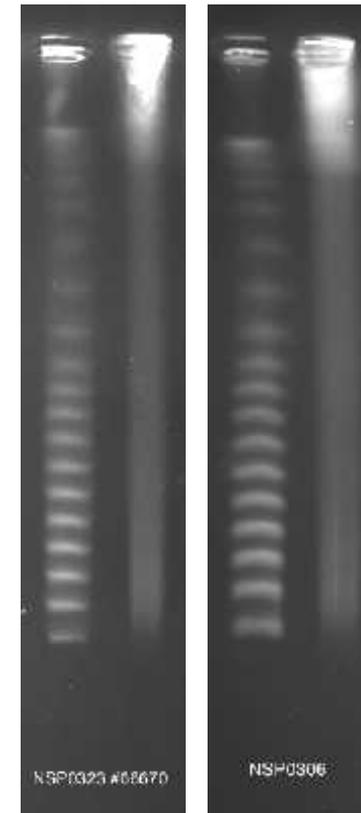
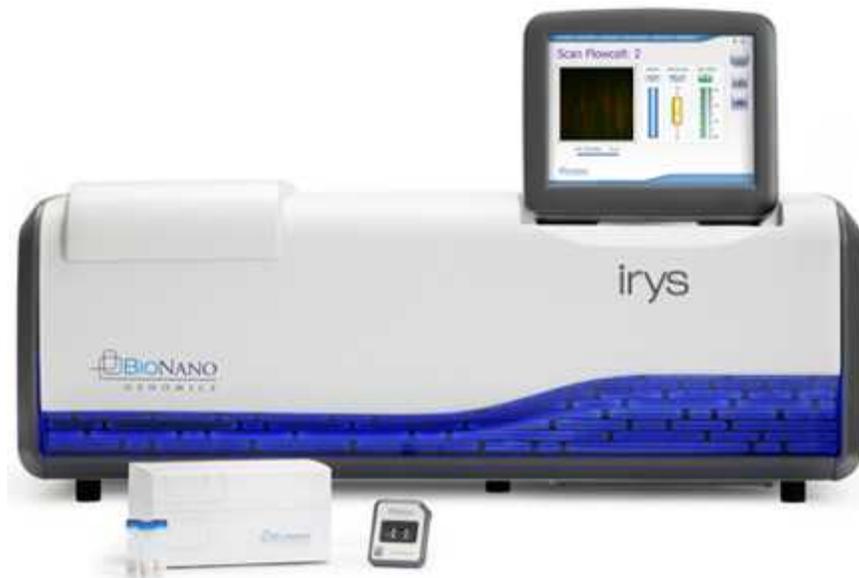
# Assembly of NCNSP-0306 genome

	Contig*		Scaffold*	
	Size (bp)	Number	Size (bp)	Number
N90	2,588	25,385	6,188	1908
N80	6,439	15,308	216,263	291
N70	11,077	10,451	590,894	172
N60	16,040	7,366	998,788	116
N50	21,531	5,149	1,377,288	77
N25	41,622	1,667	3,003,258	23
N00 (Longest)	262,642	1	10,211,095	1
Total	412,345,753	63,286	447,735,145	34,330
*Only contigs and scaffolds $\geq$ 500 bp were included in the genome assembly				

PacBio reads are being used to improve the assembly

# Build genome map using BioNano technology

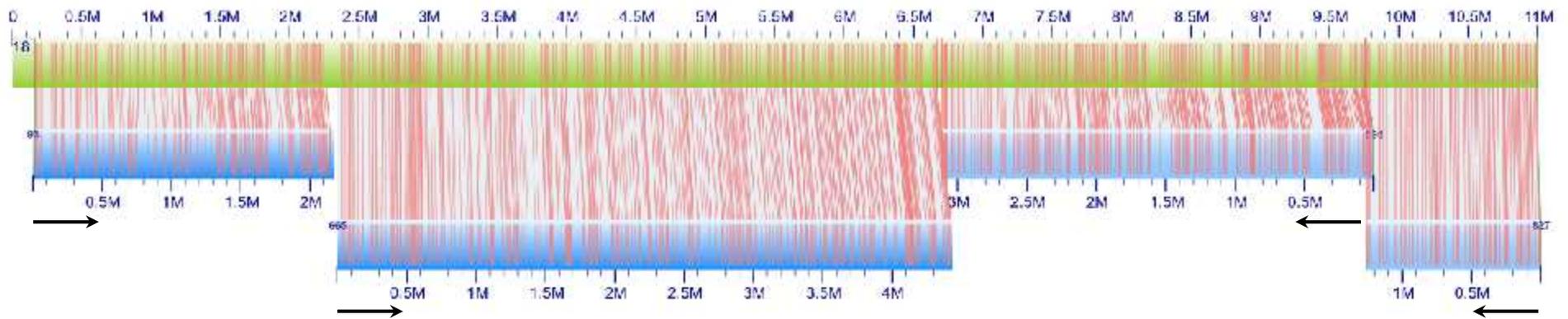
- ❖ generating whole genome maps by labeling megabase-scale genomic DNA fragments



## Initial BioNano genome (BNG) maps of *Ipomoea triloba* (NSP323)

- Total length of BNG maps: 453 Mb
- No. of BNG maps: 242
- N50 of BNG maps: 2.54 Mb

# Orders and orientations of the NGS scaffolds: case 1



BNG map 18 (11 Mb) bridges four NGS scaffolds:

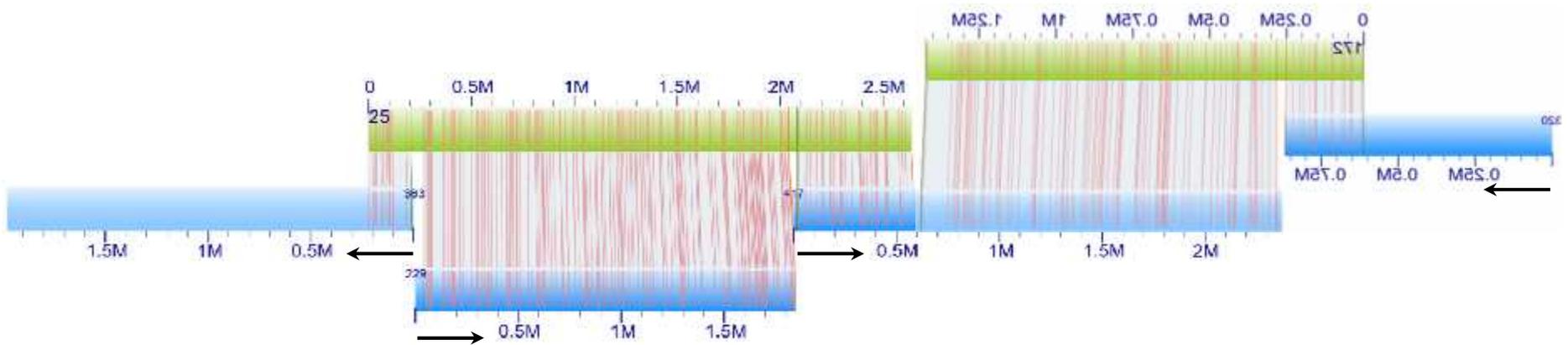
93 = scaffold7

665 = scaffold342

595 = scaffold514

827 = scaffold817

## Orders and orientations of the NGS scaffolds: case 2



BNG maps and NGS scaffolds mutually bridging.

BNG map 25 bridges three NGS scaffolds, and BNG map 172 bridges two:

363 = scaffold370

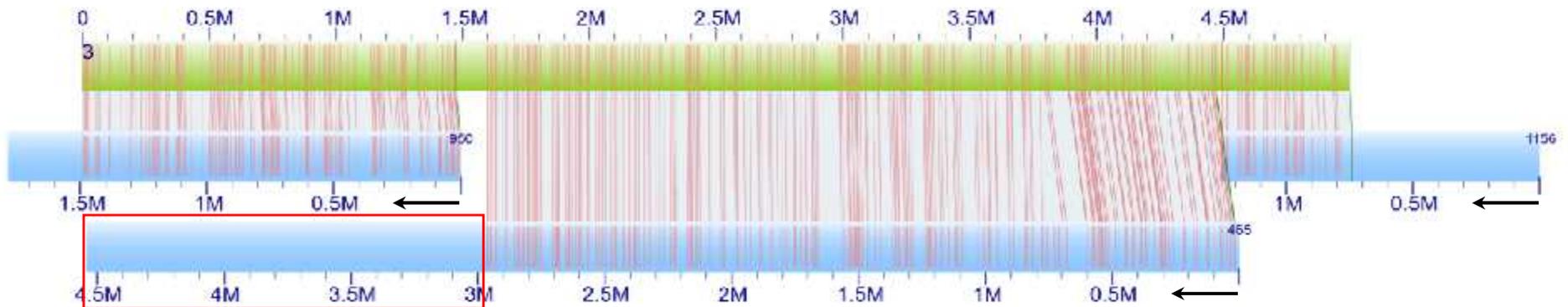
229 = scaffold228

417 = scaffold318

320 = scaffold326

Meanwhile, NGS scaffold318 bridges BNG maps 25 and 172 that a 7 Mb super scaffold can be gained.

## Mis-assembled scaffolds: case 1



BNG map 3 (5 Mb) bridges three NGS scaffolds:

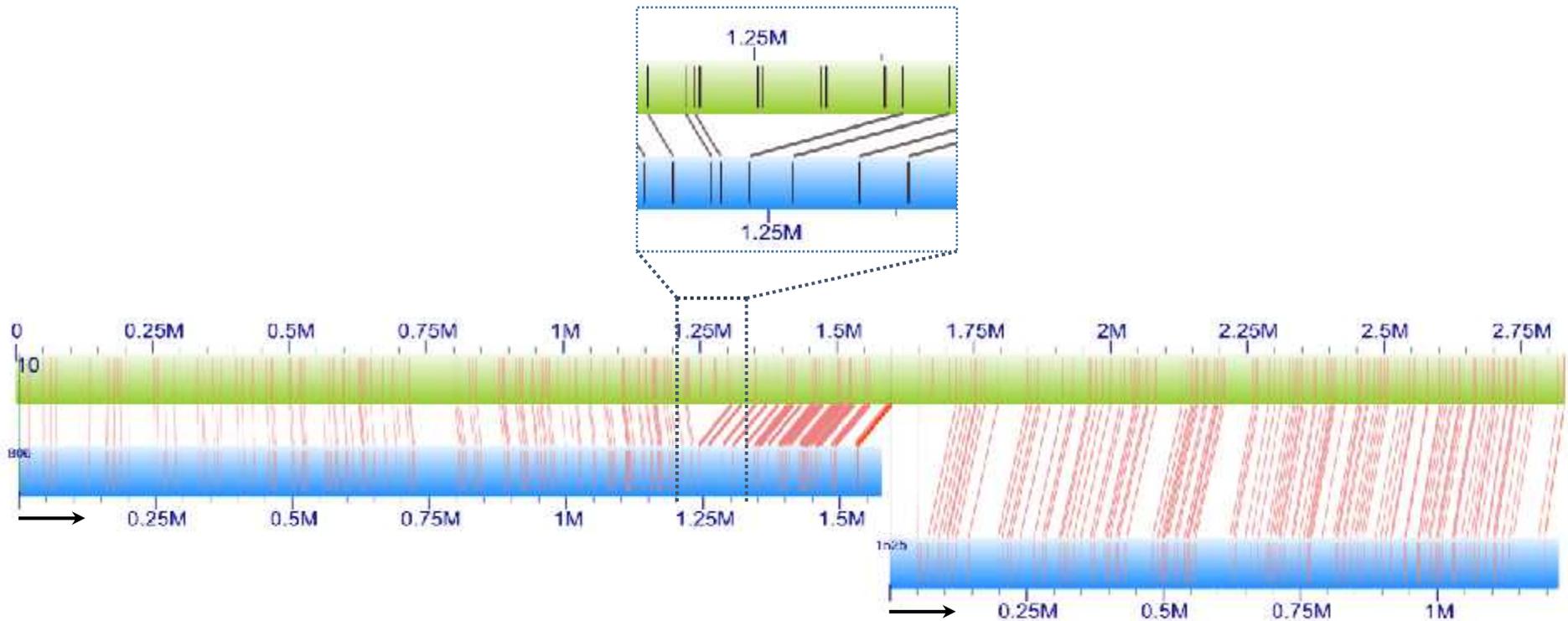
950 = scaffold904

465 = scaffold335

1156 = scaffold987

Only partial scaffolds465 could be aligned to BNG map 3, indicating a mis-assembly (red box).

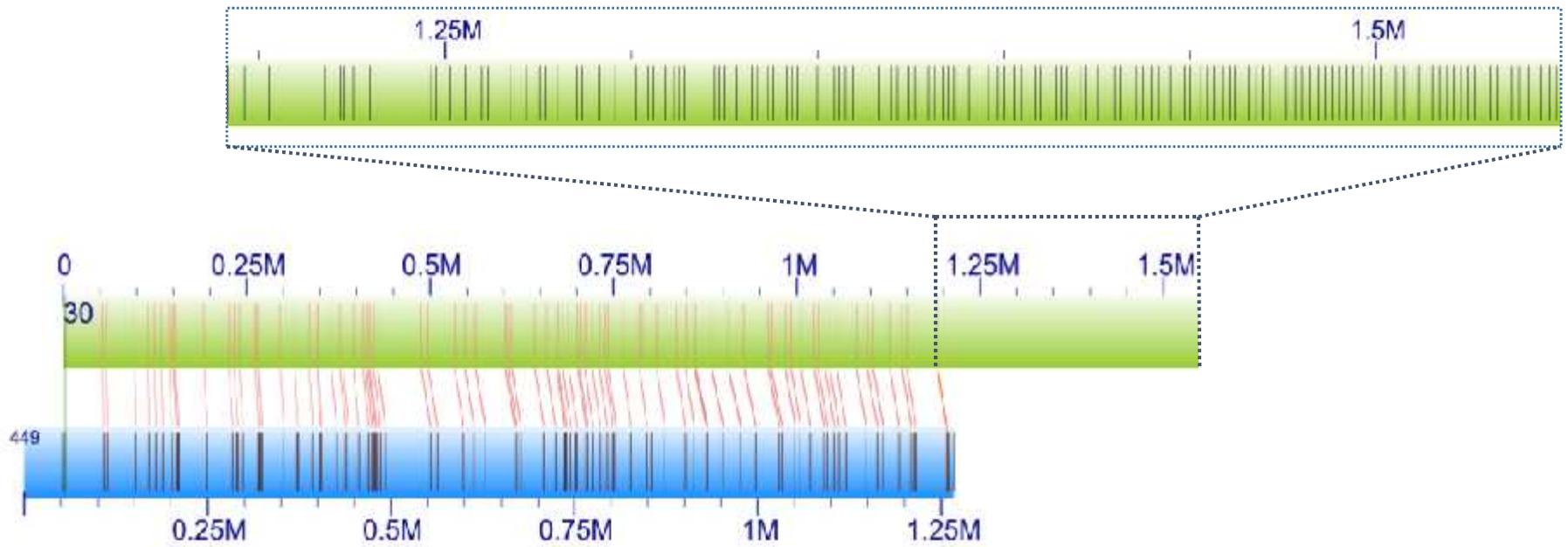
## Mis-assembled scaffolds: case 2



806 = scaffold744  
1525 = scaffold971

There's a missing part in the scaffold744 (probably repeats) and also in scaffold971.

## Region of repeats not covered by the NGS scaffolds



449 = scaffold461

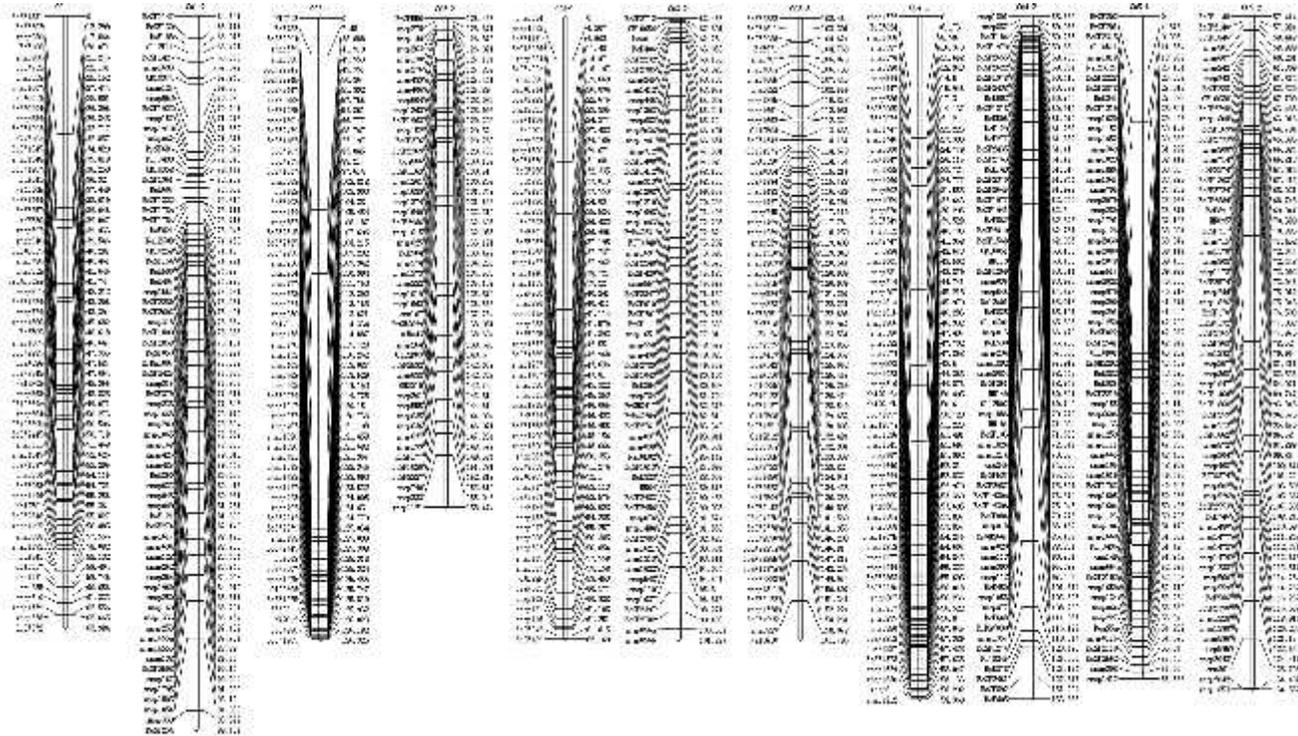
No NGS scaffolds can be aligned to the right part of BNG map 30 (~350 kb).  
This region contains only repeats.

# High density genetic map construction

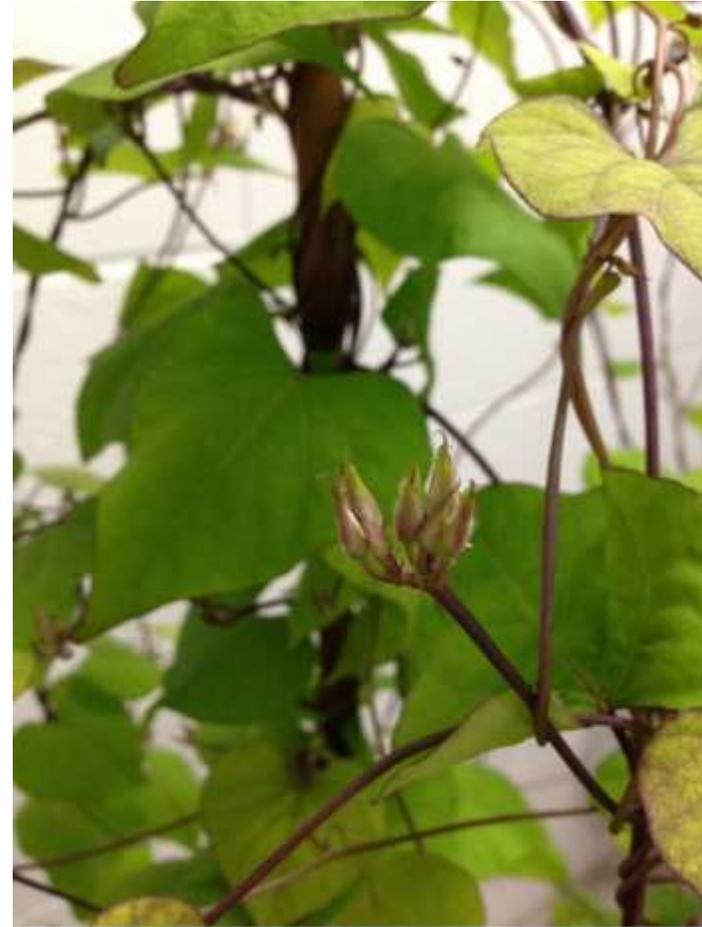
## Anchoring and ordering assembled scaffolds

Mapping population development

- ❖ CIP (Awais Khan)
- ❖ Xuzhou Sweetpotato Research Center



# Transcriptome sequencing and genome annotation



# Callus tissue generation

Joyce Van Eck at BTI



# RNA-Seq library construction

- ❖ Leaves
- ❖ Stem
- ❖ Flower bud
- ❖ Open flower
- ❖ Callus



## Other sweetpotato sequencing activities

- ❖ *I. trifida* genome sequencing by Robert Jarret at USDA
- ❖ *I. leucantha* and *I. lacunose* genome sequencing by Mark Rausher at Duke
- ❖ Cultivated sweetpotato and *I. trifida* genome sequencing by Asian Consortium
- ❖ Allen van Deyne at UC Davis
- ❖ African Orphan Crops Initiative

# *I. trifida* genome sequencing

by Bob Jarret and Jim Leebens-Mack

The screenshot shows a web browser window with the URL [npgsweb.ars-grin.gov/grn/global/accessiondetail.aspx?id=1456516](http://npgsweb.ars-grin.gov/grn/global/accessiondetail.aspx?id=1456516). The page displays the following information:

**PI 561544**  
**Ipomoea trifida (Kunth) G. Don**

Collected from:	Venezuela
Maintained by:	Plant Genetic Resources Conservation Unit, Griffin, GA
NPGS received:	02-Oct-1990
PI assigned:	1992
Inventory volume:	201
Backup location:	<a href="#">National Center for Genetic Resources Preservation</a>
Life form:	
Pedigree:	
Improvement status:	Cultivated material
Reproductive uniformity:	
Form received:	Seed

Status: Available  
Amt Distributed: 10 count  
Type Distributed: Seed

[Add to My Favorites](#)  
[Add to Order](#)

**Accession names and identifiers**

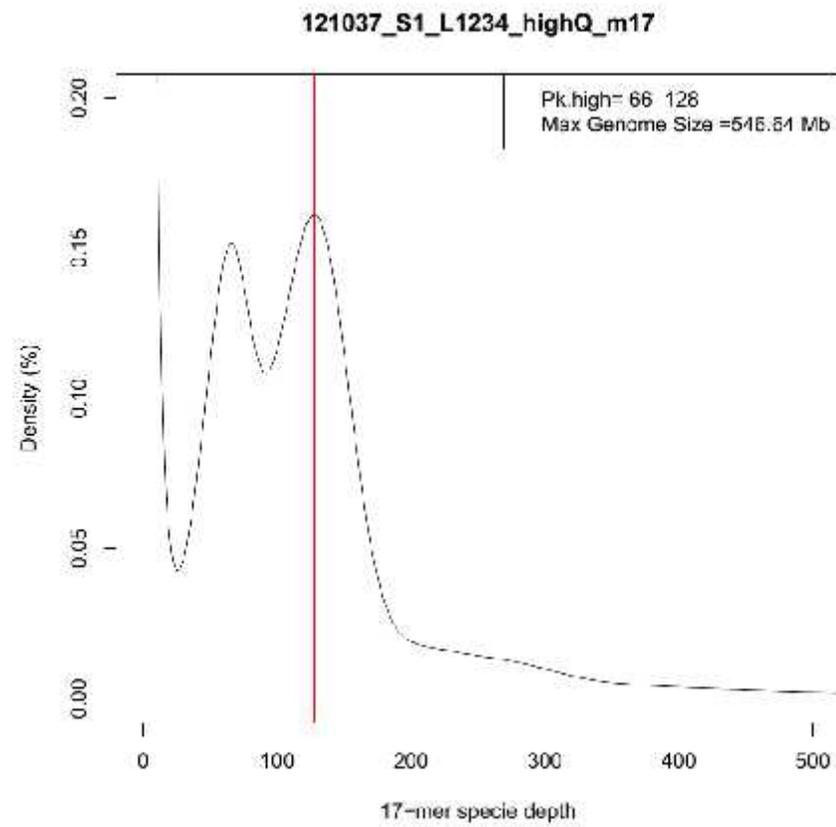
CIP 460021
------------

Type: Donor identifier

**Source History**

- Accession was collected, Venezuela
- Accession was donated, 02-Oct-1990, Peru  
Donors:
  1. de La Puente, F., International Potato Center

# k-mer distribution of PI 561544



# Asian Consortium TRAS (Trilateral Research Association of Sweetpotato)

DNA Research Advance Access published March 24, 2015

*DNA Research*, 2015, 1–9  
doi: 10.1093/dnares/dsv002  
Full Paper

OXFORD

---

Full Paper

## Survey of genome sequences in a wild sweet potato, *Ipomoea trifida* (H. B. K.) G. Don

Hideki Hirakawa<sup>1,†</sup>, Yoshihiro Okada<sup>2,†</sup>, Hiroaki Tabuchi<sup>3</sup>,  
Kenta Shirasawa<sup>1</sup>, Akiko Watanabe<sup>1</sup>, Hisano Tsuruoka<sup>1</sup>, Chiharu Minami<sup>1</sup>,  
Shinobu Nakayama<sup>1</sup>, Shigemi Sasamoto<sup>1</sup>, Mitsuyo Kohara<sup>1</sup>,  
Yoshie Kishida<sup>1</sup>, Tsunakazu Fujishiro<sup>1</sup>, Midori Kato<sup>1</sup>, Keiko Nanri<sup>1</sup>,  
Akiko Komaki<sup>1</sup>, Masaru Yoshinaga<sup>3,‡</sup>, Yasuhiro Takahata<sup>3</sup>,  
Masaru Tanaka<sup>3</sup>, Satoshi Tabata<sup>1</sup>, and Sachiko N. Isobe<sup>1,\*</sup>

<sup>1</sup>Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan, <sup>2</sup>Crop and Agribusiness Research Division, Kyushu Okinawa Agricultural Research Center, National Agriculture and Food Research Organization (NARO/KARC), Itoman, Okinawa 901-0336, Japan, and <sup>3</sup>Upland Farming Research Division, Kyushu Okinawa Agricultural Research Center, National Agriculture and Food Research Organization (NARO/KARC), Miyakonojo, Miyazaki 885-0091, Japan

\*To whom correspondence should be addressed. Tel. +81 438-52-3928. Fax. +81 438-52-3934. E-mail: sisobe@kazusa.or.jp

<sup>†</sup>These authors contributed equally to this work.

<sup>‡</sup>Present Address: NARO Headquarters, Kannondai 3-1-1, Tsukuba, Ibaraki, 305-8517, Japan.

Edited by Dr Katsumi Isono

Received 25 November 2014; Accepted 17 February 2015

**Table 1.** Statistics of the assembled genome sequences for Mx23Hm and 0431-1

Sequenced line	Mx23Hm (ITR_r1.0)	0431-1 (ITRk_r1.0)
Number of sequences	77,400	181,194
Total length (bases)	512,990,885	712,155,587
Average length (bases)	6,628	3,930
Max length (bases)	910,847	1,352,076
Min length (bases)	300	300
N50 length (bases)	42,586	36,283
A	108,919,552	155,339,270
T	108,380,339	154,432,148
G	60,024,339	86,821,603
C	60,253,902	87,276,414
N	175,412,753	228,286,152
Total (ATGC)	337,578,132	483,869,435
GC% (GC/ATGC)	35.6	36.0