BIODIVERSITY GENOMICS: APPLICATIONS FOR CLASSIFICATION OF CROP WILD RELATIVES

PRETORIA 23 NOVEMBER 2016

IN SITU CONSERVATION OF CWR IN SADC COUNTRIES

UNIVERSITY OF MAURITIUS AND MINISTRY OF AGRO INDUSTRY AND FOOD SECURITY

CWR CLASSIFICATION

GENE POOL CONCEPT

- primary gene pool (GP1),
  - the crop and all the closely related taxa, which can interbreed with the crop and produce fertile progenies - GP1A (cultivated form) and GP1B (wild or weedy forms of the crop)
- GP2

  - taxa distinct from the crop, but still capable of crossing with it and producing some fertile hybrids
- GP3
  - includes taxa distantly related to the crop

  And cannot interbreed with the crop.

Taxon group: based on taxonomic hierarchy

- Taxon group 1 (TG1) includes the crop and taxa that of the same species;
- TG2 includes taxa belonging to the same series or section as the crop;
- TG3 includes taxa of the same subgenus as the crop;
- TG4 includes taxa belonging to the same genus as the crop;
- TG5 includes taxa belonging to the same tribe as the crop, but to a different genus

Genetic diversity

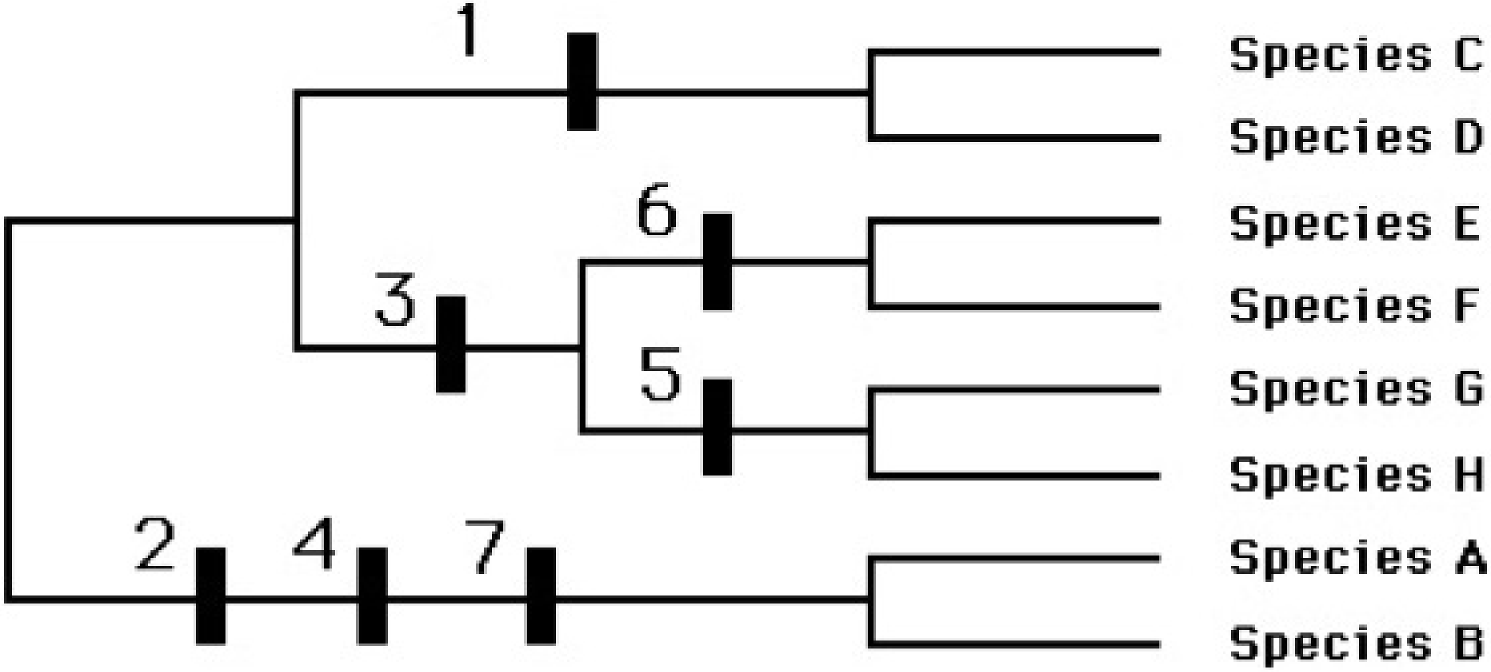
- All crop plants orginated from wild relative(s) that were domesticated-
- Which ones are they ? Are they still around?
- Vincent et al. (2013) -1667 wild species for 173 crop species

© Christopher Kaiser-Bunbur

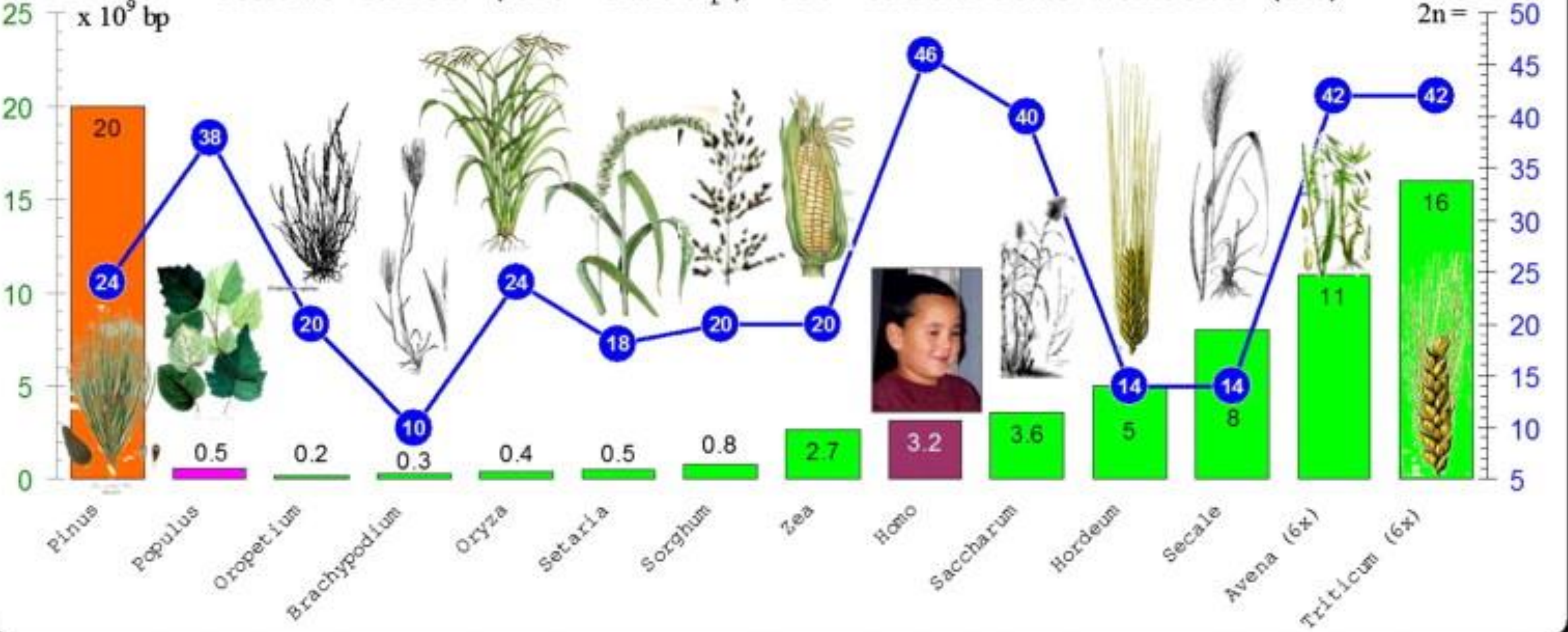
Biodiversity Genomics: using geneomic data to assess biodiversity

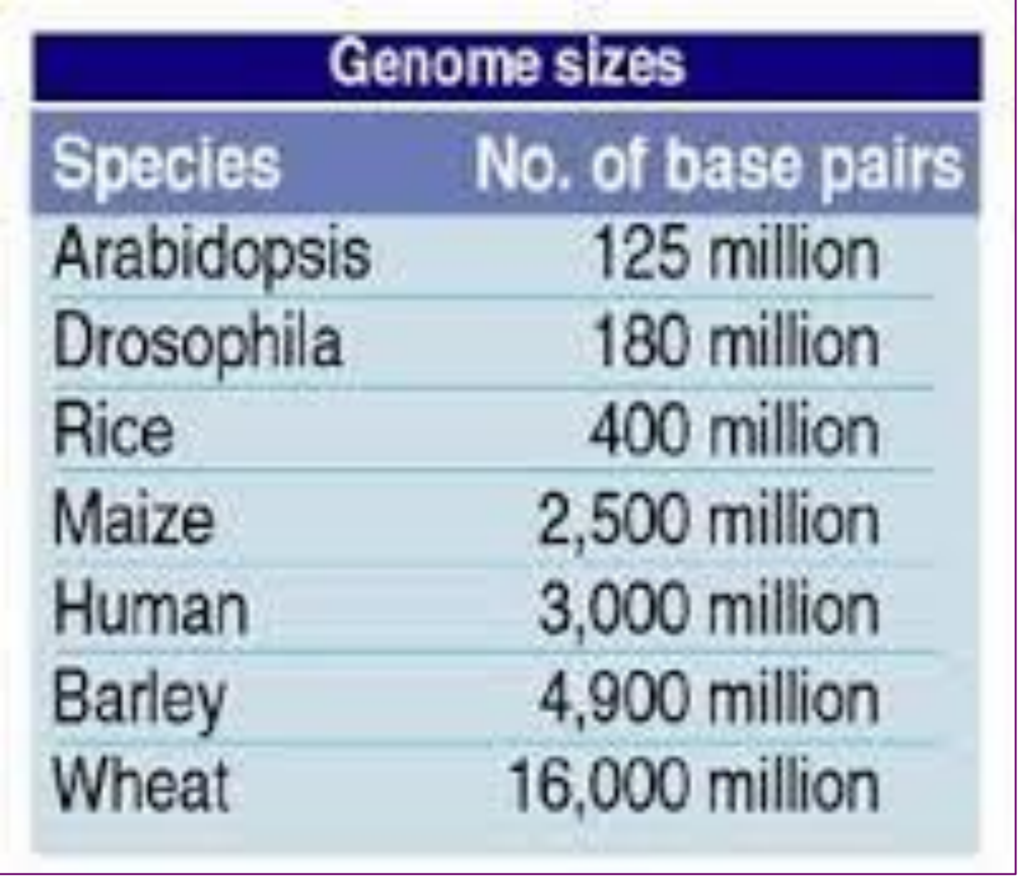
- Conservation planning requires data on diversity of species richness, endemism as well as genetic diversity
- Reconcile traditional approach based on morphology with high-resolution DNA sequence information
-

```
             1   2    3       4          5     6  7
Species A  ACCAGCCTGTGCATCGATGACGACTAAGTGATACCATAAAGACT
Species B  ACCAGCCTGTGCATCGATGACGACTAAGTGATACCATAAAGACT
Species C  ACGAGCATGTGCATCGATGCCGACTAAGTGATACCATAATGACT
Species D  ACGAGCATGTGCATCGATGCCGACTAAGTGATACCATAATGACT
Species E  ACCAGCATGTGTATCGATGCCGACTAAGTGATACCAAAATGACT
Species F  ACCAGCATGTGTATCGATGCCGACTAAGTGATACCAAAATGACT
Species G  ACCAGCATGTGTATCGATGCCGACTAAGTGCTACCATAATGACT
Species H  ACCAGCATGTGTATCGATGCCGACTAAGTGCTACCATAATGACT
```



Genome sizes (x10^9 DNA bp) *vs.* chromosome numbers (2n)





Genome sizes

Species	No. of base pairs
Arabidopsis	125 million
Drosophila	180 million
Rice	400 million
Maize	2,500 million
Human	3,000 million
Barley	4,900 million
Wheat	16,000 million

In a Conventional PCR

20-30 bases primers anneal over millions of bases -

Phylogenetic diversity: measuring genetic distance

- Units of diversity- “species”
- Classical approach: identification can be ambiguous when characters such as flower color, flower shape, and leaf morphology change over the life of an individual
- Definition of the “species”-

Taxonomic approaches

- Incorporate Chloroplast/mitochondrial and nuclear DNA sequences into taxonomic keys
- These DNA barcodes are available in GenBank database.
- DNA barcoding has greatly improved species delineation.
- Nearly 100 plant genomes available in public databases

DNA barcoding has greatly improved species delineation.

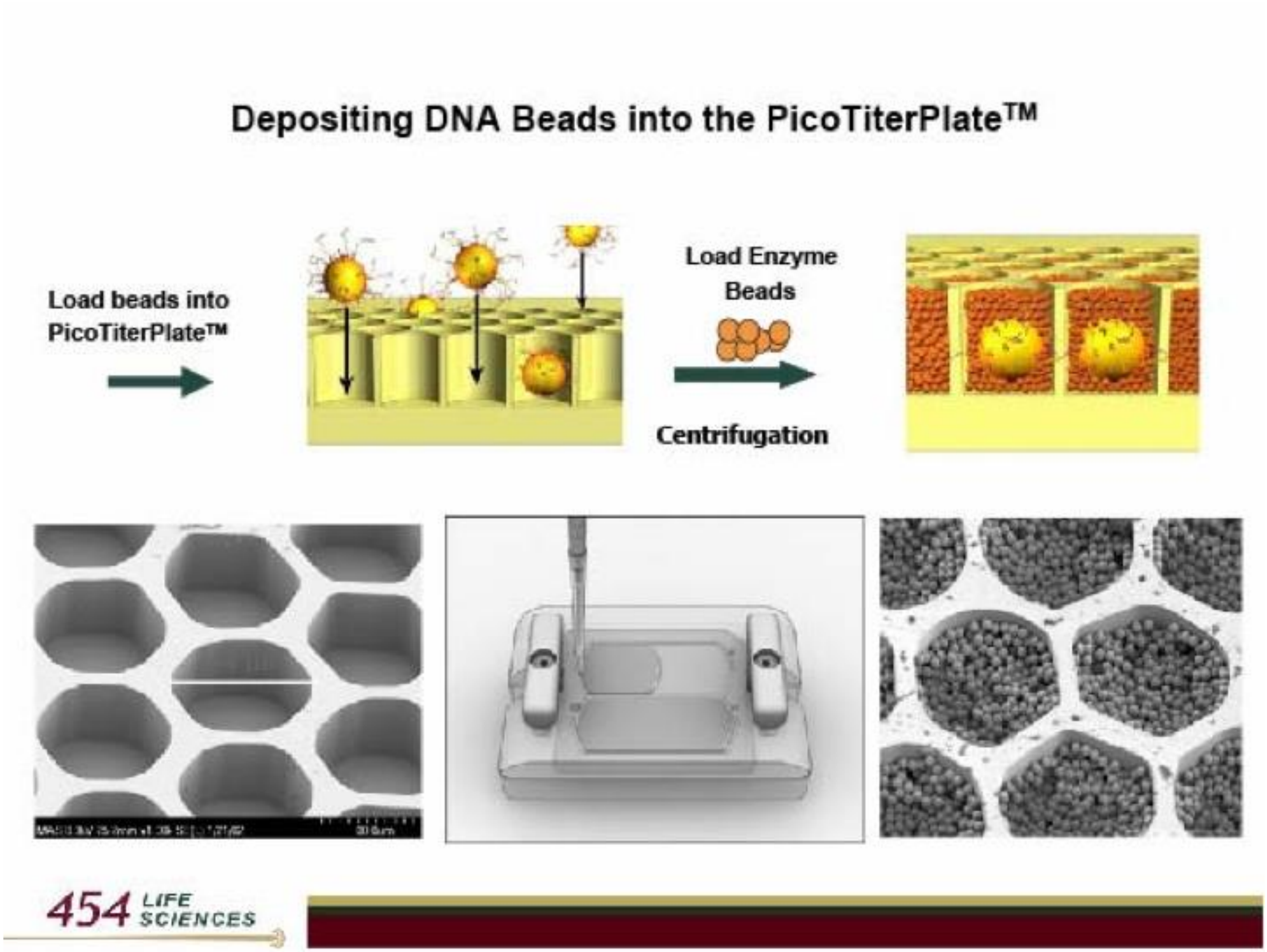
- (ITS + *trnH-psbA*, Kress et al., 2005;
- portion of *matK*, Lahaye et al., 2008),
- land plants (ITS + *rbcL*, Chase et al., 2005; *rpoC1* + *rpoB* + *matK* or *rpoC1* + *matK* + *psbA-trnH*, Chase et al., 2007; portion of *rbcL* + *trnH- psbA*, Kress and Erickson, 2007; *rbcL* + *matK*, CBOL, 2009),
- combinations of *rbcL, rpoC1, rpoB, matK, ycf5, trnL, psbA-trnH*, and ITS Gonzalez et al., 2009),
- and a 50-ha forest in Panama (*rbcL* + *matK* + *trnH-psbA*, Kress et al., 2009).

Whole genome sequencing

- Shotgun sequencing using next-generation sequencing (NGS) : provides very large volumes of high-quality sequence data in the form of short sequences.
- the majority of food and fiber crops have **at least one genome assembly** in the public domain

- **Next Generation Sequencing**
- 454 - 35 bases
- Illumina 250 bases
- 15 000 bases are available on the Pac Bio platform

Chemistry of sequencing reactions have improved to give high-throughput data



Depositing DNA Beads into the PicoTiterPlate™

Genomes of crop wild relatives completed

Arachis duranensis (A genome diploid wild peanut) and *Arachis ipaensis* (B genome diploid wild peanut) Wild ancestors of peanut, an oilseed and grain legume crop-

Solanum pimpinellifolium closest wild relative to tomato

Solanum commersoni - Wild potato relative

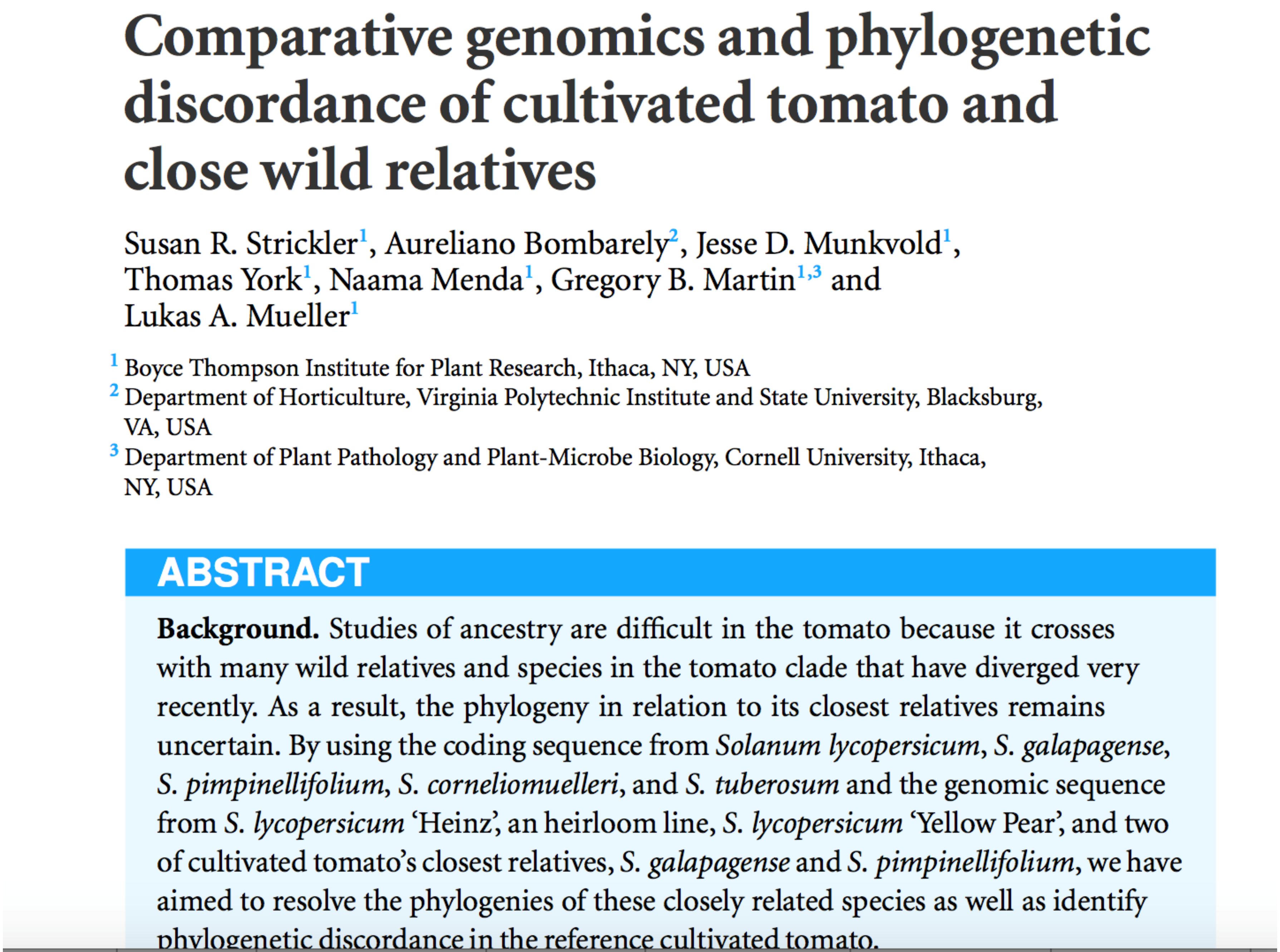
Oryza brachyantha (wild rice) -Disease resistant wild relative of rice

Oryza glaberrima (African rice) - West-African species of rice

Oryza rufipogon (red rice) - Ancestor to Oryza sativa

Molecular analysis has shown that there is frequent exchange of segments between *O. sativa* and *O. glaberrima*.

Wild relatives have poor phenotypic characteristics



Comparative genomics and phylogenetic discordance of cultivated tomato and close wild relatives

Susan R. Strickler[1], Aureliano Bombarely[2], Jesse D. Munkvold[1],
Thomas York[1], Naama Menda[1], Gregory B. Martin[1,3] and
Lukas A. Mueller[1]

[1] Boyce Thompson Institute for Plant Research, Ithaca, NY, USA
[2] Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
[3] Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY, USA

ABSTRACT

Background. Studies of ancestry are difficult in the tomato because it crosses with many wild relatives and species in the tomato clade that have diverged very recently. As a result, the phylogeny in relation to its closest relatives remains uncertain. By using the coding sequence from *Solanum lycopersicum*, *S. galapagense*, *S. pimpinellifolium*, *S. corneliomuelleri*, and *S. tuberosum* and the genomic sequence from *S. lycopersicum* 'Heinz', an heirloom line, *S. lycopersicum* 'Yellow Pear', and two of cultivated tomato's closest relatives, *S. galapagense* and *S. pimpinellifolium*, we have aimed to resolve the phylogenies of these closely related species as well as identify phylogenetic discordance in the reference cultivated tomato.

De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits

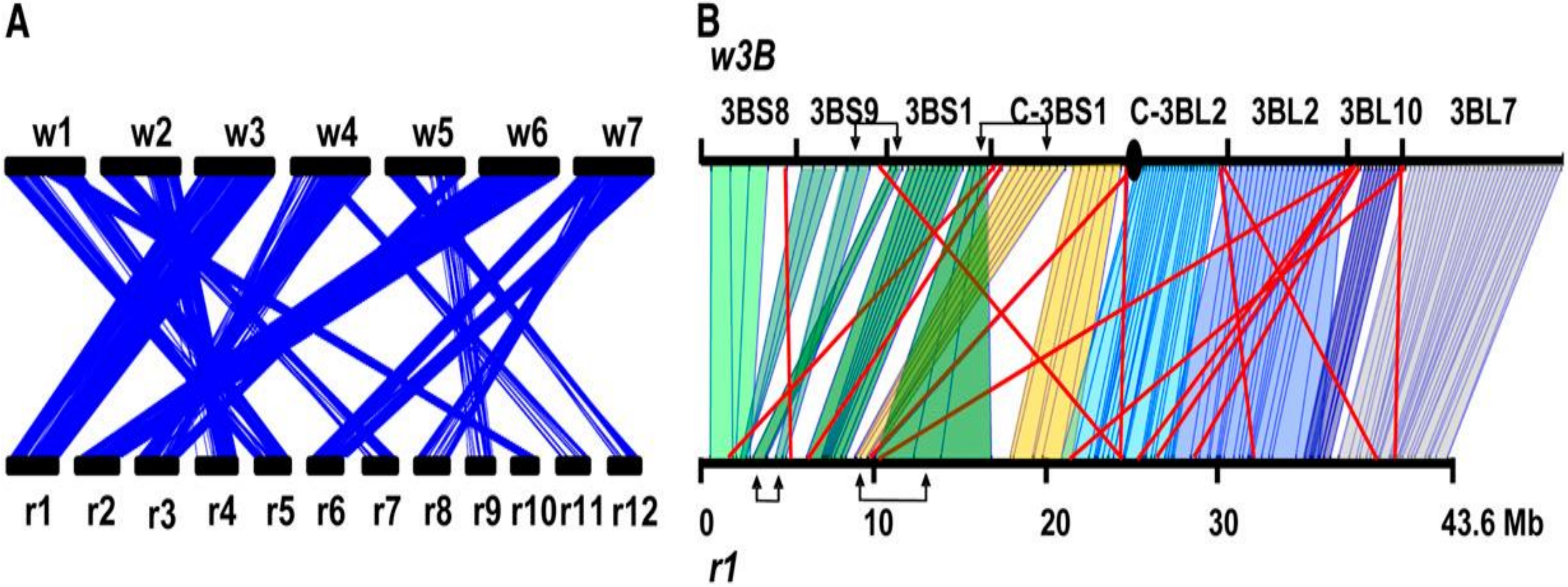
Ying-hui Li[1–3,11], Guangyu Zhou[4,10,11], Jianxin Ma[5,11], Wenkai Jiang[4,11], Long-guo Jin[1–3], Zhouhao Zhang[4], Yong Guo[1–3], Jinbo Zhang[4], Yi Sui[1–3], Liangtao Zheng[4], Shan-shan Zhang[1–3], Qiyang Zuo[4], Xue-hui Shi[1–3], Yan-fei Li[1–3], Wan-ke Zhang[6], Yiyao Hu[4], Guanyi Kong[4], Hui-long Hong[1–3], Bing Tan[1–3], Jian Song[1–3], Zhang-xiong Liu[1–3], Yaoshen Wang[4], Hang Ruan[4], Carol K L Yeung[4], Jian Liu[4], Hailong Wang[4], Li-juan Zhang[1–3], Rong-xia Guan[1–3], Ke-jing Wang[1–3], Wen-bin Li[7], Shou-yi Chen[6], Ru-zhen Chang[1–3], Zhi Jiang[4], Scott A Jackson[8], Ruiqiang Li[4,9] & Li-juan Qiu[1–3]

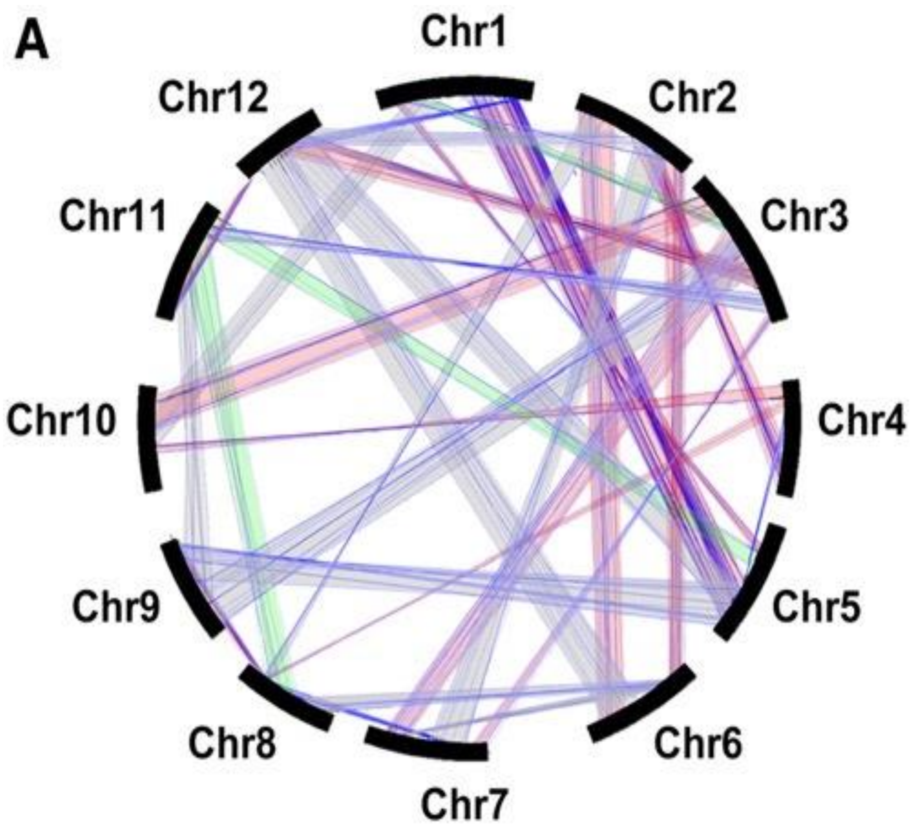
Wild relatives of crops are an important source of genetic diversity for agriculture, but their gene repertoire remains largely unexplored. We report the establishment and analysis of a pan-genome of *Glycine soja*, the wild relative of cultivated soybean *Glycine max*, by sequencing and *de novo* assembly of seven phylogenetically and geographically representative accessions. Intergenomic comparisons identified lineage-specific genes and genes with copy number variation or large-effect mutations, some of which show evidence of positive selection and may contribute to variation of agronomic traits such as biotic resistance, seed composition, flowering and maturity time, organ size and final biomass. Approximately 80% of the pan-genome was present in all seven accessions (core), whereas the rest was dispensable and exhibited greater variation than the core genome, perhaps reflecting a role in adaptation to diverse environments. This work will facilitate the harnessing of untapped genetic diversity from wild soybean for enhancement of elite cultivars.

Annual wild soybean (*Glycine soja* Sieb. & Zucc.) is the closest rela-

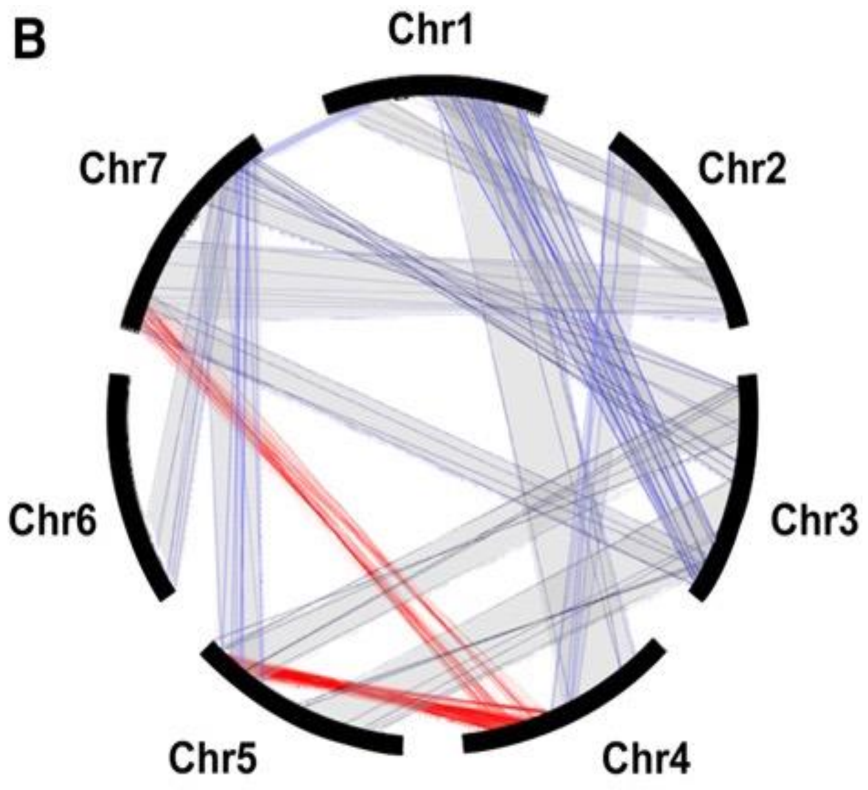
be missed[7]. Moreover, a single genome is insufficient to represent the

Orthologous Regions Between Wheat and Rice





29 duplication blocks on the 12 rice chromosomes.



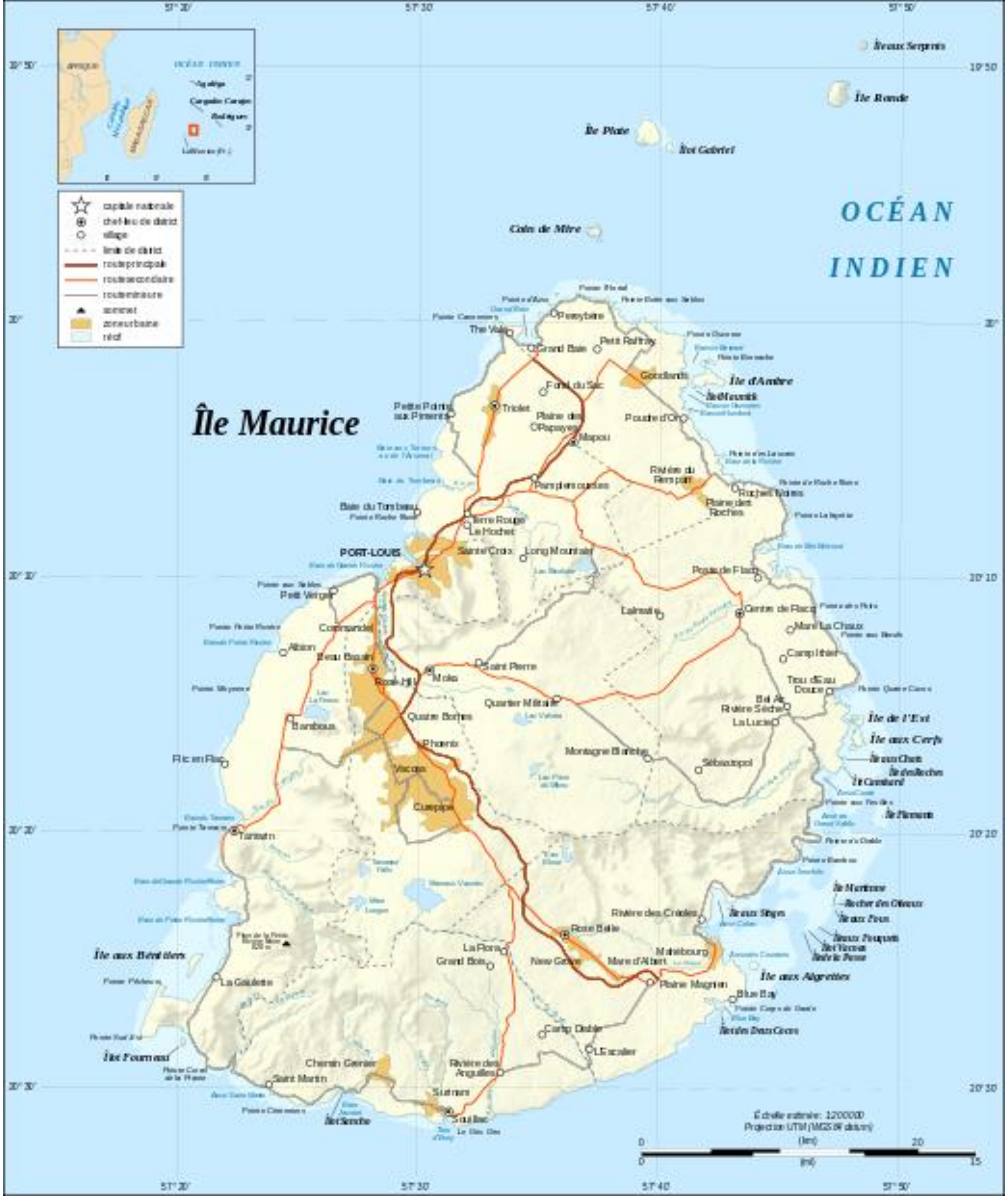
10 duplicated regions and 2 translocations identified on the seven wheat chromosome groups.

Simple Tandem Repeat Sequences

Class	Repeat size bp	Array size (no. of units)	Genomic distribution
Satellites	2-2,000	> 1,000	Centromeric, heterochromatic
Minisatellites	9-100	10-100	Sub-telomeric, dispersed
Short tandem repeats	3-5	10-100	Dispersed
Microsatellites	1-2	10-100	Dispersed

Dictyosperma in Mauritius

- Palm family (Arecaceae)
- 188 genera and approximately 2,585 species (Palmweb,2015)
- In the Mascarenes :11 species in five endemic genera (Moore & Gueho, 1984)
- Mauritius has highest palm diversity, with seven species and eight taxa – six endemic
- La Réunion has four species, two taxa endemic;
- Rodrigues has three species of which three taxa are endemic


Île Maurice
OCÉAN INDIEN
capitale nationale
chef-lieu de district
village
limite de district
route principale
route secondaire
route mineure
sommet
zone urbaine
récif
AFRIQUE
OCÉAN INDIEN
MADAGASCAR
Îlot aux Serpents
Île Ronde
Île Plate
Îlot Gabriel
Coin de Mire
Île d'Ambre
Île de l'Est
Île aux Cerfs
Île Flamants
Île aux Aigrettes
Île aux Bénitiers
Îlot Fourneau
Îlot Sancho
Pereybère
Grand Baie
Petit Raffray
Goodlands
Fond du Sac
Triolet
Plaine des Papayes
Poudre d'Or
Mapou
Pamplemousses
Rivière du Rempart
Roches Noires
Plaine des Roches
Baie du Tombeau
Terre Rouge
Le Hochet
PORT-LOUIS
Sainte Croix
Long Mountain
Poste de Flacq
Petit Verger
Lalmatie
Centre de Flacq
Mare La Chaux
Camp Ithier
Trou d'Eau Douce
Coromandel
Albion
Beau Bassin
Rose-Hill
Moka
Saint Pierre
Quartier Militaire
Bel Air Rivière Sèche
La Lucie
Quatre Bornes
Bambous
Phoenix
Montagne Blanche
Sébastopol
Flic en Flac
Vacoas
Curepipe
Tamarin
Rivière des Créoles
Rose Belle
La Flora
Grand Bois
New Grove
Mare d'Albert
Mahébourg
Plaine Magnien
Blue Bay
La Gaulette
Camp Diable
L'Escalier
Chemin Grenier
Rivière des Anguilles
Saint Martin
Surinam
Souillac
57° 20'
57° 30'
57° 40'
57° 50'
19° 50'
20°
20° 10'
20° 20'
20° 30'
Échelle estimée: 1:200000
Projection UTM (WGS 84 datum)
0
20
15
(km)
(mi)

Scientific name	Vernacular name	Endemic range	IUCN Red List category	Estimated wild population
Dictyosperma album (Bory) H. Wendl. & Drude ex Scheff. var. *album*	Palmiste blanc	Mauritius & La Réunion	CR	c. 30
Dictyosperma album var. *conjugatum* Moore & Guého	Palmiste de l’Île Ronde	Round Island	CR	1

Dictyosperma album var *album*

Dictyosperma album **var** ***conjugatum***

Ile Ronde
(Round Island)

Dictyosperma album var *conjugatum*

Ile aux Aigrettes

Ile aux Aigrettes

Re-introduction onto Round Island

Seedlings

Annals of Botany Page 1 of 7
doi:10.1093/aob/mcr158, available online at www.aob.oxfordjournals.org

ANNALS OF
BOTANY

PART OF A SPECIAL ISSUE ON PALM BIOLOGY

DNA barcoding: a new tool for palm taxonomists?

Marc L. Jeanson[1,2,*], Jean-Noël Labat[2] and Damon P. Little[1]

[1]*Lewis B. and Dorothy Cullman Program for Molecular Systematics, The New York Botanical Garden, Bronx, NY 10458-5126*
USA and [2]Département systématique et évolution, UMR 7205, Muséum National d'Histoire Naturelle, Paris 75005, France
** For correspondence. E-mail mjeanson@nybg.org*

Received: 31 October 2010 Returned for revision: 17 December 2010 Accepted: 3 May 2011

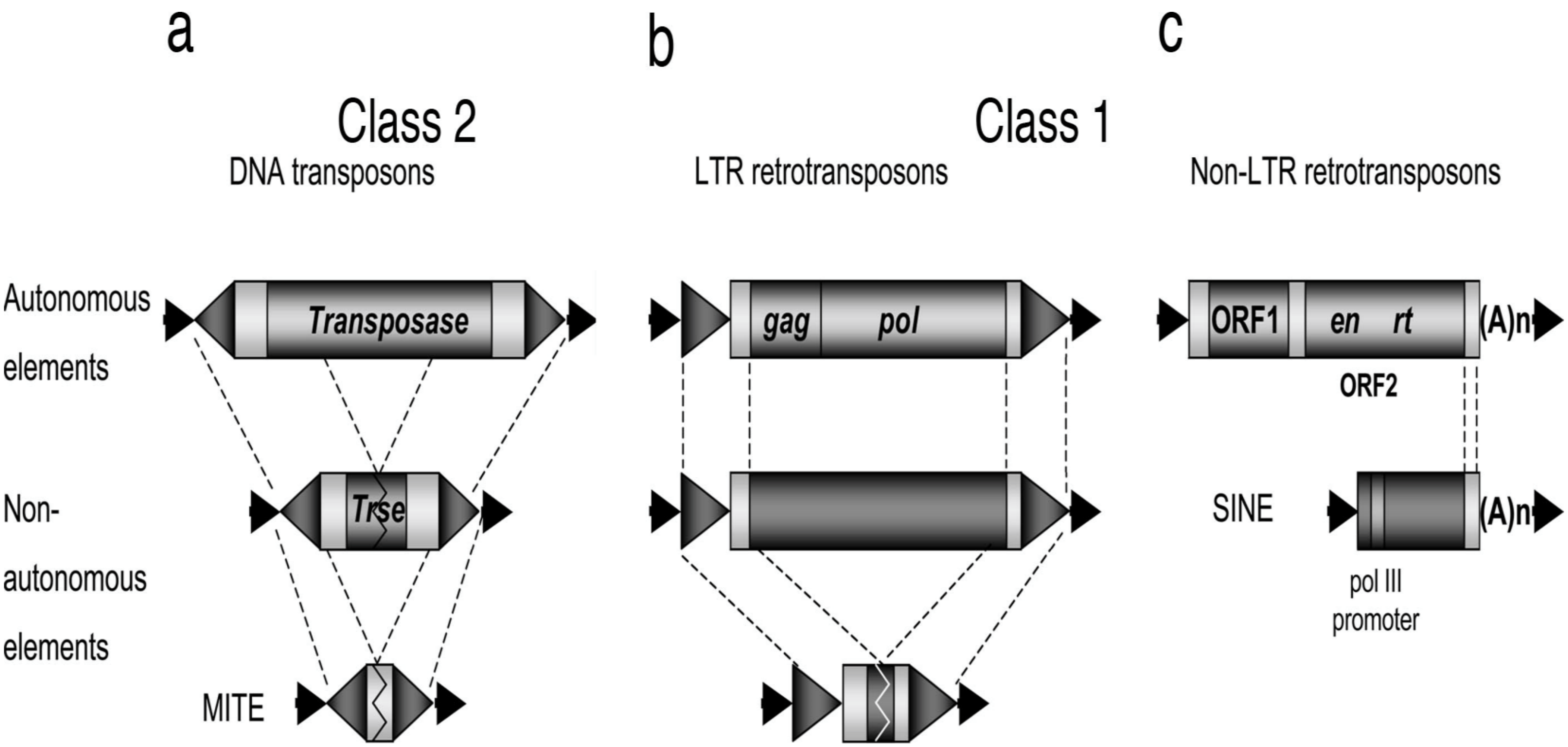
- *Background and Aims* In the last decade, a new tool – DNA barcoding – was proposed to identify species. The technique of DNA barcoding is still being developed. The Consortium for the Barcode of Life's Plant Working Group (CBOL-PWG) selected two core markers (*matK* and *rbcL*) that now must be tested in as many taxa as possible. Although the taxonomy of palms (Arecaceae/Palmae) has been greatly improved in the past decades, taxonomic problems remain. Species complexes, for example, could significantly benefit from DNA barcoding. Palms have never before been subjected to a DNA barcoding test.
- *Methods* For this study, 40 out of the 48 species of the southeast Asian tribe Caryoteae (subfamily Coryphoideae) were included. In total, four DNA markers – three plastid encoded (*matK*, *rbcL* and *psbA-trnH*) and one nuclear encoded (nrITS2) – were analysed to determine if adequate variation exists to discriminate among species.
- *Key Results* The combination of three markers – *matK*, *rbcL* and nrITS2 – results in 92 % species discrimination. This rate is high for a barcoding experiment. The two core markers suggested by the CBOL-PWG, *rbcL* and *matK*, have a low species discrimination rate and need to be supplemented by another marker. In Caryoteae, nrITS2 should be chosen over *psbA-trnH* to supplement the two 'core' markers.
- *Conclusions* For the first time a test of DNA barcoding was conducted in Arecaceae. Considering that palms have highly variable mutation rates compared with other angiosperms, the results presented here are encouraging for developing DNA barcoding as a useful tool to identify species within this ecologically important tropical plant family.

Transposable Elements (TE)

- They account for almost 50% of the human genome and 70% of the genomes of some grass species, including maize.
- Class 1 LTR-retrotransposons, such as the *Gypsy* and *Copia* superfamilies
- Class 2 MITEs (Miniature Inverted Repeat Elements) are also abundant in maize

IRAP

- interretrotransposon amplied polymorphism (IRAP) protocol (Kalendar and Schulman 2006)
- segments between two nearby retrotransposons or LTRs(Long Terminal Repeats) are amplifed using outward-facing primers, usedto determine the diversity of TE elements.

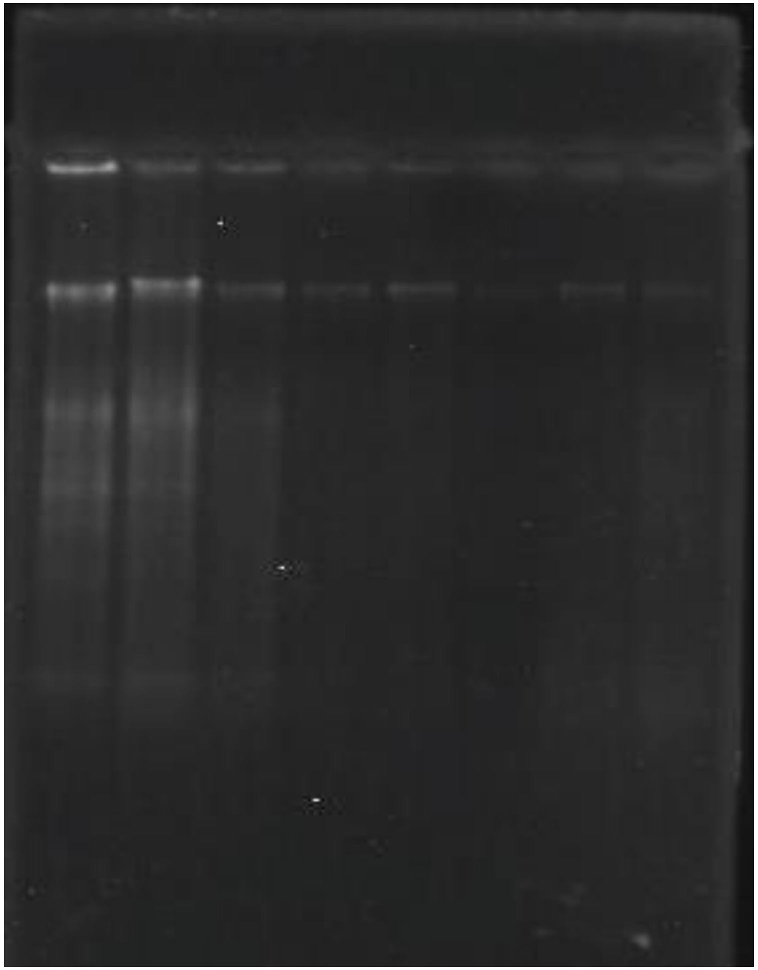

a
b
c
Class 2
Class 1
DNA transposons
LTR retrotransposons
Non-LTR retrotransposons
Autonomous elements
Transposase
gag
pol
ORF1
en
rt
(A)n
ORF2
Non-autonomous elements
Trse
SINE
(A)n
pol III promoter
MITE

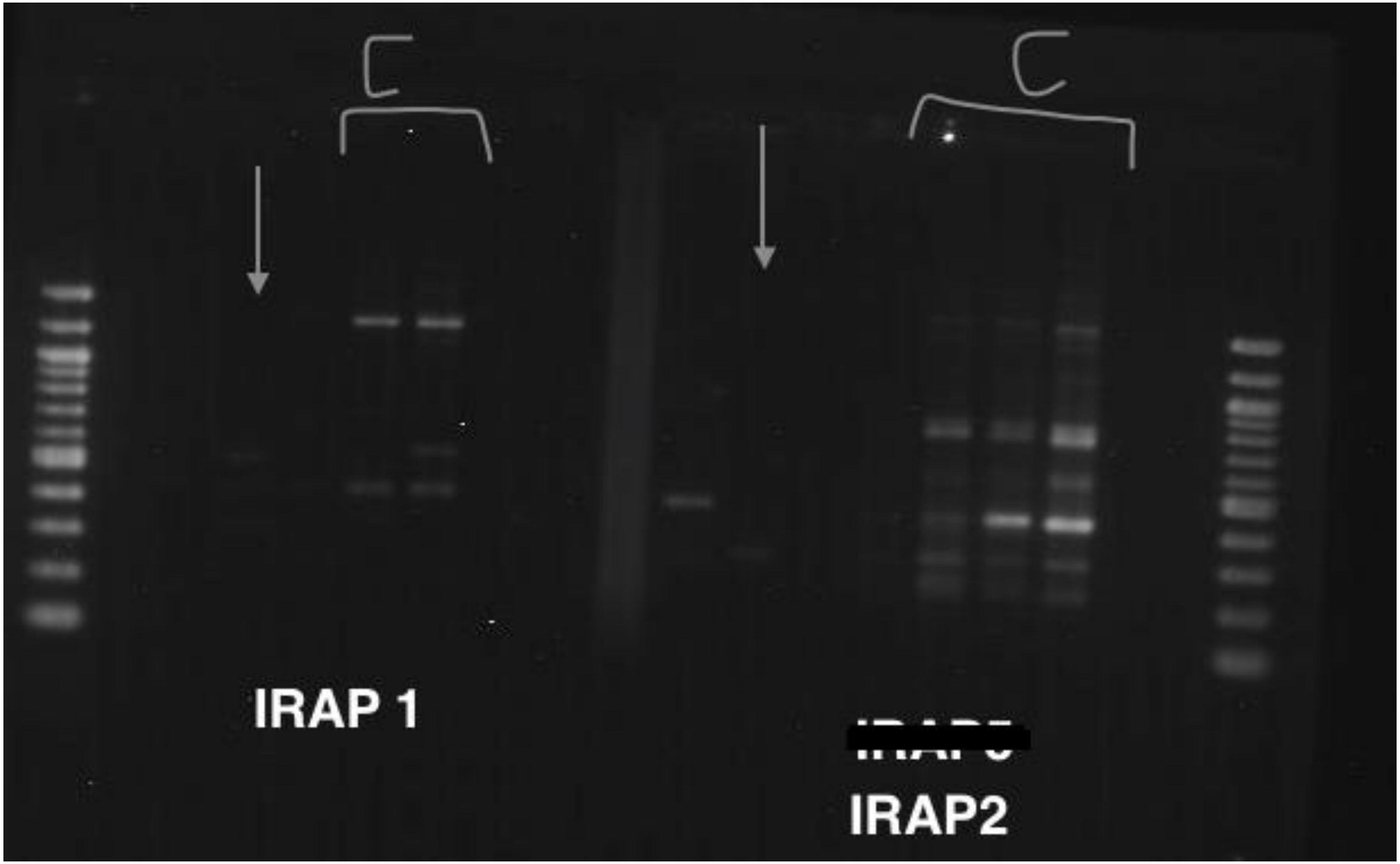
Molecular characterisation on *Dictyosperma album* var *album* and *D a* var conjugatum

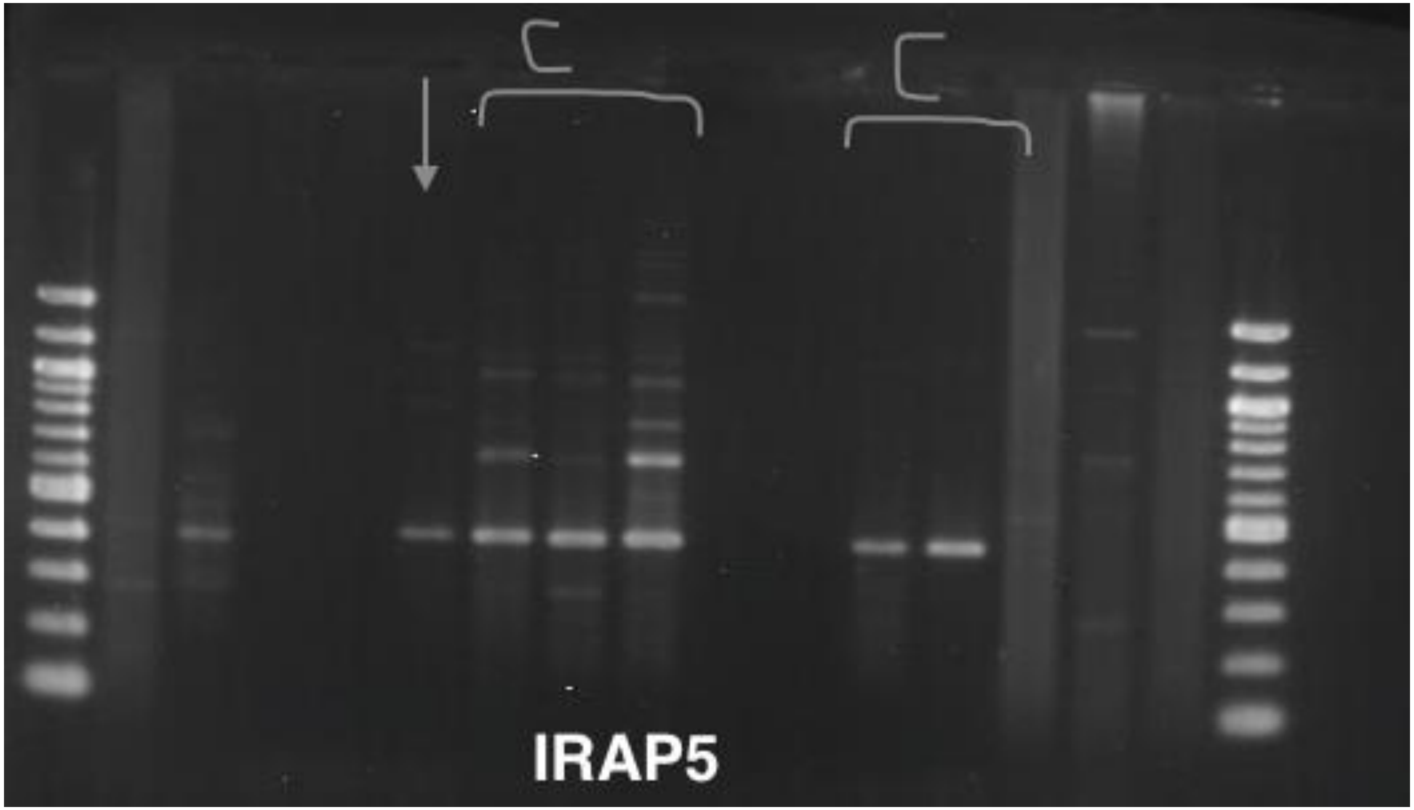
- Samples :
  - National Parks Black River (conjugatum)
  - Ile aux Aigrettes (3 conjugatum)
  - Curepipe (album)
  - Plaine Sophie (album)
  - Botanical Garden- Pamplemousse (conjugatum)

DNA extraction

- 1st attempt- Kit – not successful
- CTAB method- isopropanol




C
C
IRAP 1
[redacted]
IRAP2


IRAP5

On going work

- More samples to be tested
- Reproducibility of PCR
- DNA barcoding- amplicons will be sequenced-
- Outcome- relatedness of the two varieties that can assist in conservation management

Acknowledgements

- Bioversity International
- University of Birmingham
- University of Mauritius
- Ministry Agro Industry and Food Security
- Mauritius Herbarium
- Mauritius Wildlife Foundation
- Technician: Sunjaya Leela
- Research assistant: Lakshmi Sujeen