Regional Seminar

"Advanced research on promissory edible plants in Latin America: tools to improve Food Security in the region"

Multi-omics approach in plant characterization, domestication and breeding in a food safety context.

Dr. Gustavo H.F. Klabunde

klabunde.gustavo@gmail.com

Agronomist - Post-Doc Research Fellow – UFSC, Brasil

Dec 4th, 2017

Multi-omics approach in plant characterization, domestication and breeding in a food safety context.

Outline

Omics definition – Brief history and concept

Omics tools used by our research group on edible plants.

Genome characterization – Marker development and use
Phylogeography – *Araucaria angustifolia*
Chloroplast genomics – Native Myrtaceae species
Transcriptomics – *Acauraria angustifolia*
Phenomics – *Acca sellowiana*
Metagenomics – Bambusoideae species

Genome (Winkler 1920)

Hybrid term - Fusion of affixes

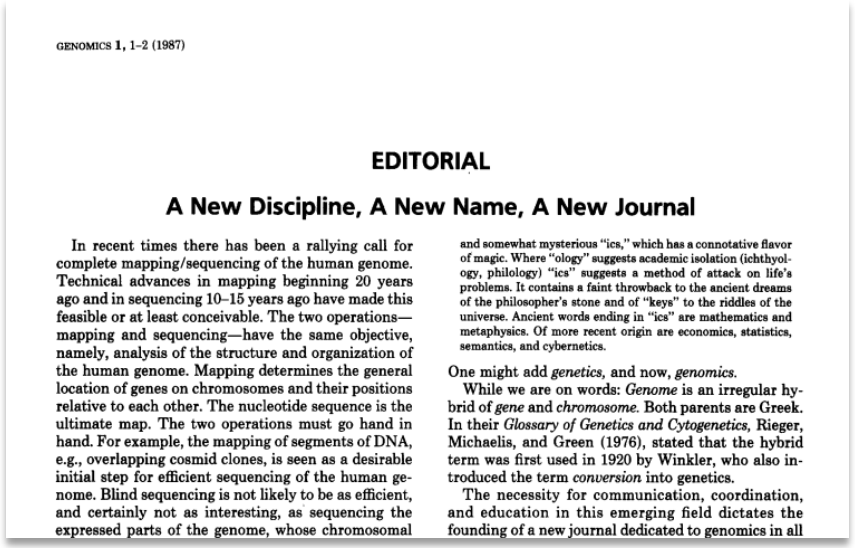
Gene (Johannsen 1909) – Evolving concept + *Chromosome* (Waldeyer-Hartz 1888)

"I propose the Genome expression for the set of haploid chromosomes, which together with their respective protoplast specifies the founding material of the species"

Winkler, H. 1920. Verbreitung und Ursache der Parthenogenesis im Pflanzen-und Tierreiche. Verlag Fischer, Jena.

Genomics

Term created in 1987
First "Omics"
New discipline
New peer-reviewed journal


GENOMICS 1, 1–2 (1987)

EDITORIAL

A New Discipline, A New Name, A New Journal

In recent times there has been a rallying call for complete mapping/sequencing of the human genome. Technical advances in mapping beginning 20 years ago and in sequencing 10–15 years ago have made this feasible or at least conceivable. The two operations—mapping and sequencing—have the same objective, namely, analysis of the structure and organization of the human genome. Mapping determines the general location of genes on chromosomes and their positions relative to each other. The nucleotide sequence is the ultimate map. The two operations must go hand in hand. For example, the mapping of segments of DNA, e.g., overlapping cosmid clones, is seen as a desirable initial step for efficient sequencing of the human genome. Blind sequencing is not likely to be as efficient, and certainly not as interesting, as sequencing the expressed parts of the genome, whose chromosomal and somewhat mysterious "ics," which has a connotative flavor of magic. Where "ology" suggests academic isolation (ichthyology, philology) "ics" suggests a method of attack on life's problems. It contains a faint throwback to the ancient dreams of the philosopher's stone and of "keys" to the riddles of the universe. Ancient words ending in "ics" are mathematics and metaphysics. Of more recent origin are economics, statistics, semantics, and cybernetics.

One might add *genetics*, and now, *genomics*.

While we are on words: *Genome* is an irregular hybrid of *gene* and *chromosome*. Both parents are Greek. In their *Glossary of Genetics and Cytogenetics*, Rieger, Michaelis, and Green (1976), stated that the hybrid term was first used in 1920 by Winkler, who also introduced the term *conversion* into genetics.

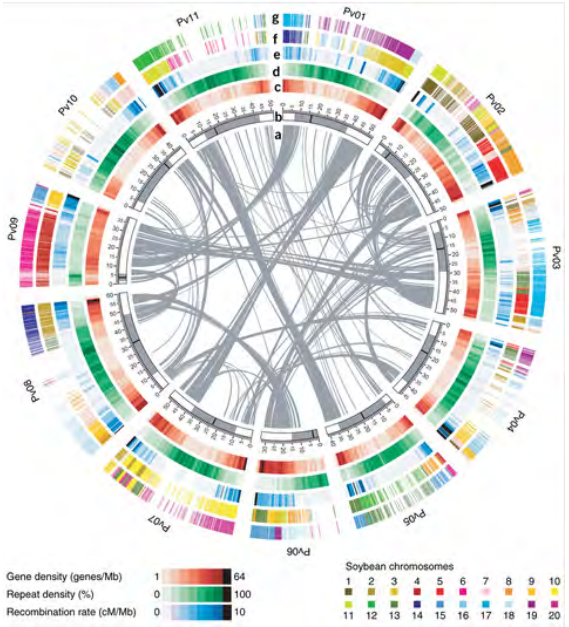
The necessity for communication, coordination, and education in this emerging field dictates the founding of a new journal dedicated to genomics in all

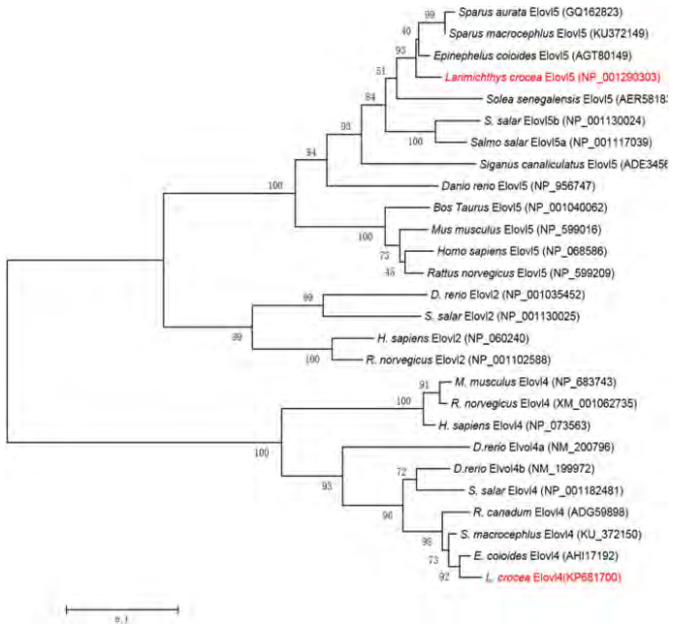
Genomics 1:1-2 (1987)

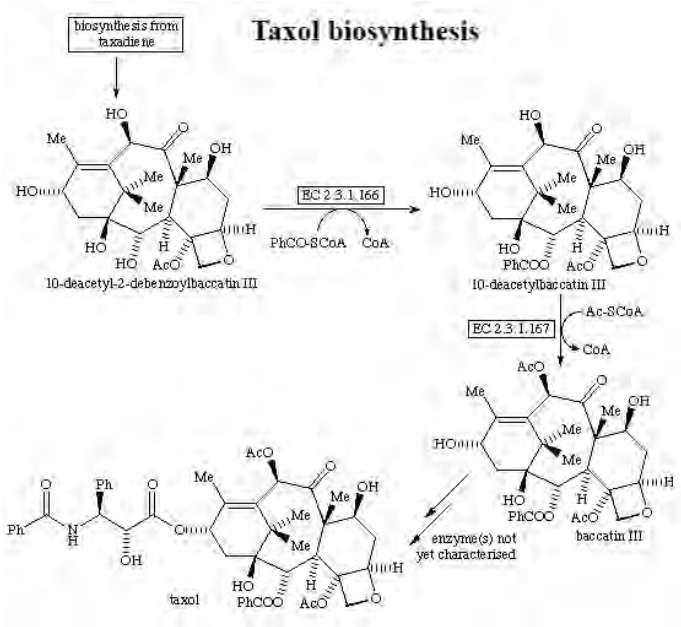
ISSN:0888-7543

Genomics

Structure **Evolution** **Function / Variation**

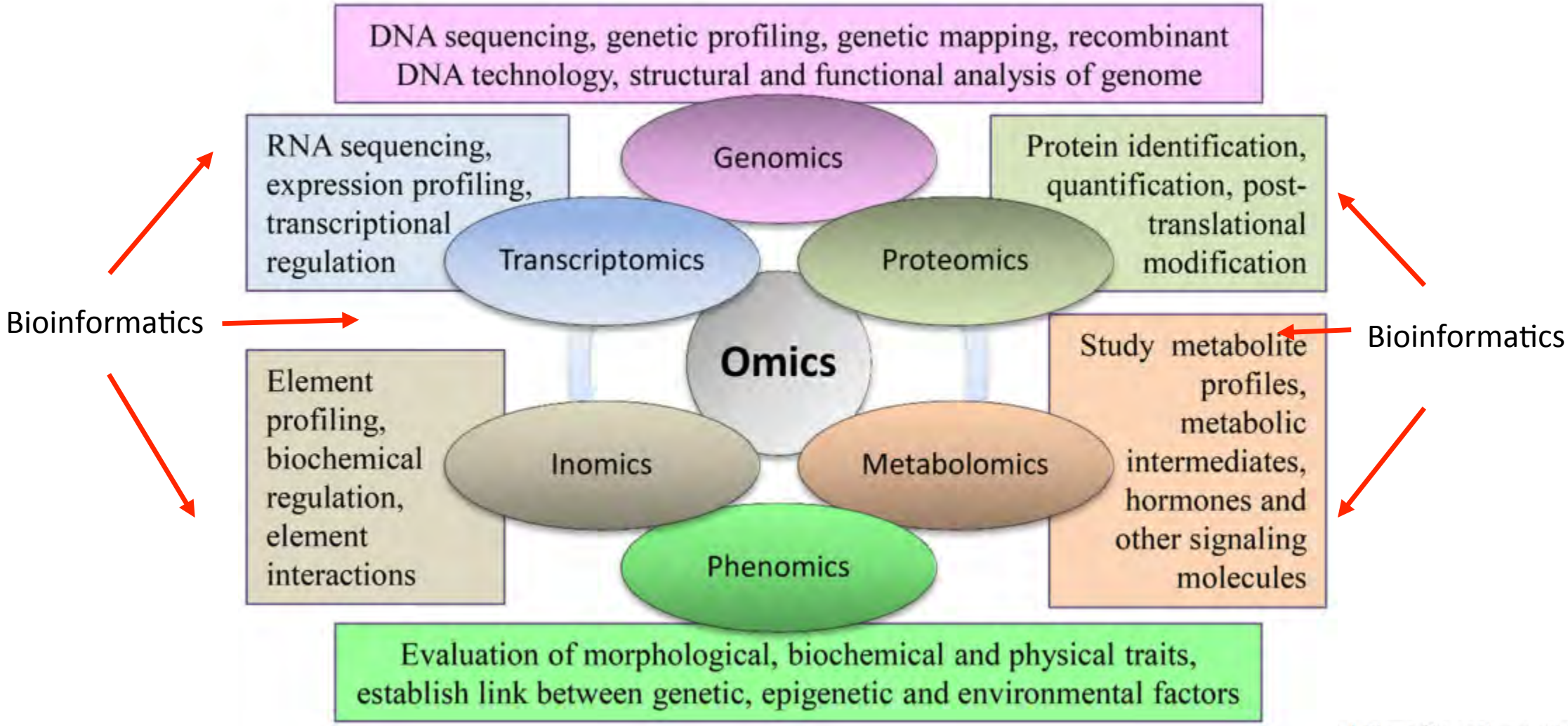






Training – “*capacity building*”

Omics



Frontiers in Plant Science | Plant Genetics and Genomics
Deshmukh et al. June 2014 | Volume 5 | Article 244

Short Tandem Repeats – STR

or Microsatellite - SSR

Molecular Marker Development

Acca sellowiana and *Campomanesia xanthocarpa* (Native Myrtaceae)

Short Tandem Repeats Development STR
or SSR – Microsatellite

- Development of specific nuclear and cpDNA molecular markers for poorly characterized tree species
- Requires classical Library construction (bacterial transformation) OR a single low throughput NGS run (only to identify repetitive mofits)
- Requires samples from at least 4 distant populations to maximize the probability of variant (alleles) finding
- Requires access to DNA genotyping platforms (C.E. or P.A.G.E.)
- Time to develop and validate markers: from 6 to 12 months
- When ready, STR markers are able to access genetic diversity from undercharacterized tree species

Applications in Plant Sciences 2014 2(6): 1400020

PRIMER NOTE

CHARACTERIZATION OF 10 NEW NUCLEAR MICROSATELLITE MARKERS IN *ACCA SELLOWIANA* (MYRTACEAE)[1]

GUSTAVO H. F. KLABUNDE[2,4], DENISE OLKOSKI[2,4], VINICIUS VILPERTE[2], MARIA I. ZUCCHI[3], AND RUBENS O. NODARI[2,5]

[2]Post Graduation Program in Plant Genetic Resources, Universidade Federal de Santa Catarina, CP 476 88034-000 Florianópolis, Brazil; and [3]Polo Centro Sul, Agência Paulista de Tecnologia dos Agronegócios, CP 13400 970 Piracicaba, Brazil

Acca sellowiana – Feijoa - Native Myrtaceae
Fruits with unique flavour and texture – High Market acceptance
In domestication and breeding process
Genetic diversity – Unknown for most populations
(Lack of molecular tools)

Samples collected for marker valitation

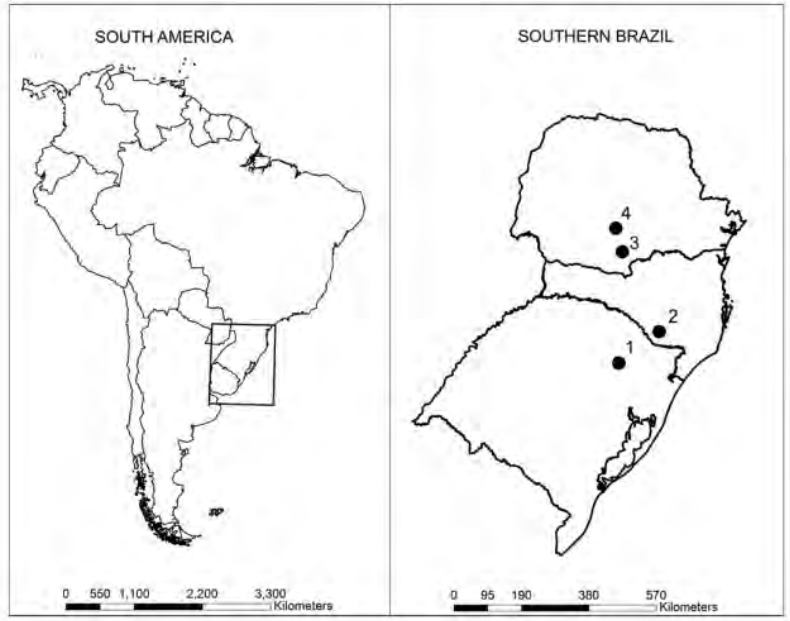


Fig. 1. Geographic localities of the populations sampled in this paper: 1 = Ipê, Rio Grande do Sul (28°75′14″S, 51°35′39″W); 2 = Urubici, Santa Catarina (28°04′18″S, 50°29′70″W); 3 = Bituruna, Paraná (26°02′52″S, 51°70′20″W); and 4 = Guarapuava, Paraná (25°26′29″S, 51°39′17″W).

Short Tandem Repeats Development STR

or SSR – Microsatellite

TABLE 2. Results of initial primer screening in four populations of *Acca sellowiana*.

	Ipê, Rio Grande do Sul (n = 15)				Urubici, Santa Catarina (n = 15)				Bituruna, Paraná (n = 15)				Guarapuava, Paraná (n = 15)			
Locus	A	H_e	H_o	PIC	A	H_e	H_o	PIC	A	H_e	H_o	PIC	A	H_e	H_o	PIC
Fse04^{ns}	6	0.671	0.667	0.596	9	0.809**	0.733	0.761	7	0.744**	0.813	0.676	9	0.874**	0.667	0.826
Fse06*	10	0.877	0.750	0.822	11	0.932	0.800	0.874	5	0.788	0.667	0.680	8	0.894**	0.714	0.846
Fse08*	5	0.458	0.400	0.421	5	0.743	0.733	0.672	4	0.599	0.500	0.493	5	0.637**	0.400	0.541
Fse09^{ns}	4	0.690**	0.667	0.610	5	0.533	0.467	0.479	4	0.651	0.688	0.578	5	0.729**	0.467	0.650
Fse10*	3	0.530	0.333	0.424	2	0.467**	0.200	0.332	3	0.654	0.273	0.553	2	0.536	0.250	0.359
Fse11^{ns}	7	0.724	0.800	0.654	8	0.832**	0.867	0.781	7	0.778	0.813	0.720	6	0.699	0.933	0.635
Fse12*	14	0.940	1.000	0.902	13	0.906	0.800	0.866	7	0.740	0.563	0.686	12	0.887**	0.867	0.844
Fse16^{ns}	12	0.913	1.000	0.872	15	0.949	1.000	0.912	12	0.911	1.000	0.871	8	0.855	1.000	0.806
Fse17^{ns}	12	0.931	0.933	0.891	15	0.915	0.800	0.875	14	0.883	0.875	0.845	11	0.864**	0.733	0.821
Fse21^{ns}	7	0.825**	0.867	0.769	6	0.848**	0.867	0.794	7	0.851**	0.875	0.800	5	0.789**	0.533	0.724

Note: A = number of alleles sampled; H_e = expected heterozygosity; H_o = observed heterozygosity; n = number of individuals sampled; ns = nonsignificant values for null alleles analysis ($P > 0.05$); PIC = polymorphism information content.

*Significant values for null alleles analysis ($P < 0.05$).

**Indicates deviation from Hardy–Weinberg equilibrium ($P < 0.05$).

New markers

Alleles revealed

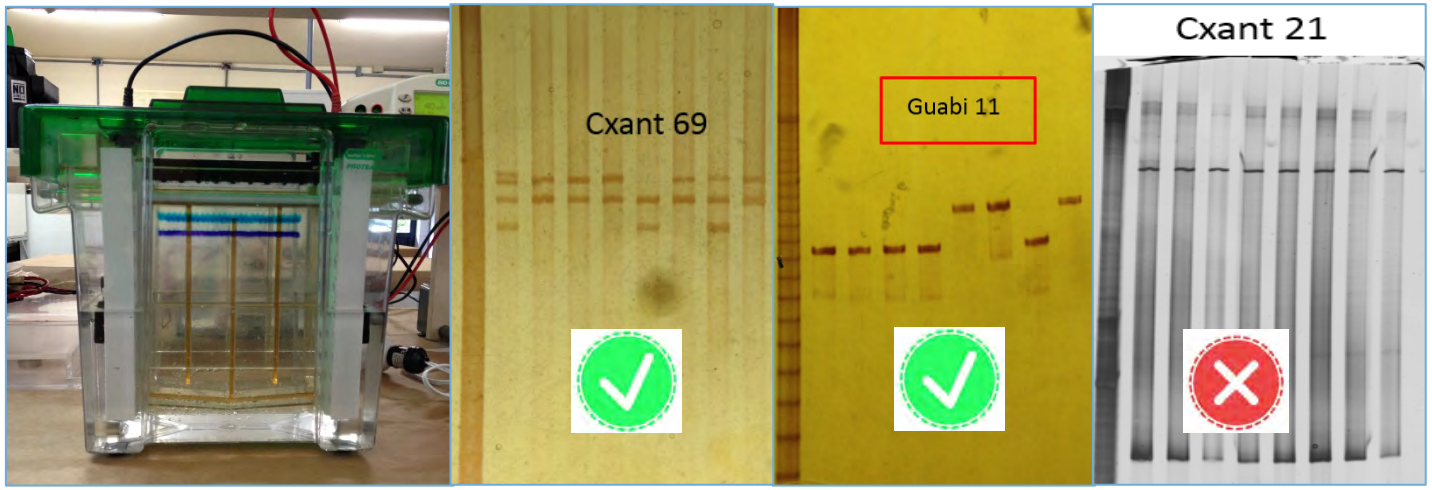
High Polimorphysm Information Content (PIC)

Short Tandem Repeats Development STR
or SSR – Microsatellite

Laboratory phase 1

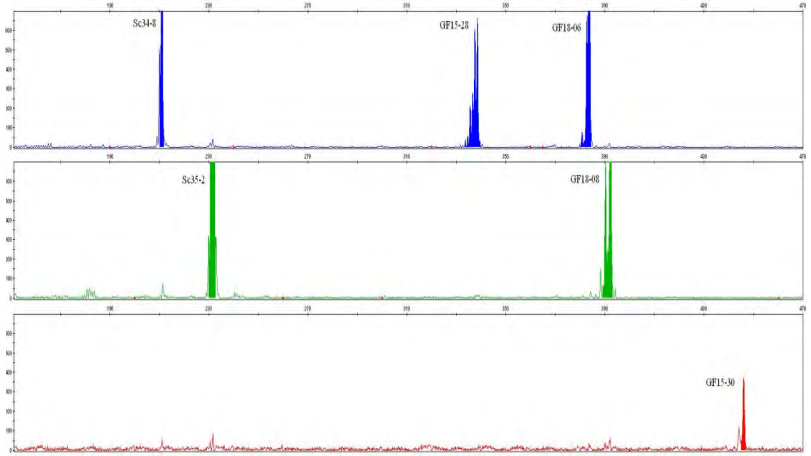
Polyacrylamide Gel Electrophoresis P.A.G.E.
Cost effective

Polymorphism screening



Laboratory phase 2
Automated genotyping
Cappilary Electrophoresis E.C.

ABI 3500 XL DNA Sequencer



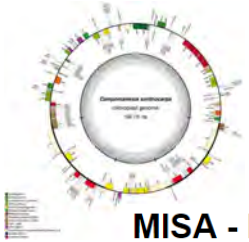
*Multiplex genotyping Panels (until 24 loci per run)
*96 plants genotyped in less than 120 minutes

Short Tandem Repeats Development STR
or SSR – Microsatellite

***Campomanesia xanthocarpa* (Myrtaceae) "guabiroba"**

Genetic diversity – Unknown (Lack of molecular tools)
Medicinal properties – Fruits
Cultivation option

Chloroplast Genome



MISA - MIcroSAtellite identification tool Primer3

12 Putative motifs detected

2 Polymorphic CpSSR

Polymorphism screening 4 native pops.

Nuclear genome Illumina run

77 Putative motifs detected

8 Polymorphic CpSSR

Locus	Primer sequences (5'-3')	Repeat motif	Allele size (bp)	T_a (°C)	5´ Dye	GenBank accession no.
Cxant22	F: GCTTGGTGGTGCCTCTCTC	$(TCCA)_3$	209	55	VIC	MG557627
	R: GCTCTTCCCTTTGCCTCTCT					
Cxant26	F: ATGCAAAATCCCTACGTGCT	$(ATCG)_3$	162	57	NED	MG557629
	R: ATGACACATTTCGGCTGTGA					
Cxant36	F: TTCCCGCCTAACCCTAATG	$(AAG)_4$	323	57	PET	MG557631
	R: GTCAAATCTCGCTCCTCCAA					
Cxant50	F: CGCACAACCAGCACAAAAC	$(CTTT)_3$	477	66	6FAM	MG557634
	R: CTATCACCGAGGGAGGCAAG					
Cxant59	F: GAGGGACTTTCAGTTTGTGTGTC	$(GA)_{10}$	230	55	NED	MG557635
	R: GACCGTTTCCAACATTTCCA					
Cxant66	F: GCGAGACCATAAGCCACTAC	$(AGA)_4$	211	57	NED	MG557636
	R: TGAGAAGGAGACACACACAAAT					
Cxant69	F: CCCAACACTCTCCACAATCC	$(GA)_7$	294	55	6FAM	MG557637
	R: TCCTTCCCTCTTCTCTCCATC					
Cxant76	F: ATGTTTTTGTGCGTTCTGG	$(AAG)_6$	336	57	VIC	MG557638
	R: TTGACCTTTGTTCCTCTTCCT					
Guabi05	F: TTCTCGTGATTTGTATCCAAGG	$(T)_{10}$	201	66	PET	MG557633
	R: TGCTTCAATCTTTCCTATCGAA					
Guabi11	F: TTGATTCAGGGAACAAATTCAA	$(ATTA)_4$	225	55	6FAM	MG557632
	R: TGGCTAGTGTGGTTCATTCAG					

Short Tandem Repeats – STR

Molecular Marker Characterization

Genetic characterization of wild populations of *Araucaria angustifolia* (Native conifer






Short Tandem Repeats Characterization

- Requires previous marker Development and Validation
- STRs markers reveal present genetic diversity and gene flow indexes
- In some cases, STR data may be incorpored to Phylogeographyc analysis
- STR Analysis reveals several Population Genetics questions, such as derived from:
  - Inbreeding / Mating systems
  - Migration / Habitat Fragmentation Effects
  - Random Genetic Drift
  - Natural or Artificial Selection

Conserv Genet (2013) 14:601–613
DOI 10.1007/s10592-013-0451-5

RESEARCH ARTICLE

Realized pollen and seed dispersal within a continuous population of the dioecious coniferous Brazilian pine [*Araucaria angustifolia* (Bertol.) Kuntze]

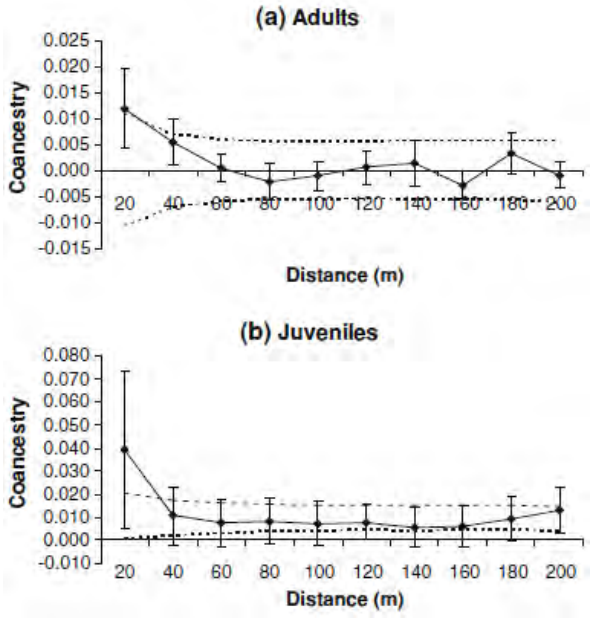
Cristina S. Sant'Anna · Alexandre M. Sebbenn · Gustavo H. F. Klabunde · Ricardo Bittencourt · Rubens O. Nodari · Adelar Mantovani · Maurício S. dos Reis

Multiple purpose study:

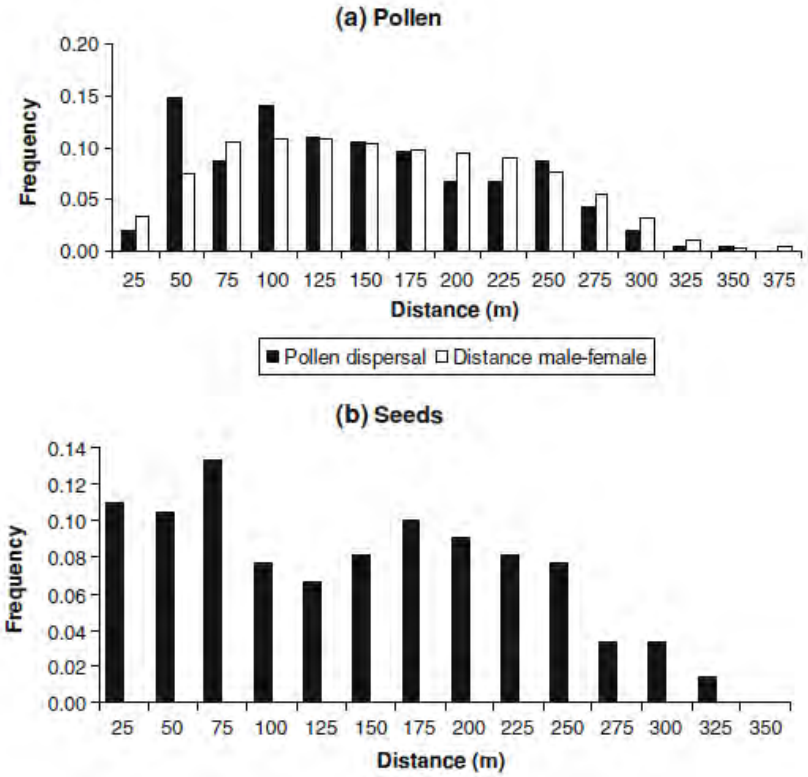
To access genetic diversity of Federal Conservation Areas and evaluate its "efficiency in conservate genetic resources".

To determine seed and pollen dispersal ranges. Increasing concern about the high level of forest fragmentation and associated loss of diversity.

Spatial genetic Structure



Frequence distribution of observed distance dispersal events of (a) pollen and (b) seed dispersal, **determined by STR parentage analysis**



Chloroplast Genomics Plastome

Whole plastome sequence provides:

Data for evolutionary studies

Filogeny

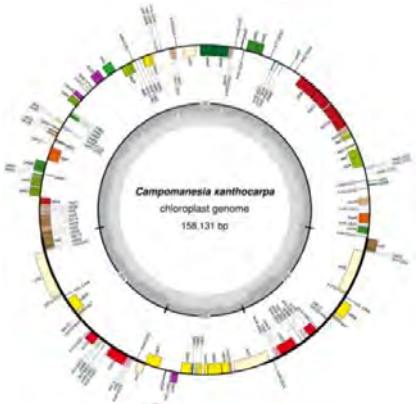
Ecology

Source of new molecular markers (SSR, SNP and intergenic spacers)

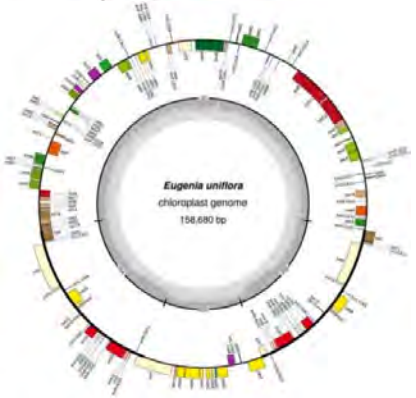
Genetic diversity

Phylogeography

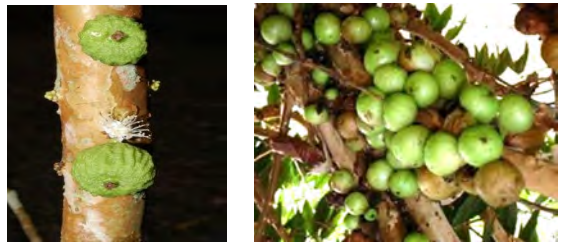
Campomanesia xanthocarpa

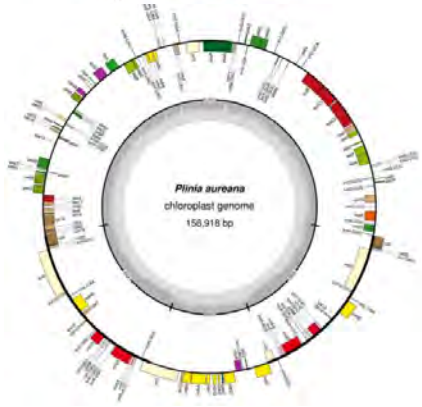


Eugenia uniflora



Plinia aureana

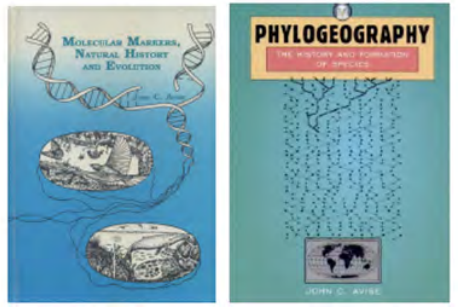




Tree Phylogeography *Araucaria angustifolia*

- Study of principals and processes governing the geographical distribution of individuals
- Use of Chloroplast DNA (cpDNA) variation to reconstruct historical events due to its low mutation rates
- Extensive field sampling work (whole distribution range) and DNA sequencing (Sanger Sequencing 1st gen.)
- Search for barriers, historical gene flow tendencies, glacial refugia and actual gene diversity status for conservation purposes

John C. Avise



Sampling area x Range distribution
40 natural populations (n=594 trees)

Annual Reviews
www.annualreviews.org/aronline

Ann. Rev. Ecol. Syst. 1987. 18:489–522

INTRASPECIFIC PHYLOGEOGRAPHY: The Mitochondrial DNA Bridge Between Population Genetics and Systematics

John C. Avise[1], Jonathan Arnold[1], R. Martin Ball[1], Eldredge Bermingham[1,2], Trip Lamb[1,3], Joseph E. Neigel[1,4], Carol A. Reeb[1], and Nancy C. Saunders[1,5]

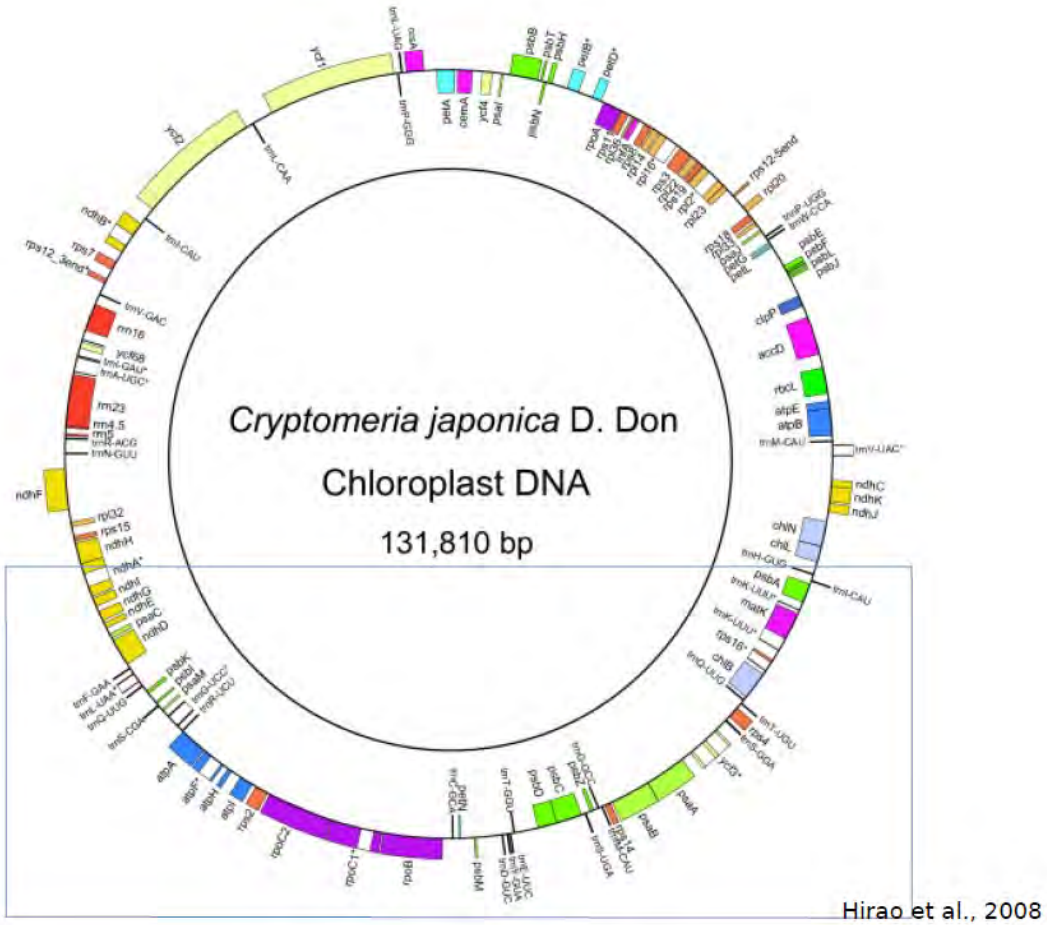
[1]Department of Genetics, University of Georgia, Athens, Georgia 30602; [2]NMFS/CZES, Genetics, 2725 Montlake Boulevard East, Seattle, Washington 98112; [3]Savannah River Ecology Laboratory, Drawer E, Aiken, South Carolina 29801; [4]Department of Microbiology and Immunology, School of Medicine, University of California, Los Angeles, California 90024; [5]School of Veterinary Medicine, Virginia Tech University, Blacksburg, Virginia 24046

BAR 21º13'S 43º46'W

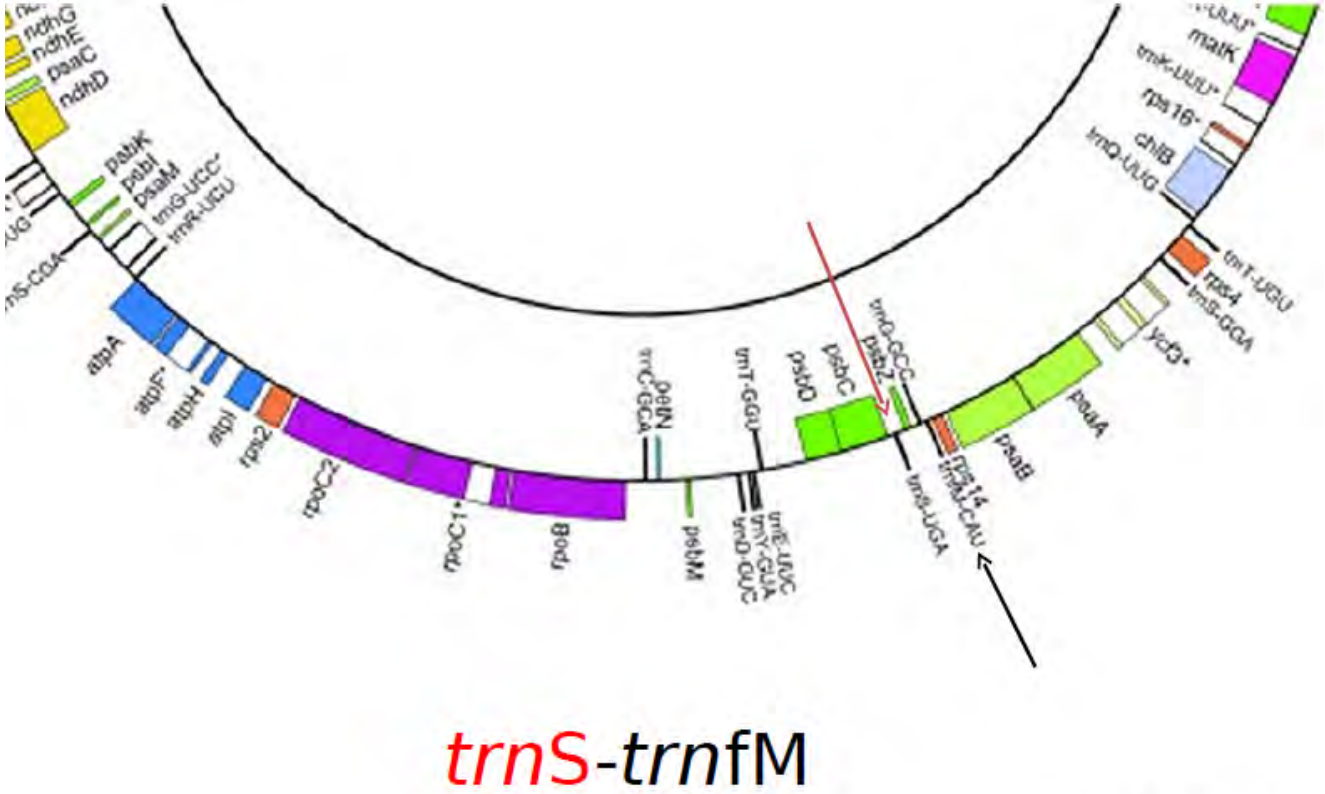
CPV- 30º30'S 53º29'W

Guerra et al., 2002

Chloroplast Variation



Hirao et al., 2008

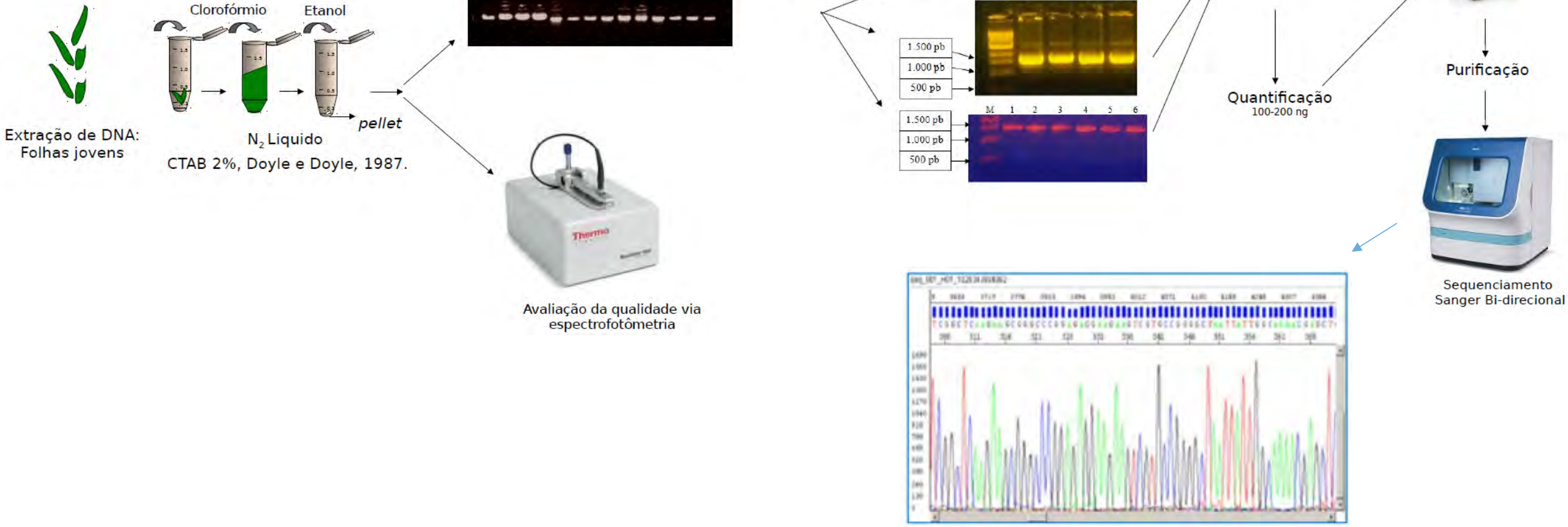


Hirao et al., 2008

Several Chloroplast Intergenic Spacers – Sanger sequencing

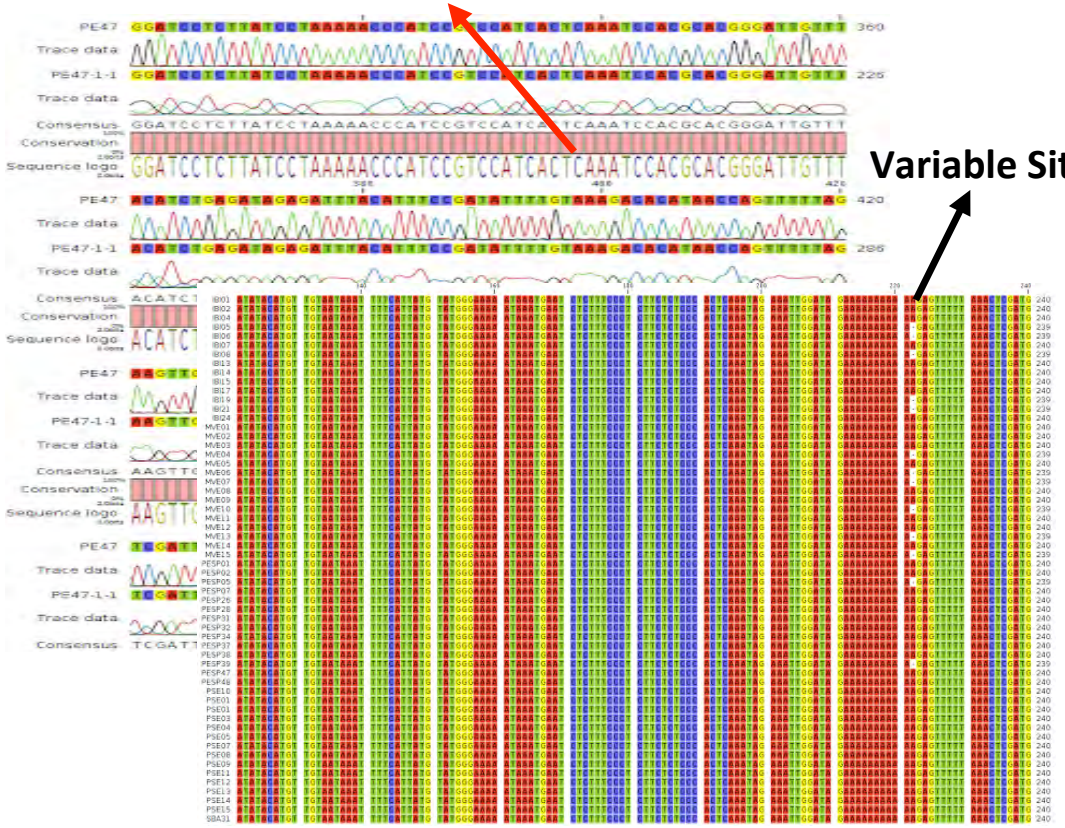
Laboratory workflow

Extensive lab work



Forward / - Reverse Alignment *Bioinformatics*

Bidirecional sequencing
Final Consensus sequence



29 variable sites in 2.492bp

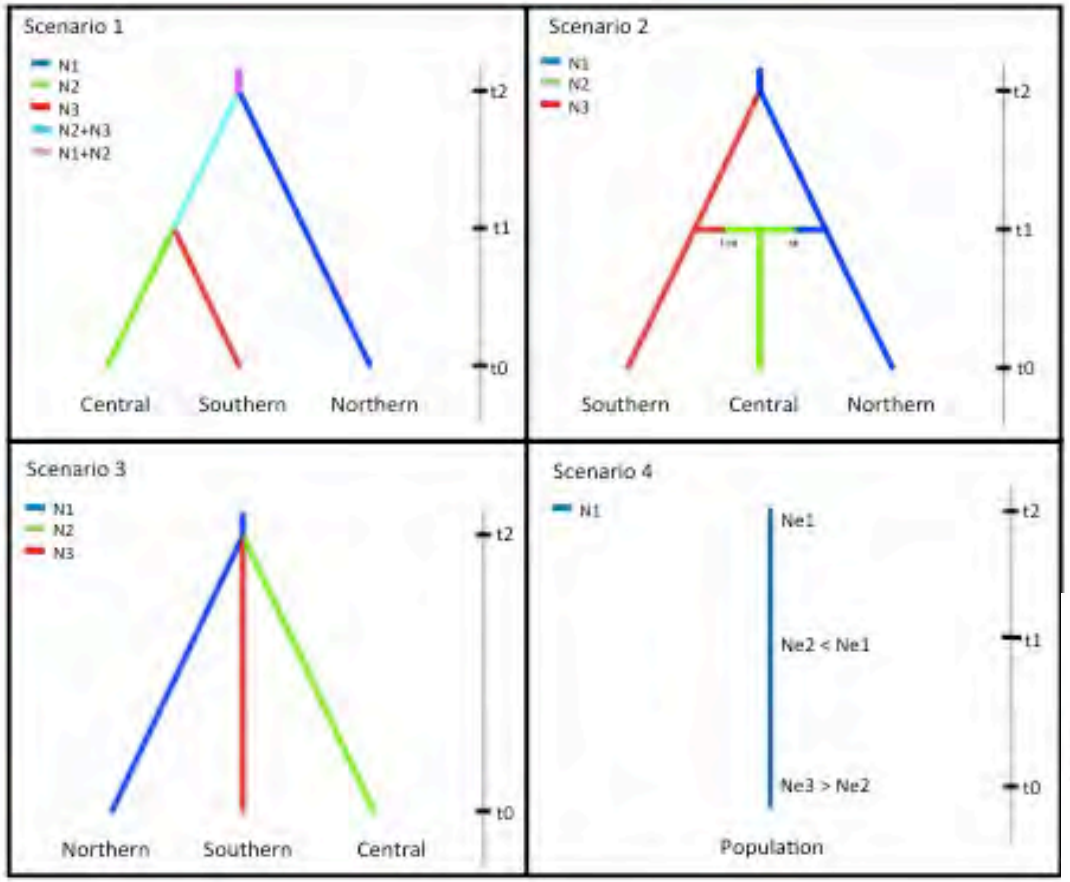
Table 2: Estimations of genetic diversity parameters for the pooled dataset and for the lineages

	Pooled dataset (n = 594)	Northern lineage (n = 116)	Central lineage (n = 195)	Southern lineage (n = 283)
Polymorphic sites	29	11	04	09
Indels	05	01	01	01
Haplotypes	H1-H24	H1-H3, H10, H17-H24	H4, H9, H12, H15 and H16	H4-H9 and H11-H14
Private Haplotypes	--	H1-H3, H10, H17-H24	H15 and H16	H5-H8, H11, H13 and H14
Overall θ_π^	5.133	7.018	1.821	1.157
Overall θ_S^	4.167	5.069	3.933	3.862

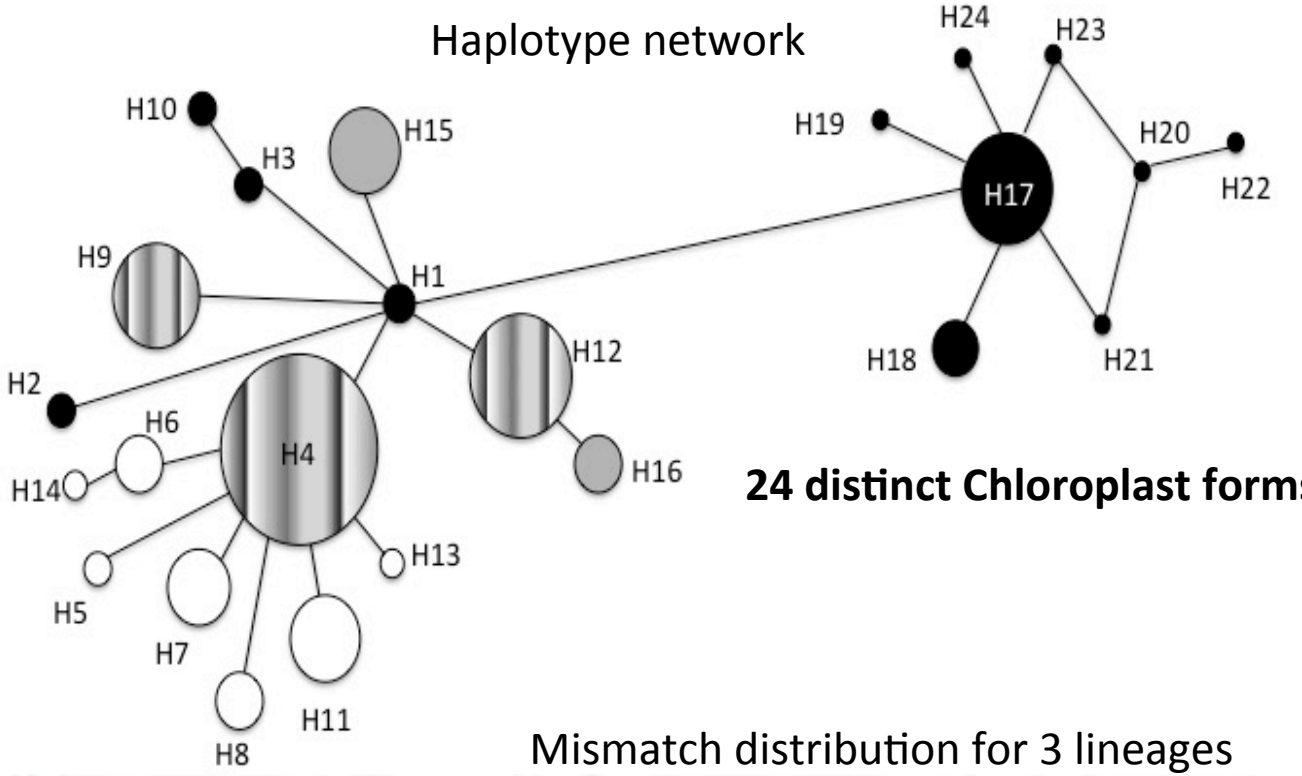
^ estimations performed grouping all individuals as a single population for the whole range or for each lineage

Araucaria angustifolia – Phylogeography

Demographic scenarios – ABC analysis

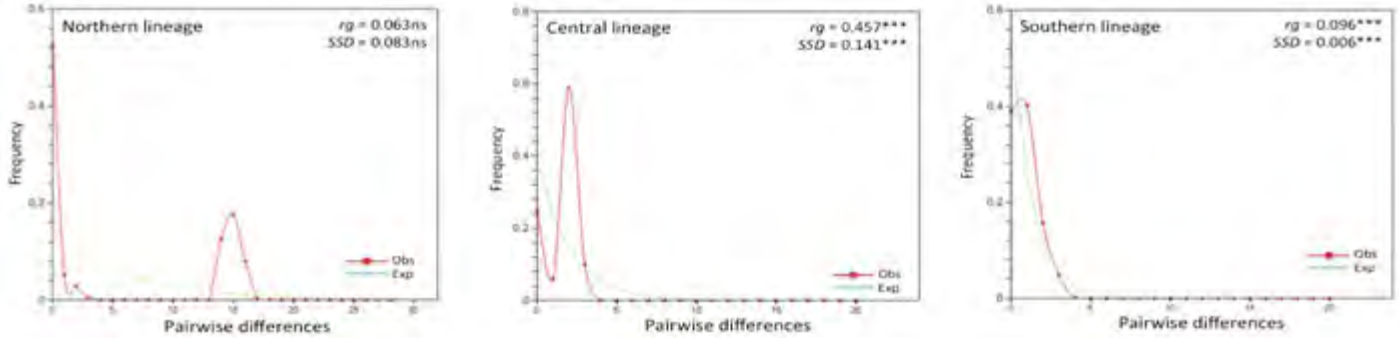


Haplotype network

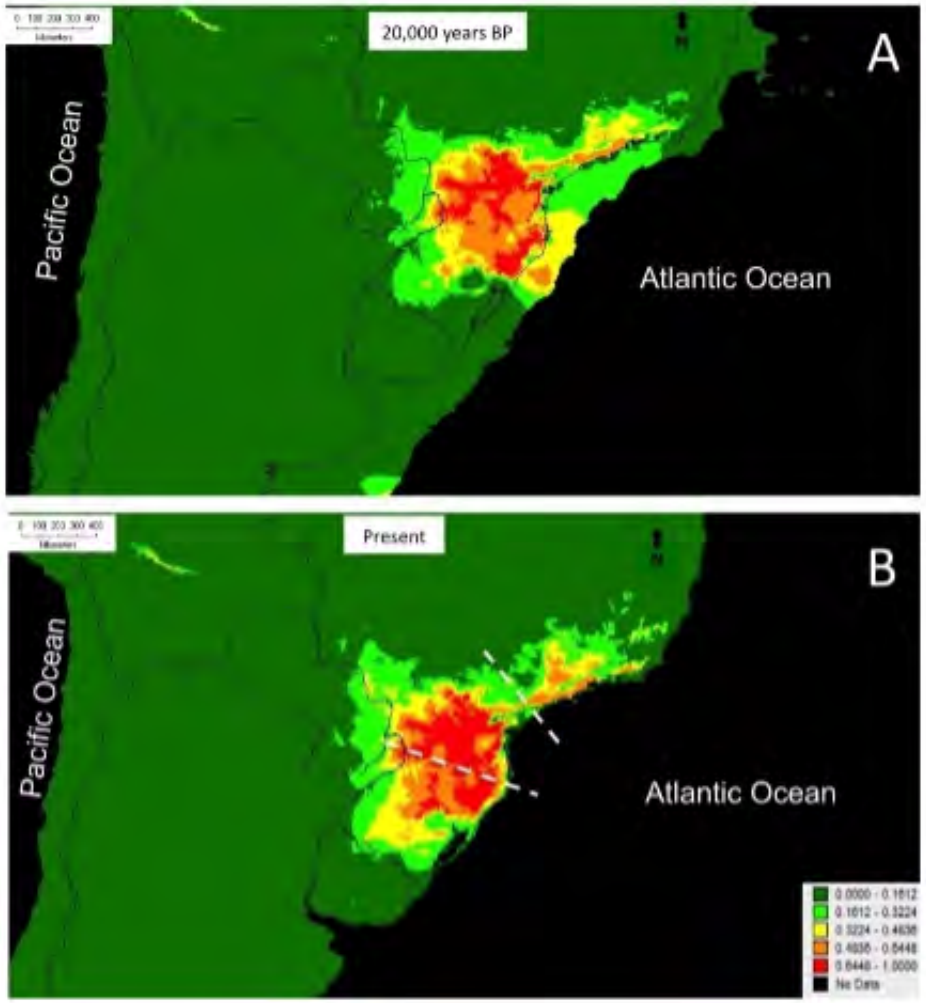


24 distinct Chloroplast forms

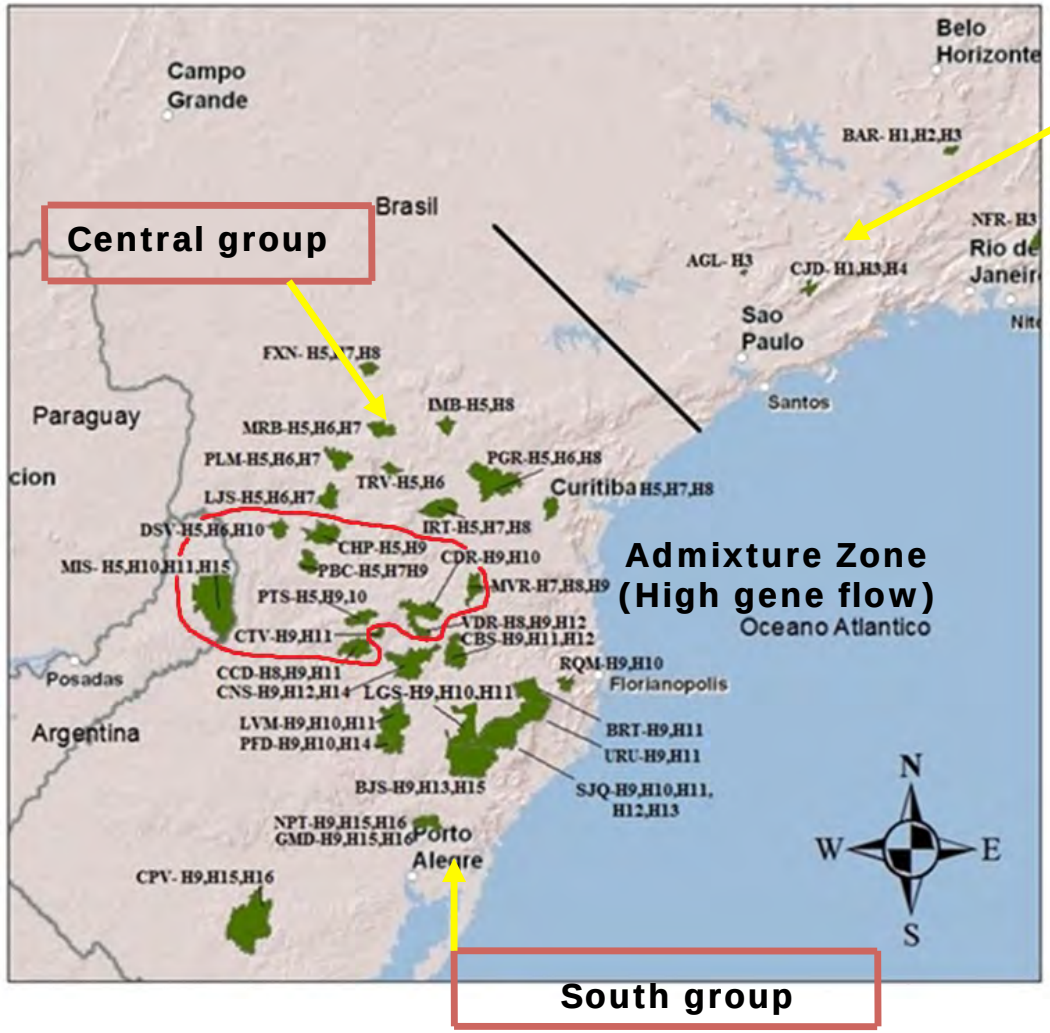
Mismatch distribution for 3 lineages



Araucaria angustifolia – Phylogeography



Species distribution modeling of *A. angustifolia*



North group

Barrier

Northern haplotypes are not shared with Central and Southern Variants

High Genetic differentiation between natural populations

Multiple Glacial Refugia

Central group

South group

Transcriptome Sequencing
RNA-Seq

Native Conifers
Araucaria angustifolia
Podocarpus lambertii

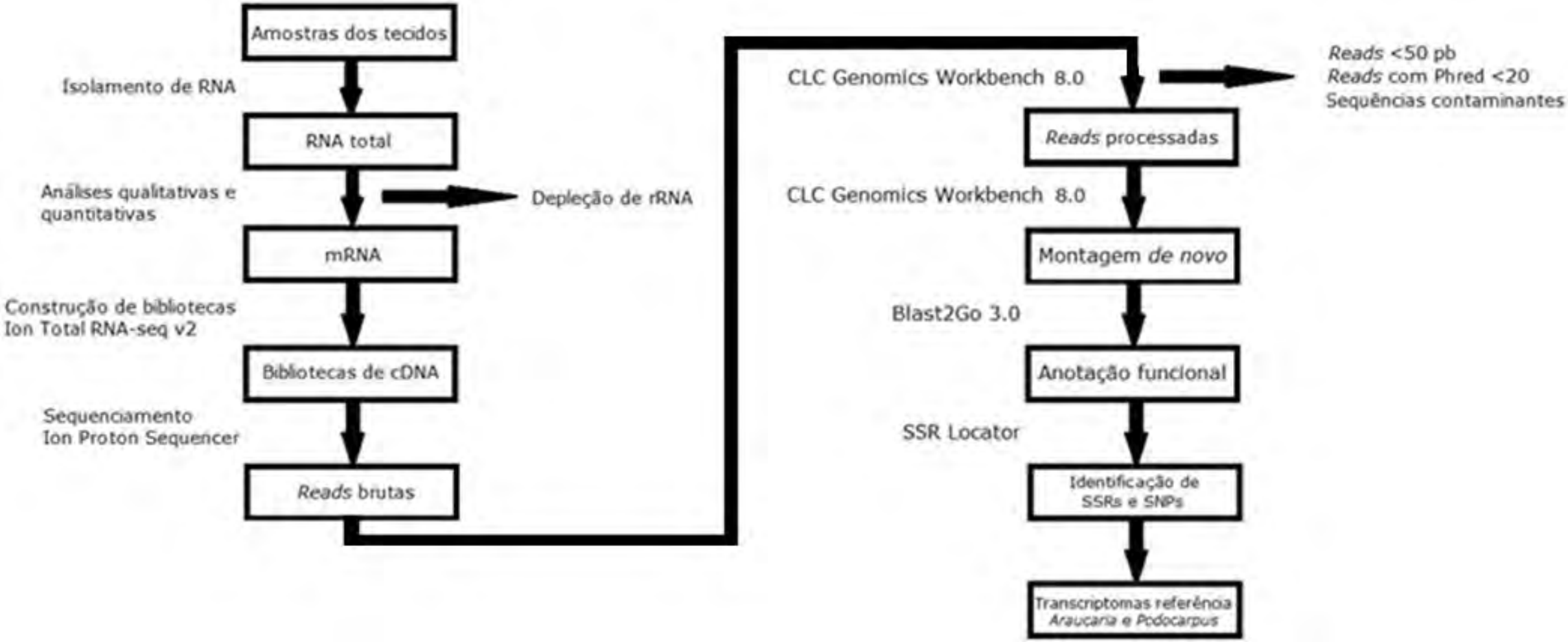




Transcriptome Sequencing

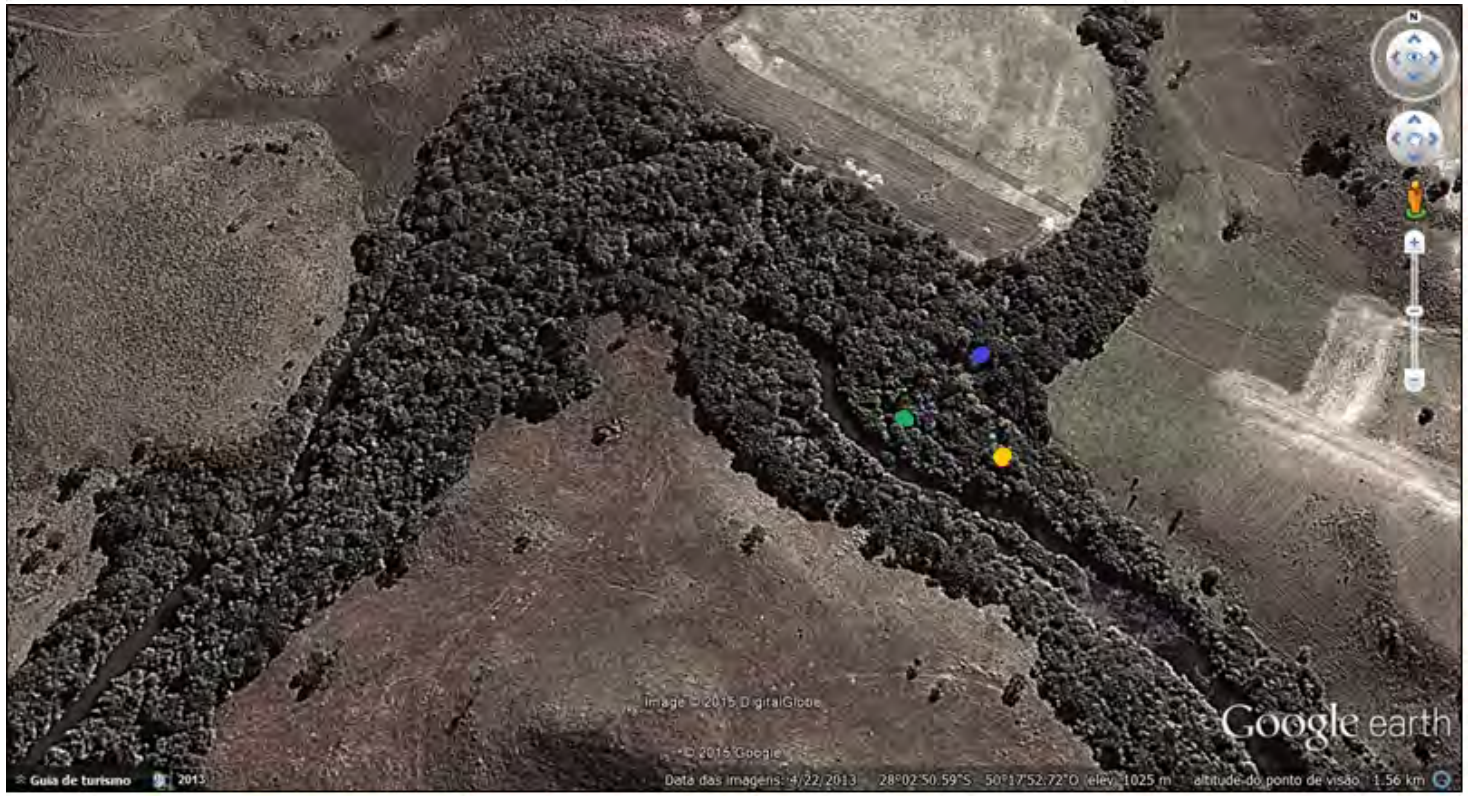
- Analysis of all mRNA expressed in a given tissue and condition
- Gene expression complete profiles (including miRNA and other RNAs)
- Highly dinamic profile and dependent on physiological and environmental conditions, cells, tissues, age, method of sampling, stabilization and storage
- Use of Next Generation Sequencing (NGS) Platforms for deep sequencing
- Alternative when the genome of the species of interest is not available (Creation of a gene catalog based on transcripts)
- Requires extensive bioinformatic training and analysis
- Excellent tool for the characterization of diversity and search for genes of interest for breeding

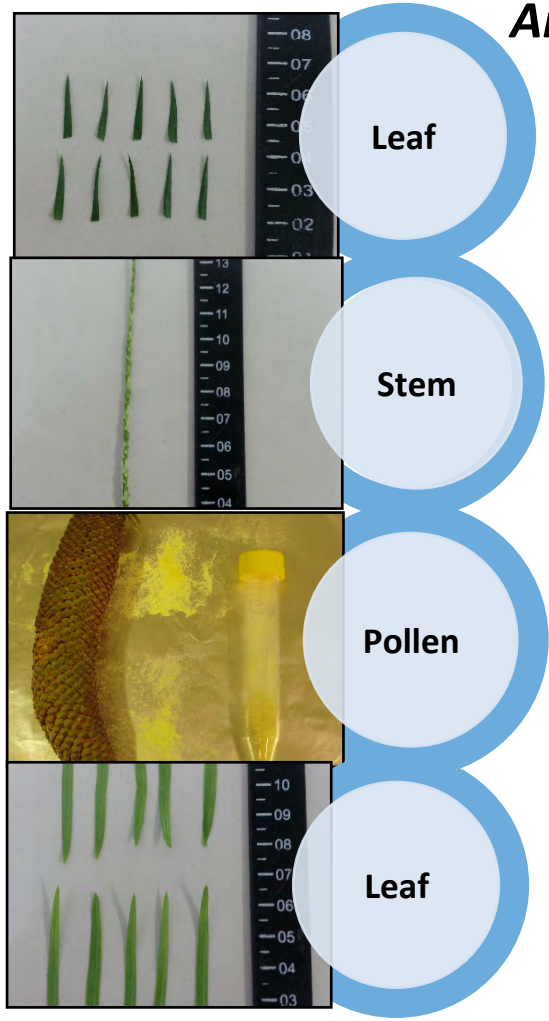
RNA-Seq Workflow
From field samples to complete transcriptomes



Field samples – Species Natural Environment

Tecido	Latitude / Longitude
Folhas - *Podocarpus lambertii*	28° 02′ 49′′ S / 50° 17′ 47′′ O
Folhas e hastes - *Araucaria angustifolia*	28° 02′ 50′′ S / 50° 17′ 50′′ O
Pólen - *Araucaria angustifolia*	28° 02′ 52′′ S / 50° 17′ 47′′ O



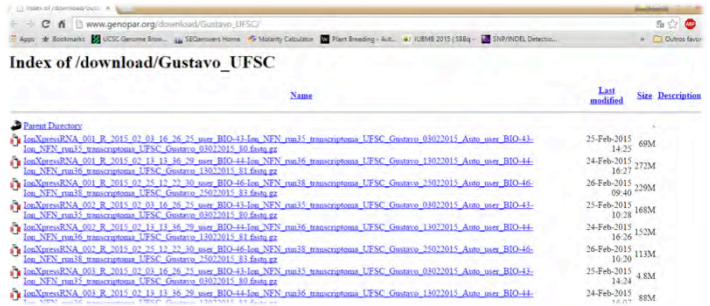


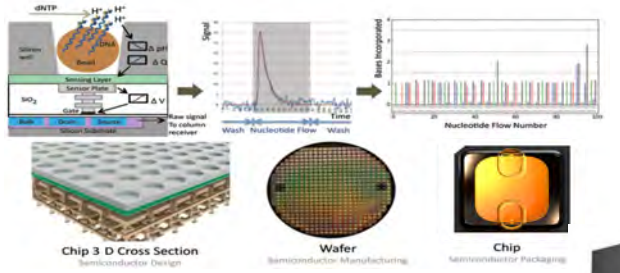
Araucaria angustifolia

Podocarpus lambertii

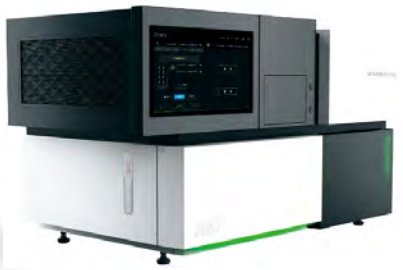
Transcriptomics

*100 Millions of raw reads per tissue
*Read Length = 80pb to (Ion Proton)
**De novo* and Refference based assembly (*Picea abies*)





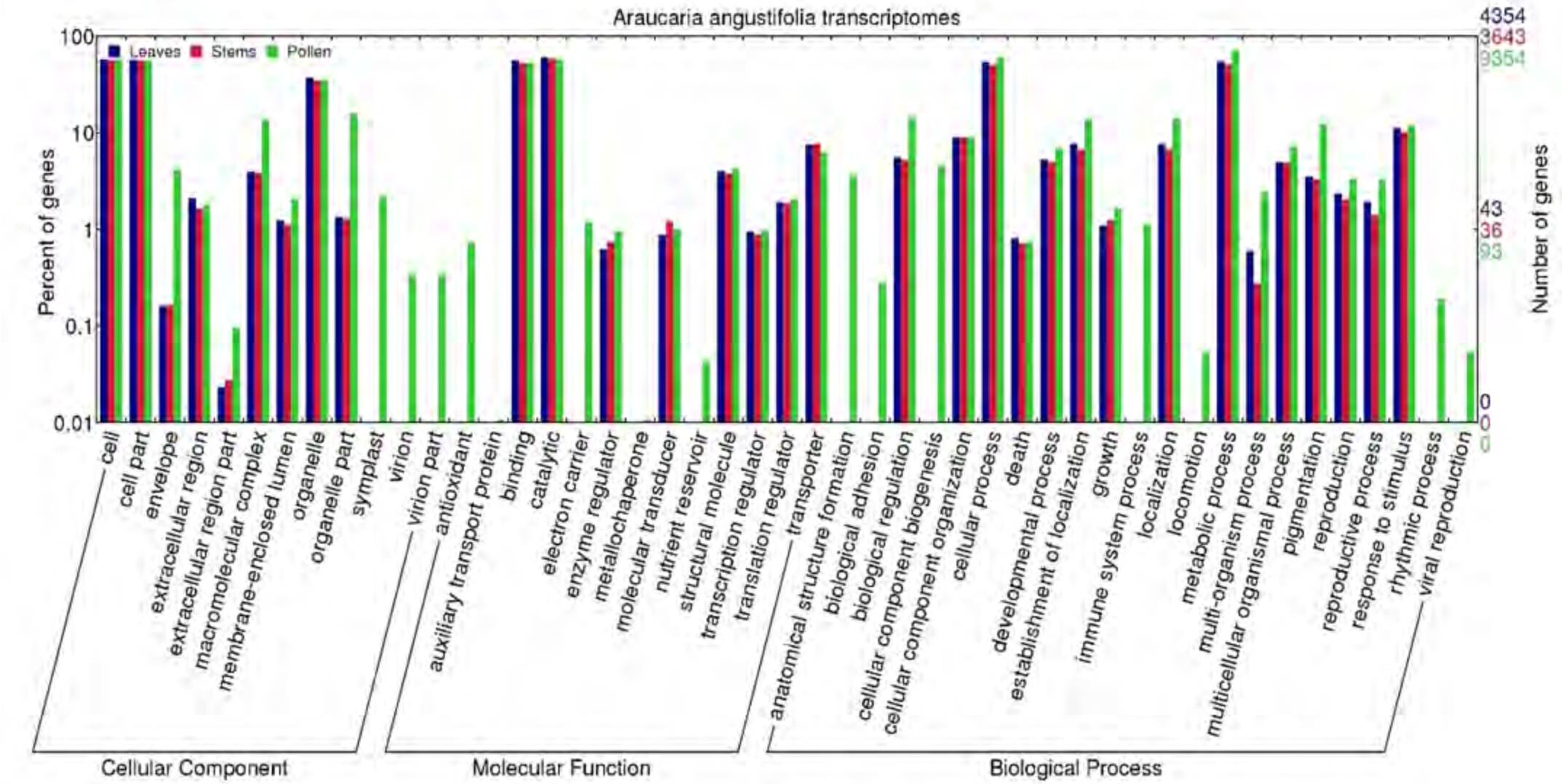
*Several NGS platforms available



Complete gene sequences from RNA-Seq data

	Annotated genes	Annotation Efficiency	GO lvl 2	GO lvl 3
P. lambertii Leaf	4.649	28,43 %	CC: 9	CC: 17
			MF: 8	MF: 12
			PB: 15	PB: 42
A. angustifolia Leaf	4.354	29,25 %	CC: 9	CC: 17
			MF: 8	MF: 12
			PB: 15	PB: 42
A. angustifolia Stem	3.643	28,43 %	CC: 9	CC: 17
			MF: 8	MF: 12
			PB: 15	PB: 42
A. angustifolia Pollen	9.354	33,45 %	CC: 12	CC: 40
			MF: 13	MF: 71
			PB: 22	PB: 196

Transcriptomics Functional annotation



Phenomics

- Large scale phenotype characterization
- Integration between factors

  environmental, genetic, epigenetic

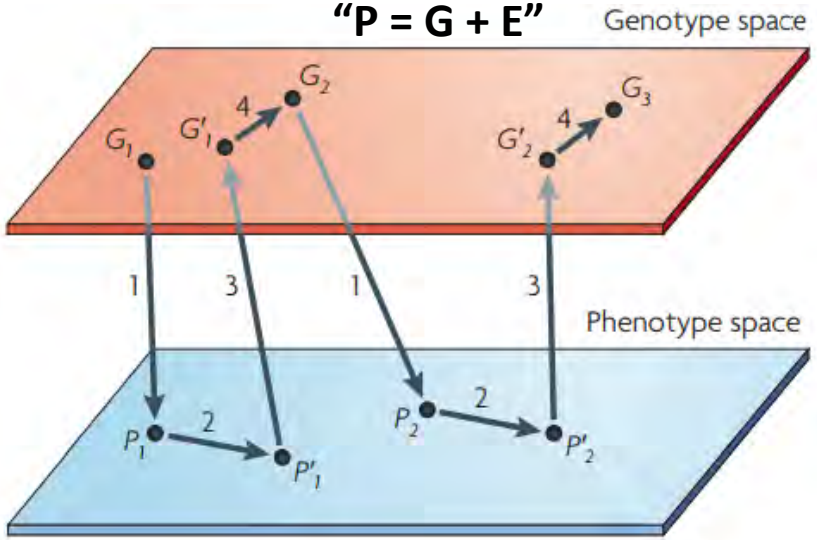
How genetic diversity affects the phenotypes?

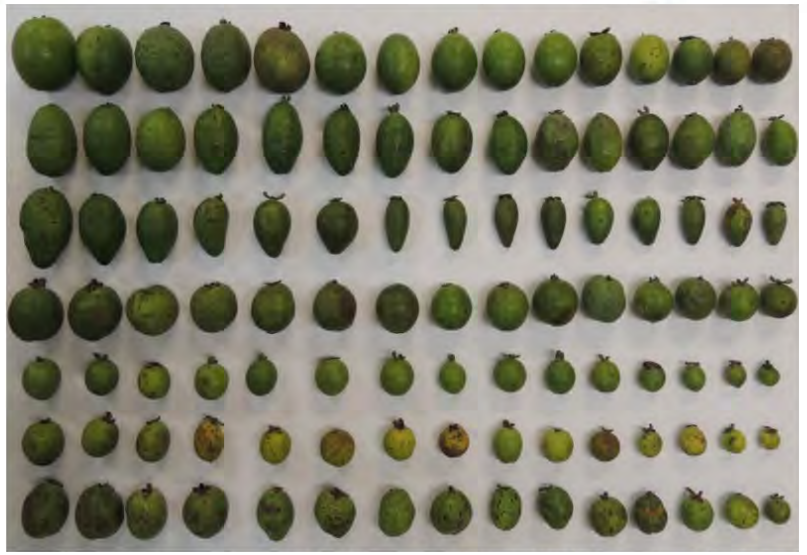
NATURE REVIEWS | GENETICS VOLUME 11 | DECEMBER 2010 | 855

Phenomics: the next challenge

David Houle, Diddahally R. Govindaraju[‡] and Stig Omholt[§||]*

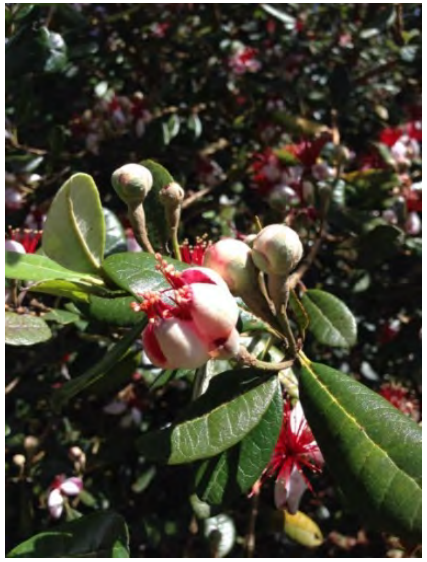
Abstract | A key goal of biology is to understand phenotypic characteristics, such as health, disease and evolutionary fitness. Phenotypic variation is produced through a complex web of interactions between genotype and environment, and such a 'genotype–phenotype' map is inaccessible without the detailed phenotypic data that allow these interactions to be studied. Despite this need, our ability to characterize phenomes — the full set of phenotypes of an individual — lags behind our ability to characterize genomes. Phenomics should be recognized and pursued as an independent discipline to enable the development and adoption of high-throughput and high-dimensional phenotyping.

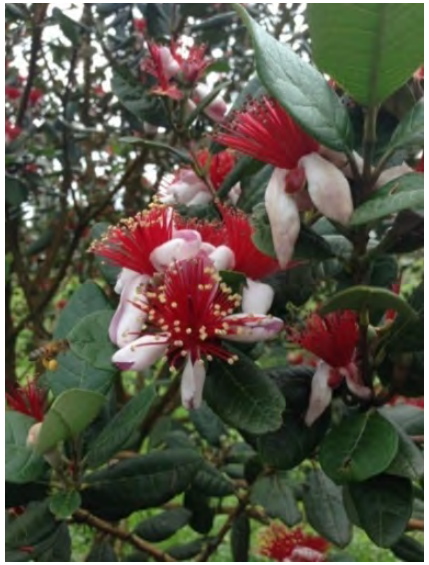




Borsuk, 2015

Acca sellowiana - Phenomics Tree analysis - Field





Evaluation period - 10 years

Variables

- Beggining of budding
- Beggining of flowering (10% open)
- End of flowering (90% fallen)
- Beggining of harvest
- End of harvest
- Flower index
- Fruit index

N= 235 genotypes

- Daily climate data

Acca sellowiana - Phenomics Fruit analysis - Lab







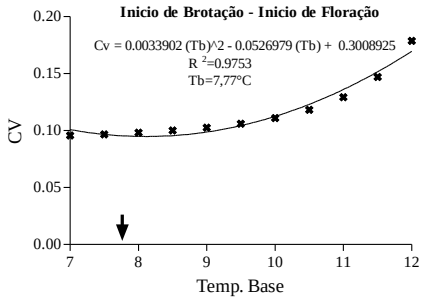
Evaluation period - 10 years

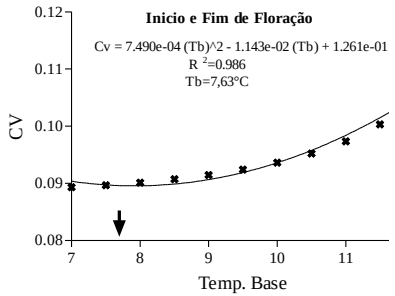
Variables

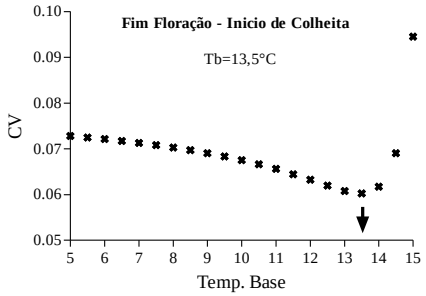
- Fruit diameter (cm),
- Fruit length (cm),
- Fruit weight (g),
- Peel weight (g),
- Pulp weight (g),
- Pulp yield (%),
- Peel thickness (cm),
- Brix - Bxº
- pH

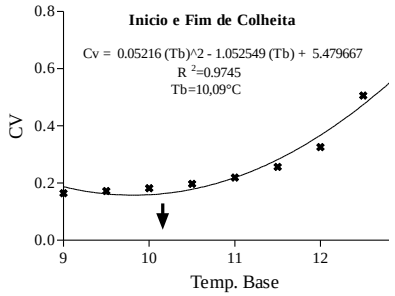
N= 235 genotypes

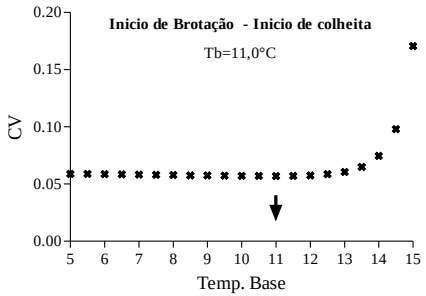
Acca sellowiana - Phenomics

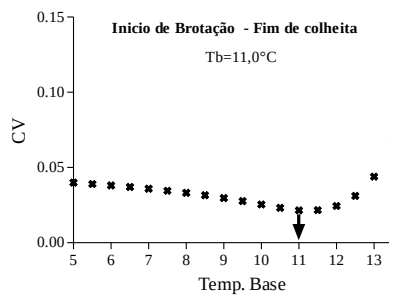


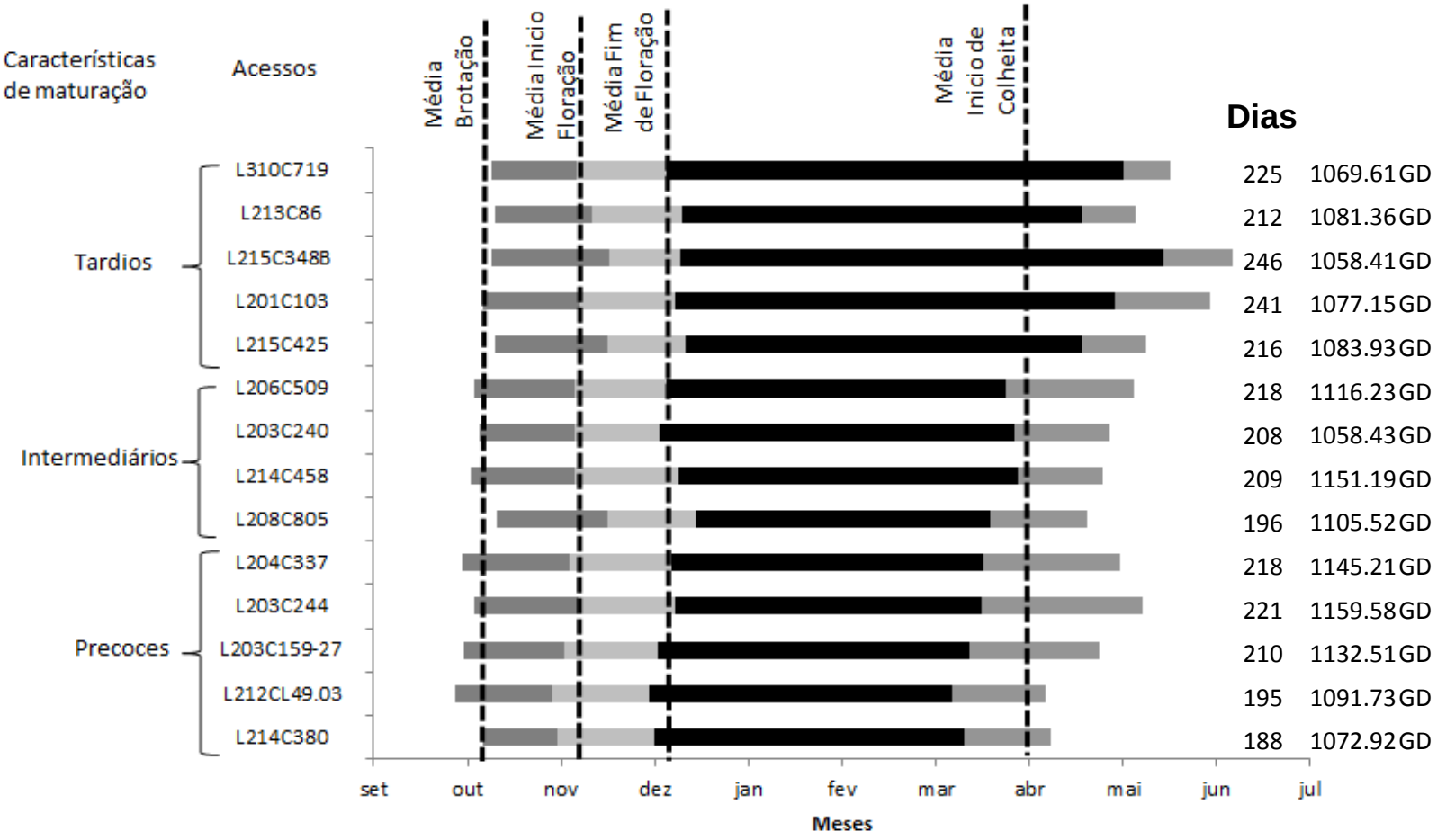












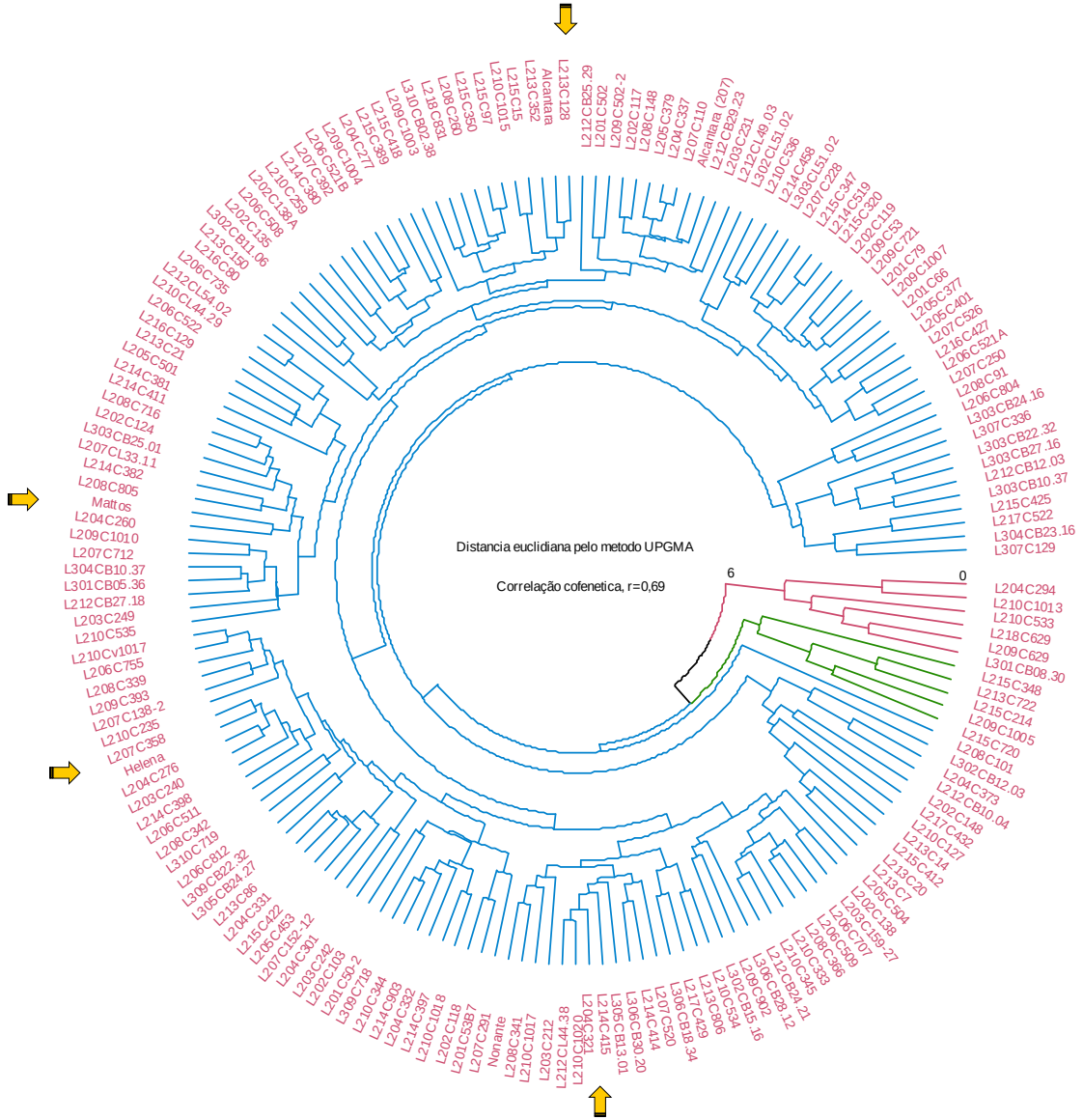
Acca sellowiana - Phenomics

Com a ACP e CLUSTER | Com o IS Método dos Ranks

Clones	Group
Alcantara	1
Helena	1
Mattos	1
Nonante	1
Outros 163 acessos	1
L204C294	2
L210C1013	2
L210C533	2
L218C629	2
L209C629	2
L301CB08.30	2
L215C[illegible]48	3
L213C[illegible]	3
L215C2[illegible]	3
L209[illegible]	3
L215C[illegible]20	3

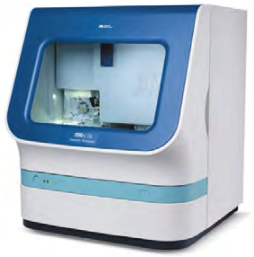
<u>Primeiro ranqueamento</u>:
L307C336,
L301CB08.30,
L209C629,
L217C522,
L303CB24.16,
L303CB10.37,
L310C719,
L208C716,
L206C812 e,
L215C425.

<u>Segundo ranqueamento</u>:
L210C534,
L217C522,
L303CB10.37,
L213C806,
L215C425,
L307C336,
L303CB27.16, L212CB12.03,
L203C249 e, L304CB23.16.

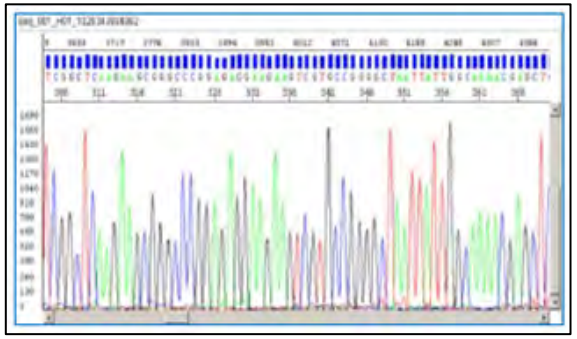


Phenomics + Genomics

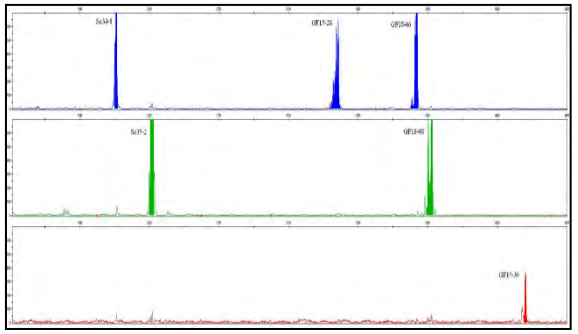
All Phenotypically evaluated cultivars/accessions are being sequenced and genotyped



Extensive set of informations for plant breeders and researchers

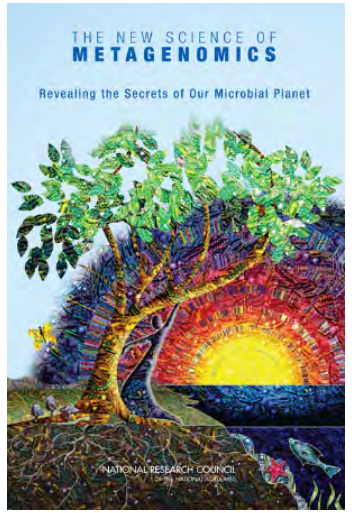


Less time to release new and adapted cultivars

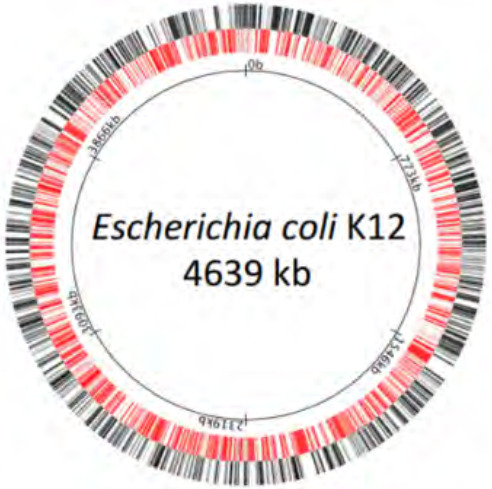


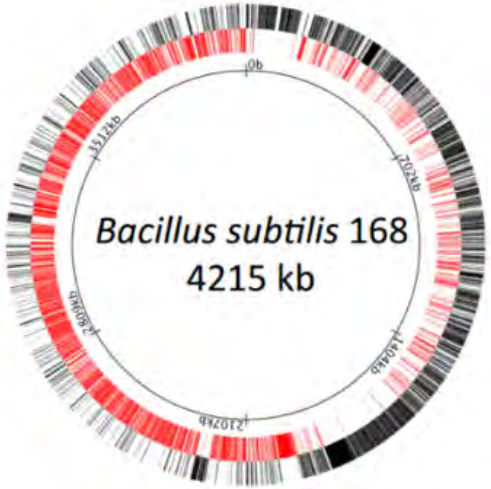
Metagenomics Associated genomes

- Application of genomic techniques for the study of microbial communities directly from entire samples;
- Dispensing the need of classic microbiologic techniques;

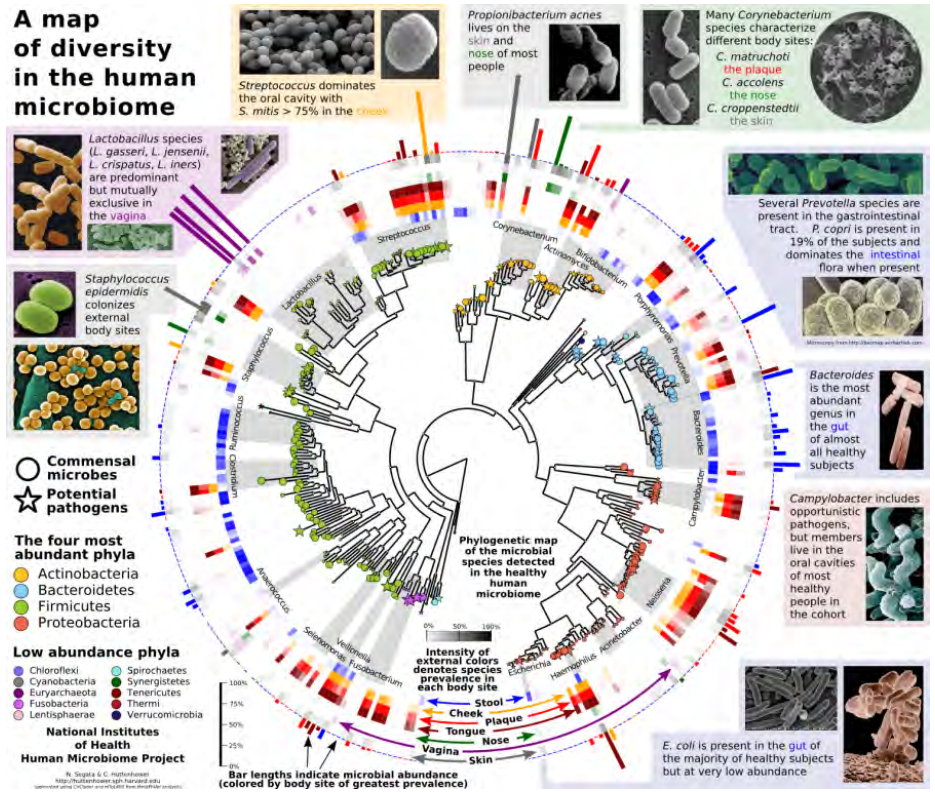


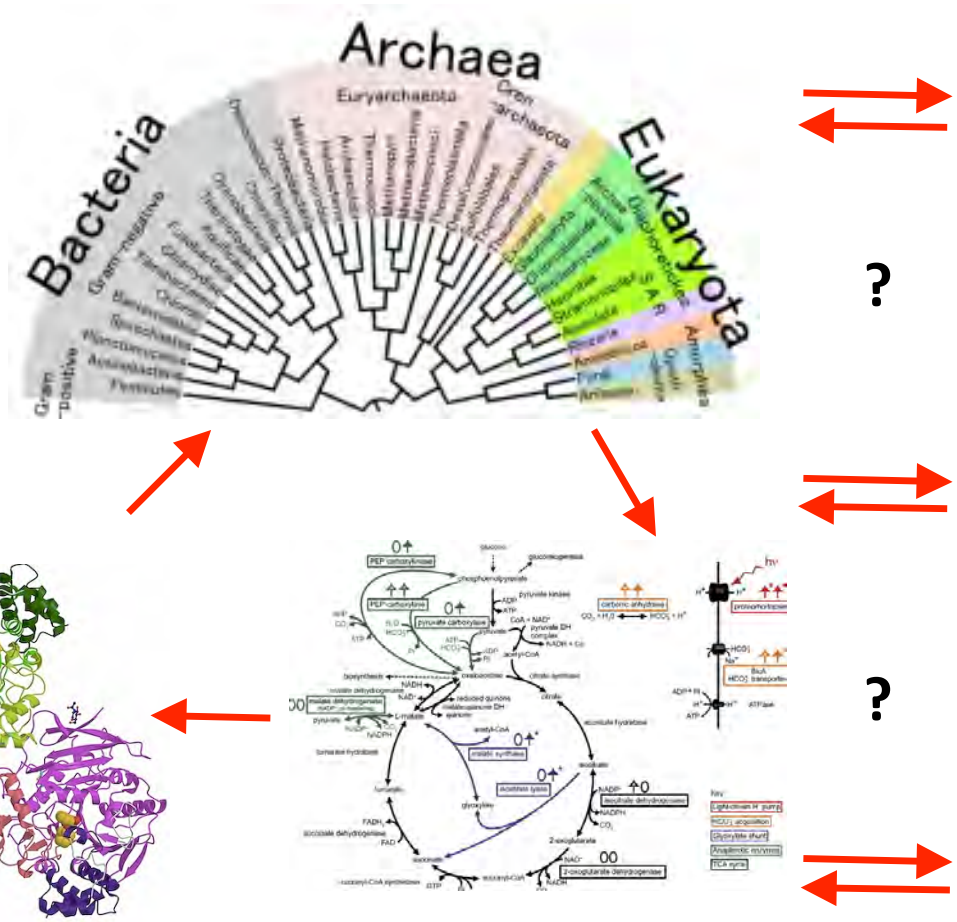
Complete genomes **or** **Metabarcode** 16S / ITS





Associated genomes metagenome



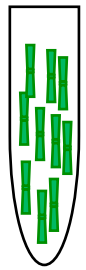


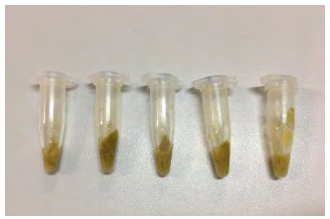
Associated genomes metabarcode workflow

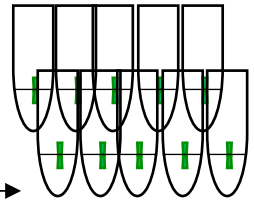
In vivo: 10 nodal segments/tube

In vitro: 10 nodal segments

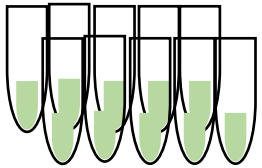
-80°C



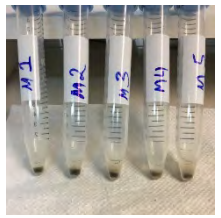




MS medium



30 days



Total DNA isolation and QC

PCR

Type	Region	Fragment
Arquea	16S	V4
Bacteria	16S	V3-V4
Fungi	ITS	ITS2

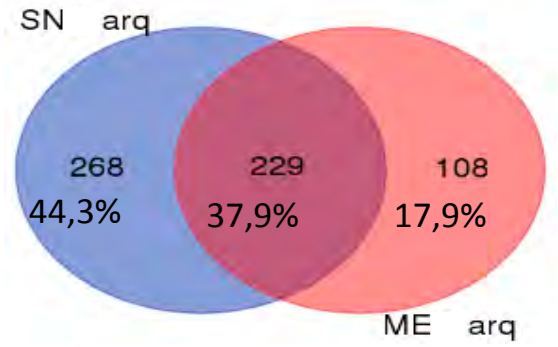
NGS - Illumina MiSeq

NGS Data processing and OTU characterization

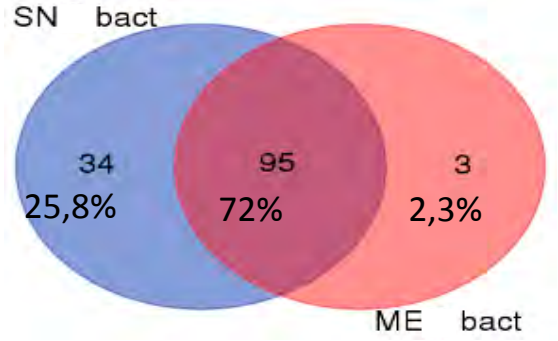
Associated genomes metagenome

Bacterial preliminary results

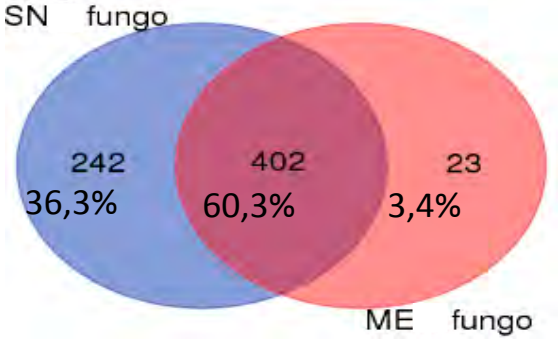
A

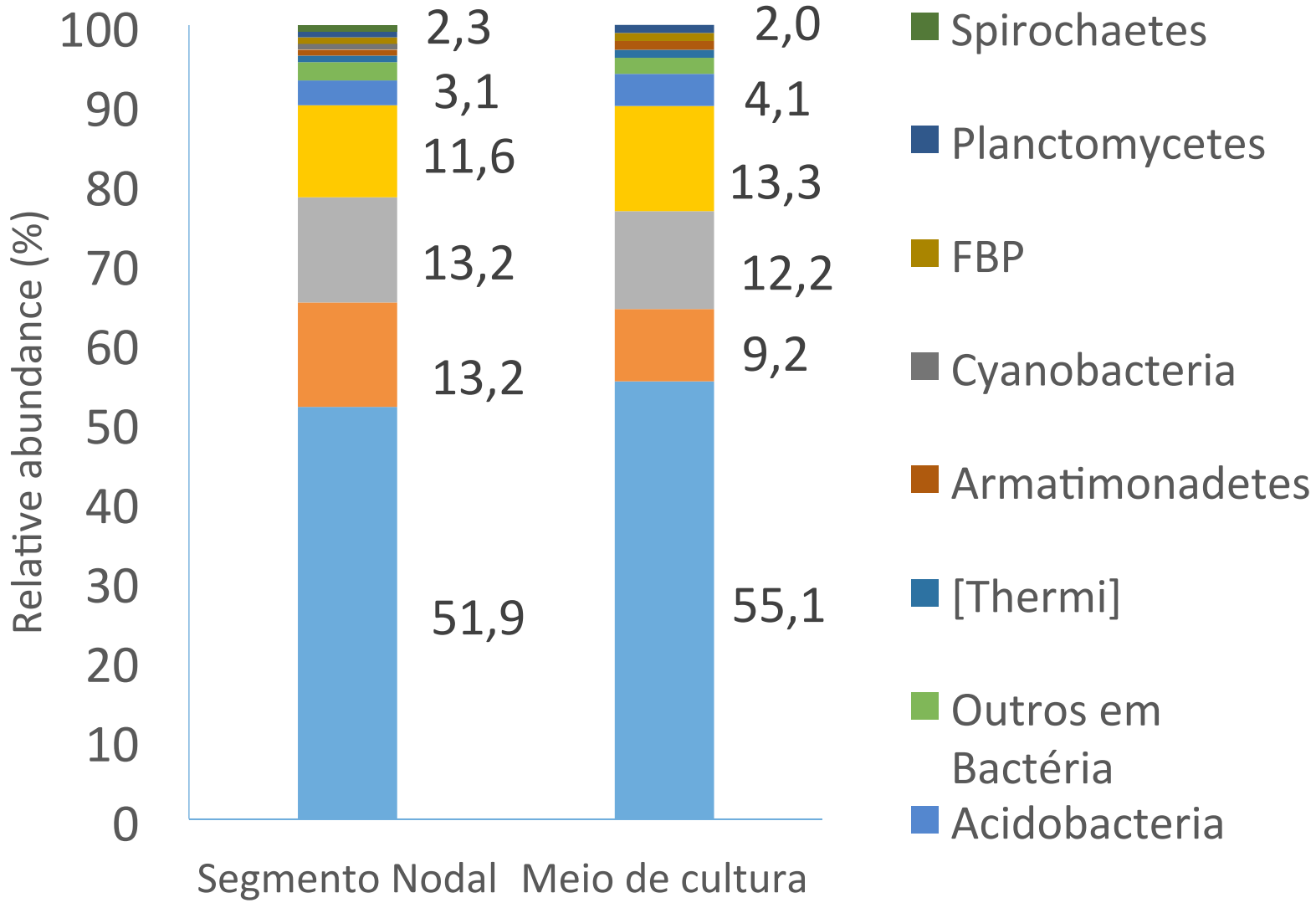


B



C





Thank you for your attention

Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal - LFDGV

Rgv.ufsc.br

Lfdgv.paginas.ufsc.br

Regional Seminar

"Advanced research on promissory edible plants in Latin America: tools to improve Food Security in the region"

Multi-omics approach in plant characterization, domestication and breeding in a food safety context.

Dr. Gustavo H.F. Klabunde

klabunde.gustavo@gmail.com

Agronomist - Post-Doc Research Fellow – UFSC, Brasil

Dec 4th, 2017