

Genética de poblaciones y mejoramiento de cultivos utilizando tecnologías de secuenciación de alto rendimiento

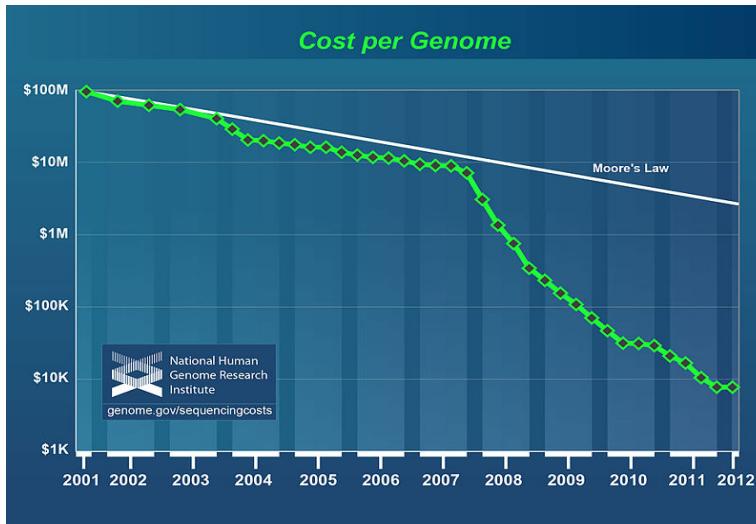
Jorge Duitama



The 9 billion-people question

A special report on feeding the world | February 26th 2011

<http://www.agriculturesnetwork.org/news/the-9-billion-people-question>



Especie	Tamaño del genoma (Mbp)	# de muestras por US\$ 10K*
Humanos	3200	2
Yuca	760	7
Frijol	520	9
Arroz	394	10

* La bioinformática se vende por separado

<http://money.cnn.com/2013/06/25/technology/enterprise/low-cost-genome-sequencing/index.html>
http://en.wikipedia.org/wiki/File:Genome_sequencing_costs,_May_2012.jpg



AA^C_G^G_G^A_G^A_T_G_C_C_G_C_A_A_T_T_C_A
TGT^C_G_T_C_G_C_C_A_T_T_C_A_T_G_G_G_T_G_A
CTTGG^A_G_T_C_T_G_G_G_T_A_G_A_C_A_A_G
AA^T_T_C_A_A_G_A_G_C_C_A_A_G_T_A_A_G_C_G
GGGT^A_T_G_G_A_A_T_C_G_C_C_A_T_C_G_A_C
CCATT^G_A_A_A_G_T_G_G_G_T_T_A_A_G_G_C
GAAGGAC^B_G_G_G_T_T_T_C_G_A_T_C_C_G_C_A_C
GATAGGGAAAGAGAGAGGGTCCGAAAAA
CAGTCCC^G_G_G_T_G_G_A_A_G_T_G_A_A_A_C_G_G_A_G_G_C
GGAGTGTTATGCTACCGCTGCTG

Infraestructura



Integración

Genomic variability

Phenotypes

Gene networks

Related species

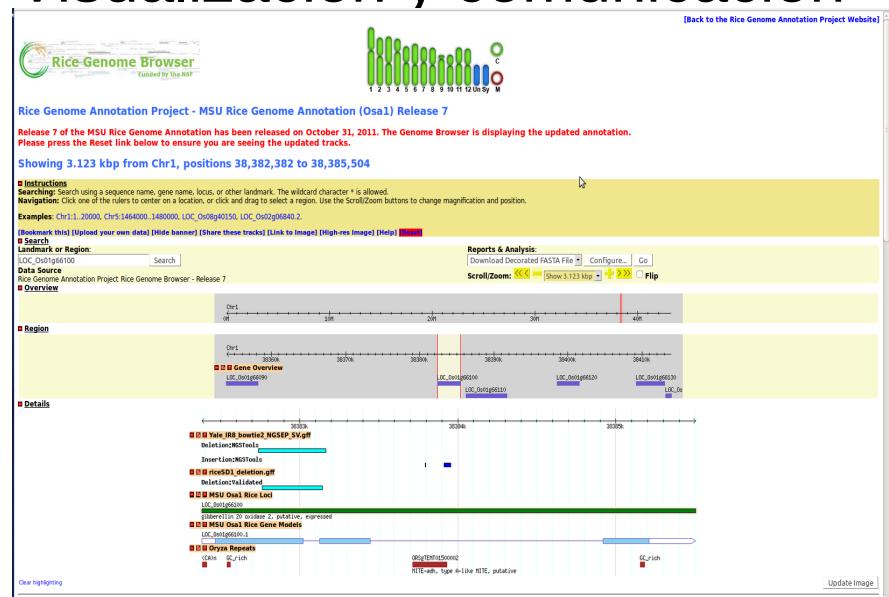
Environment

...

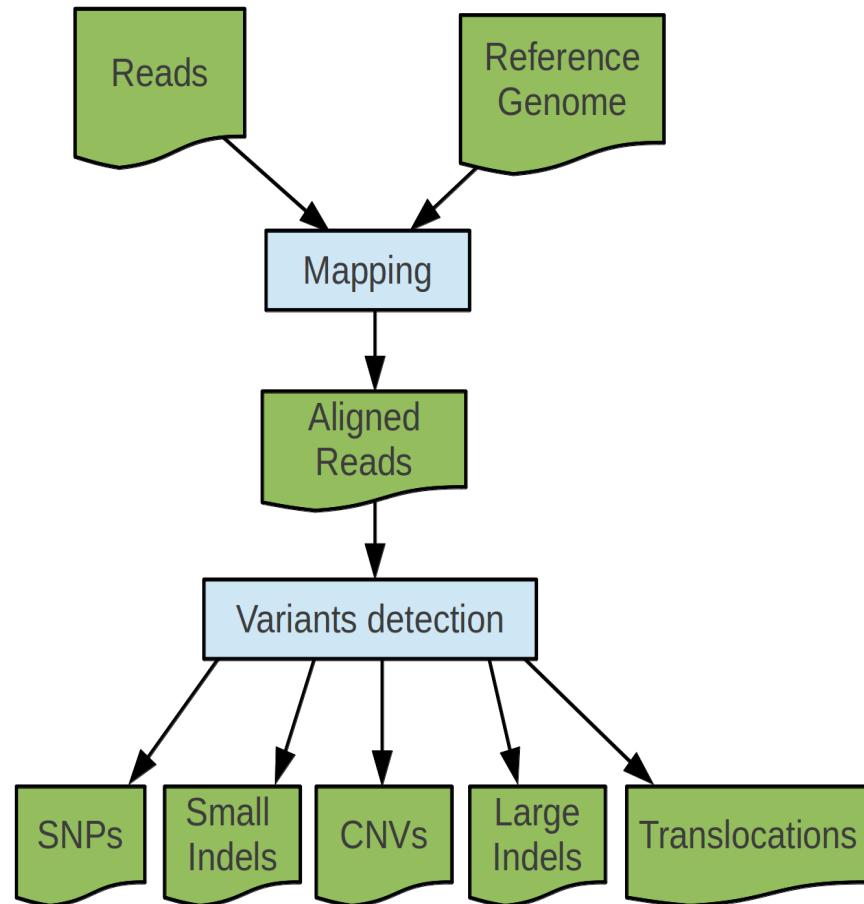
Análisis



Visualización y comunicación

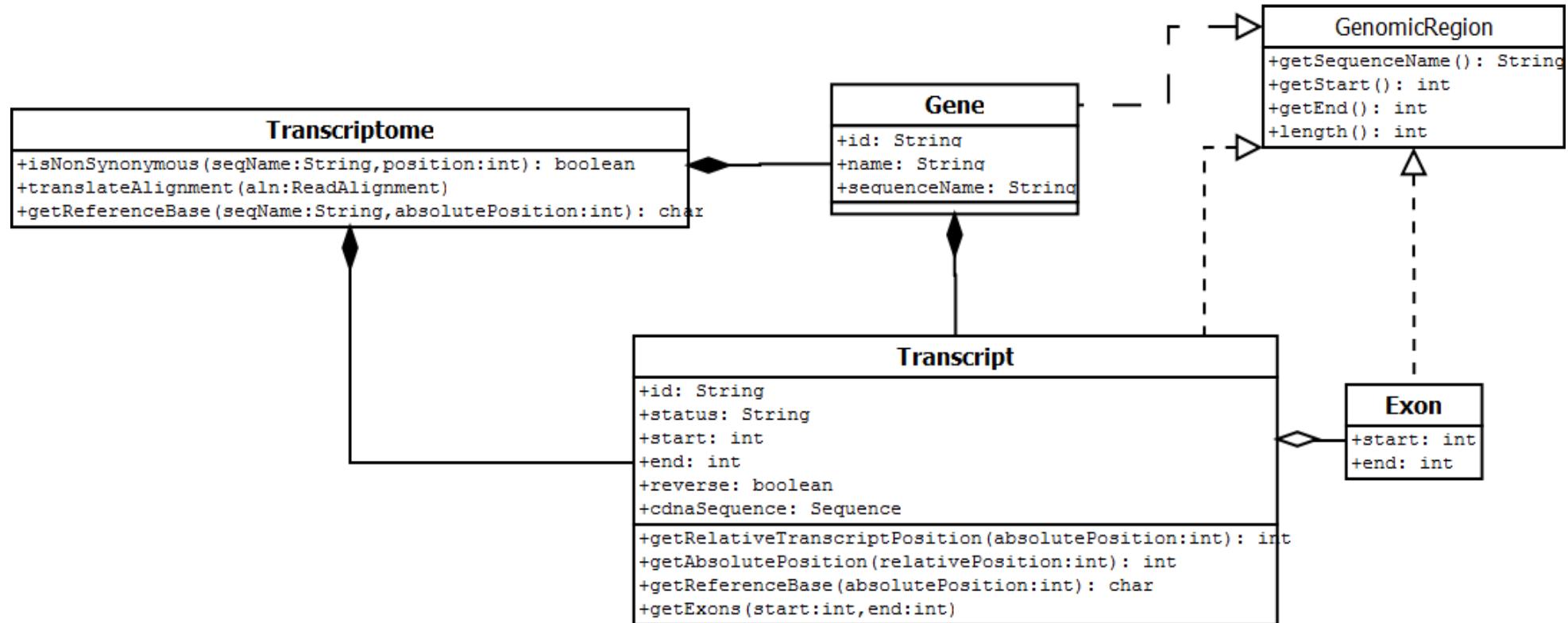


Análisis de lecturas de HTS

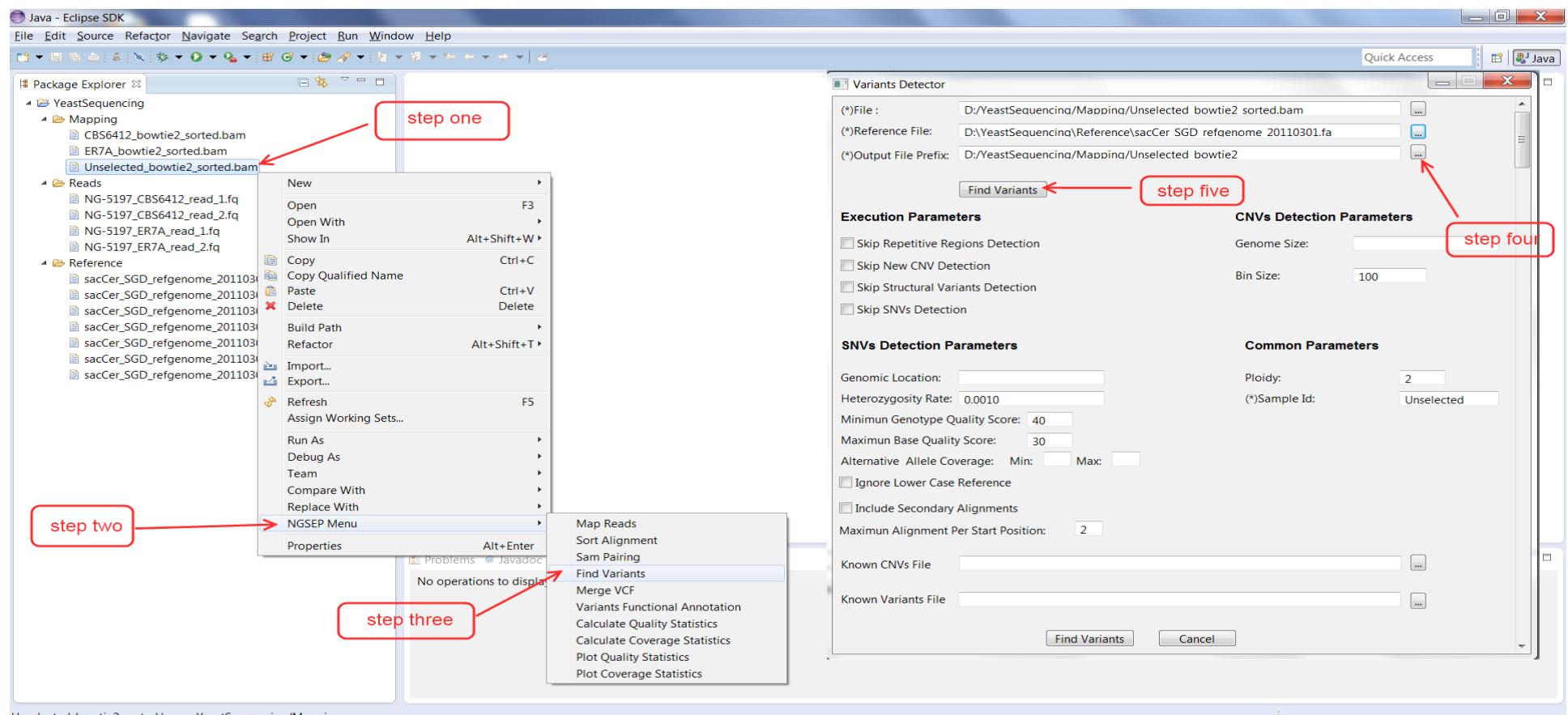


- Diferentes tipos de software
- Lenguajes de programación heterogéneos
- Problemas de:
 - Calidad
 - Soporte
 - Extensibilidad
 - Integración
 - **Usabilidad**

Análisis de lecturas de HTS



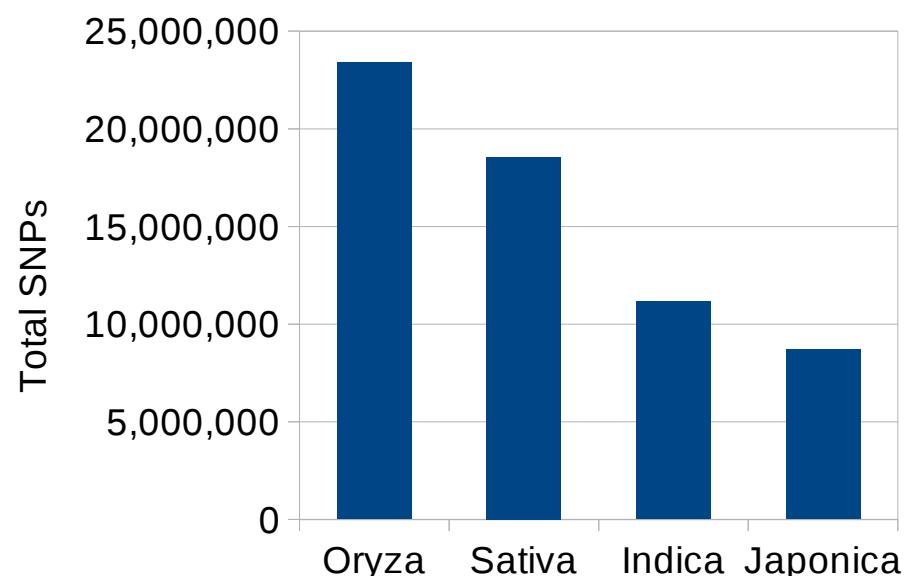
NGSEP



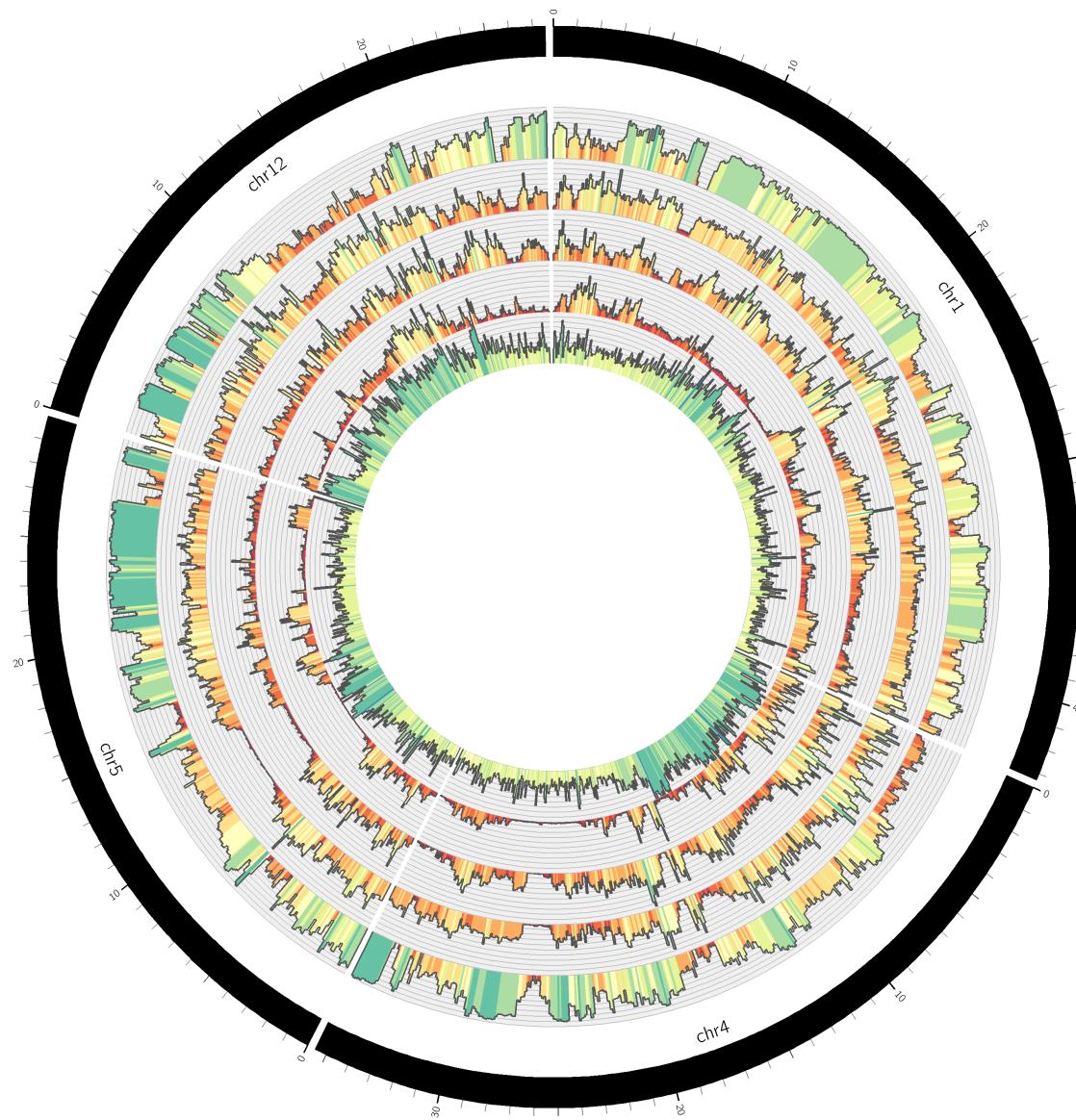
Software disponible en <http://sourceforge.net/projects/ngsep/>
 Artículo publicado en NAR methods online:
<http://nar.oxfordjournals.org/content/42/6/e44.full>

Secuenciación de genoma completo en Arroz

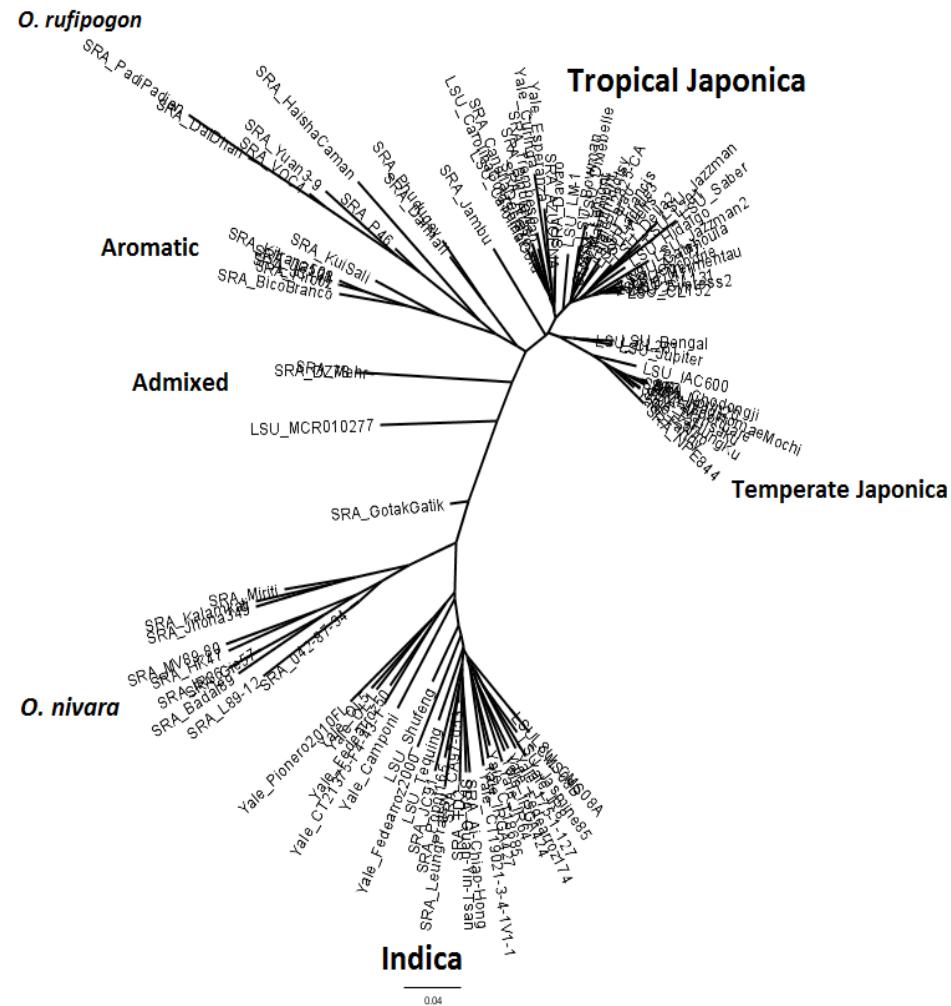
	CIAT	LSU	CAS*	Total
Indica	15	5	6	26
Tropical Japonica	4	23	9	36
Temperate Japonica	2	1	9	12
Admixed	0	4	4	8
Aromatic	0	0	6	6
Aus	0	0	6	6
Rufipogon	0	0	5	5
Nivara	0	0	5	5
Total	21	33	50	104



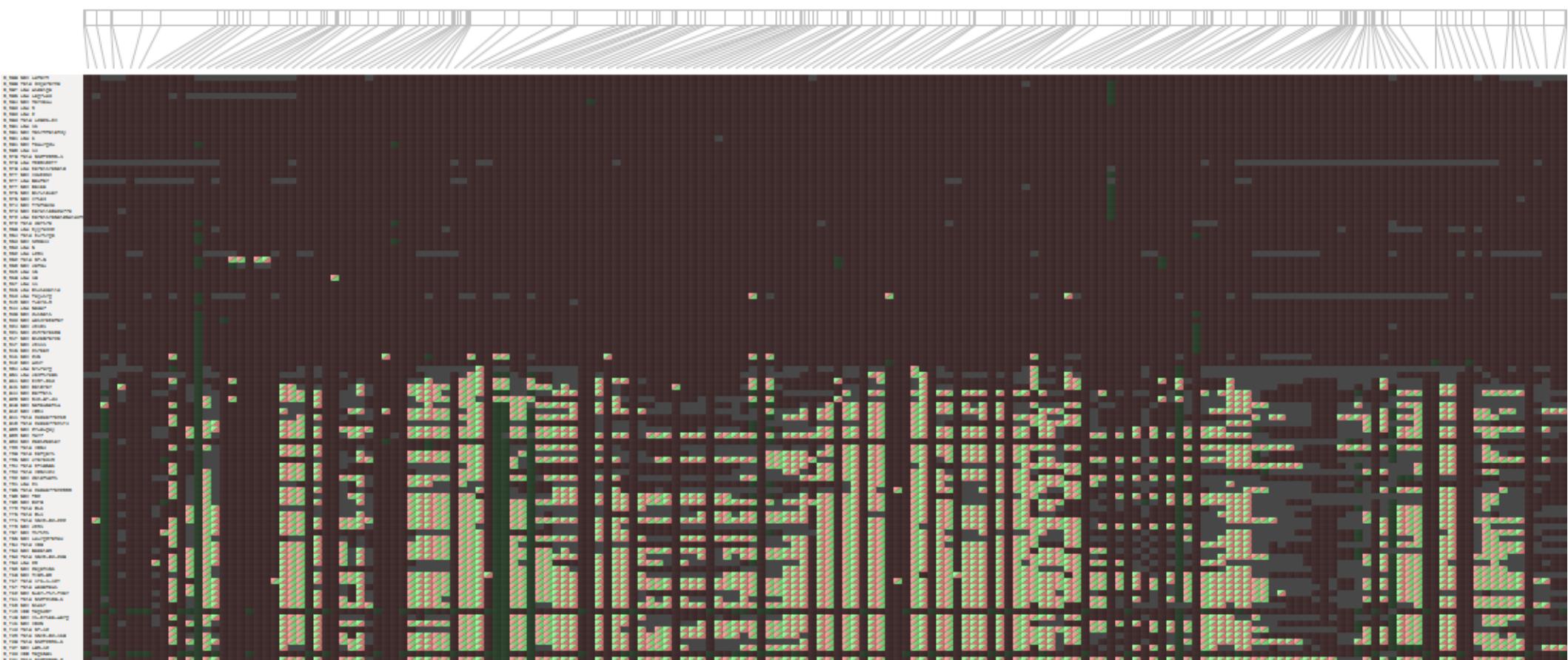
* Xu,X., Liu,X., Ge,S., Jensen,J.D., Hu,F., Li,X., Dong,Y., Gutenkunst,R.N., Fang,L., Huang,L., et al. (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat. Biotechnol., 30, 105–111.



Filogenia



Variación de Número de Copias (CNVs)



Estructura Poblacional de Arroz

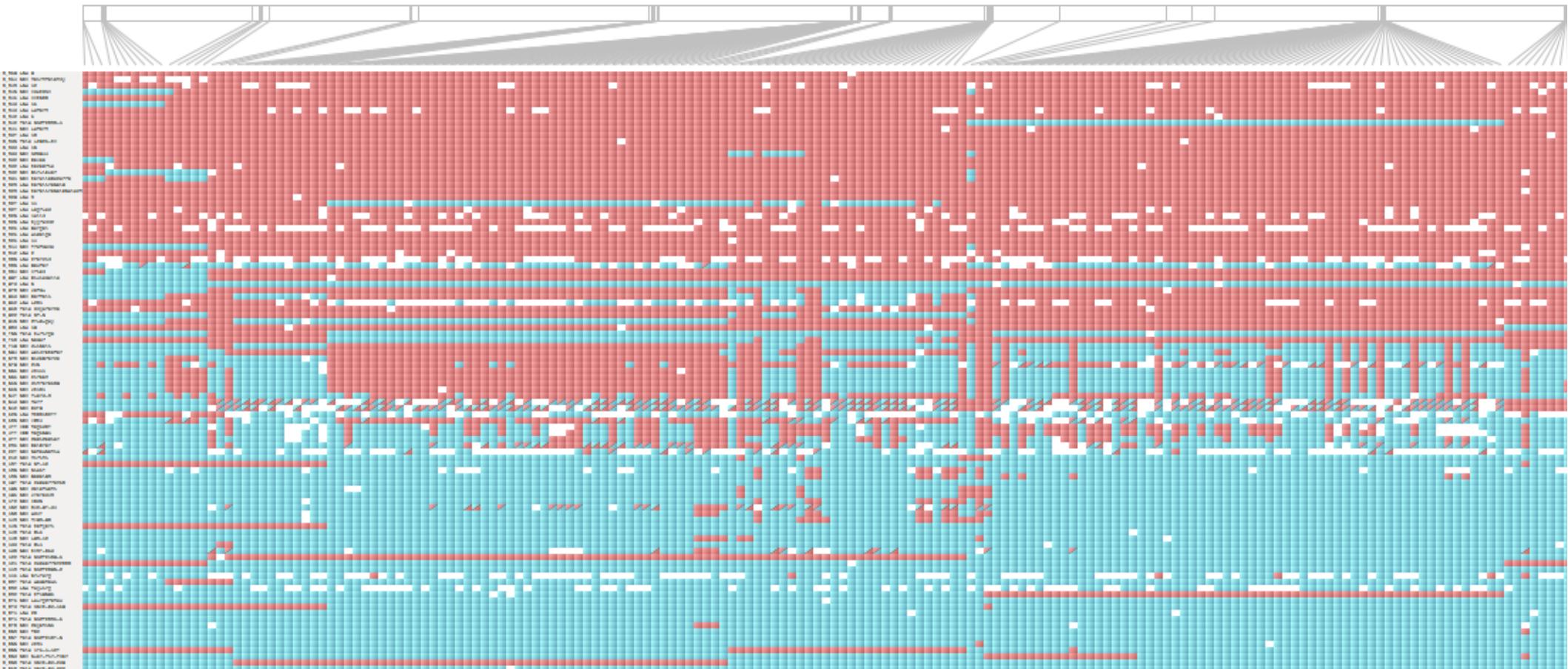


Temperate
Japonica

Tropical
Japonica

Aro Aus Indica

Introducciones individuales



SNPs para selección asistida por marcadores

Rice SNPs in grain quality and plant height genes.

Trait	Gene	# SNPs	Status
Gelatinization Temperature	Alk	5	* 3 SNPs associated with GT.
	ISA	19	
	SBE1	8	
	SBE3	24	
	SSI	24	* 1 SNPs associated with amylose variation.
Amylose content	SSIII	26	
	Waxy	8	* 6 SNPs associated with amylose variation.
White belly	APL2	9	
	PPDKB	20	
Plant height	sd1	1	*, but related to indica/japonica origin
	Total	144	

Rice Hoja Blanca Virus Project

Trait	# SNPs	Status
Resistance to RHBV and/or vector mechanical damage	172	Parental genotypes testing in process

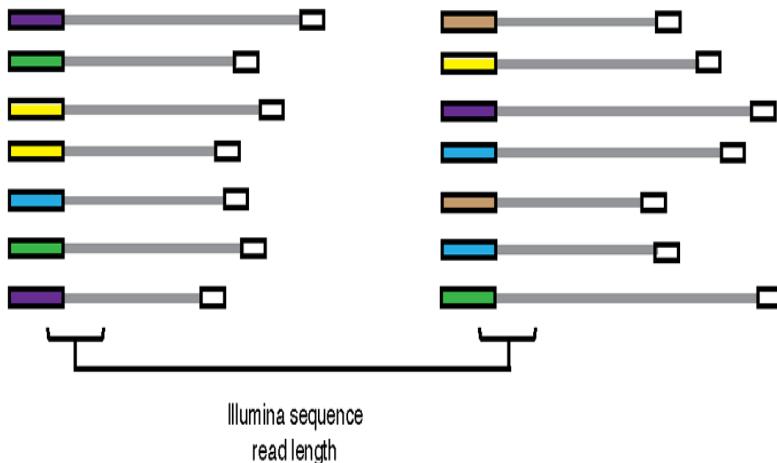
Breeding *O. sativa* x *O. glaberrima*.

188 SNPs for background selection

Yield

Trait	Gene	# SNPs	Status
Grain number/panicle	Gn1a/OsCKX2	37	Gene introgression in process.

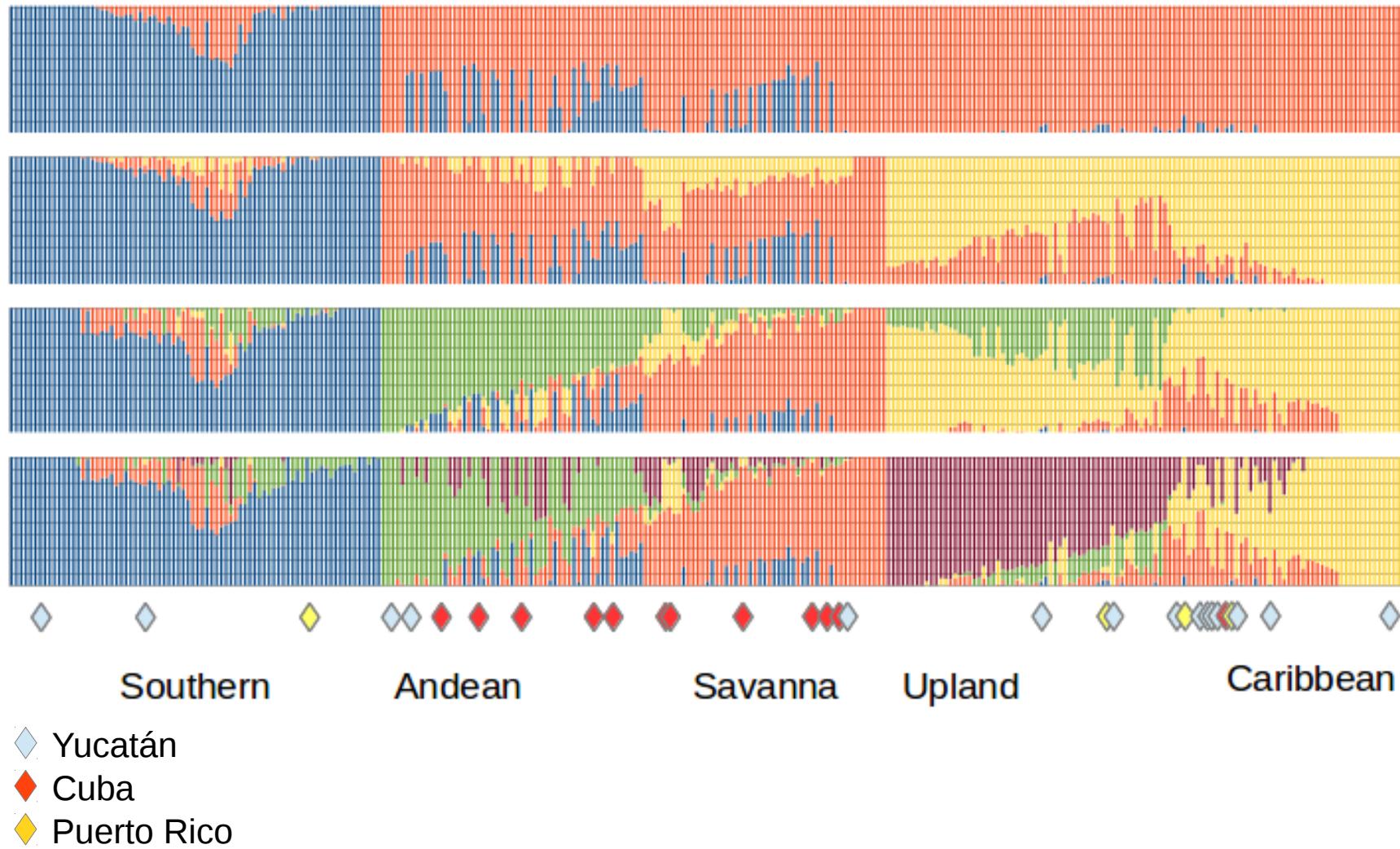
Secuenciación RAD en yuca



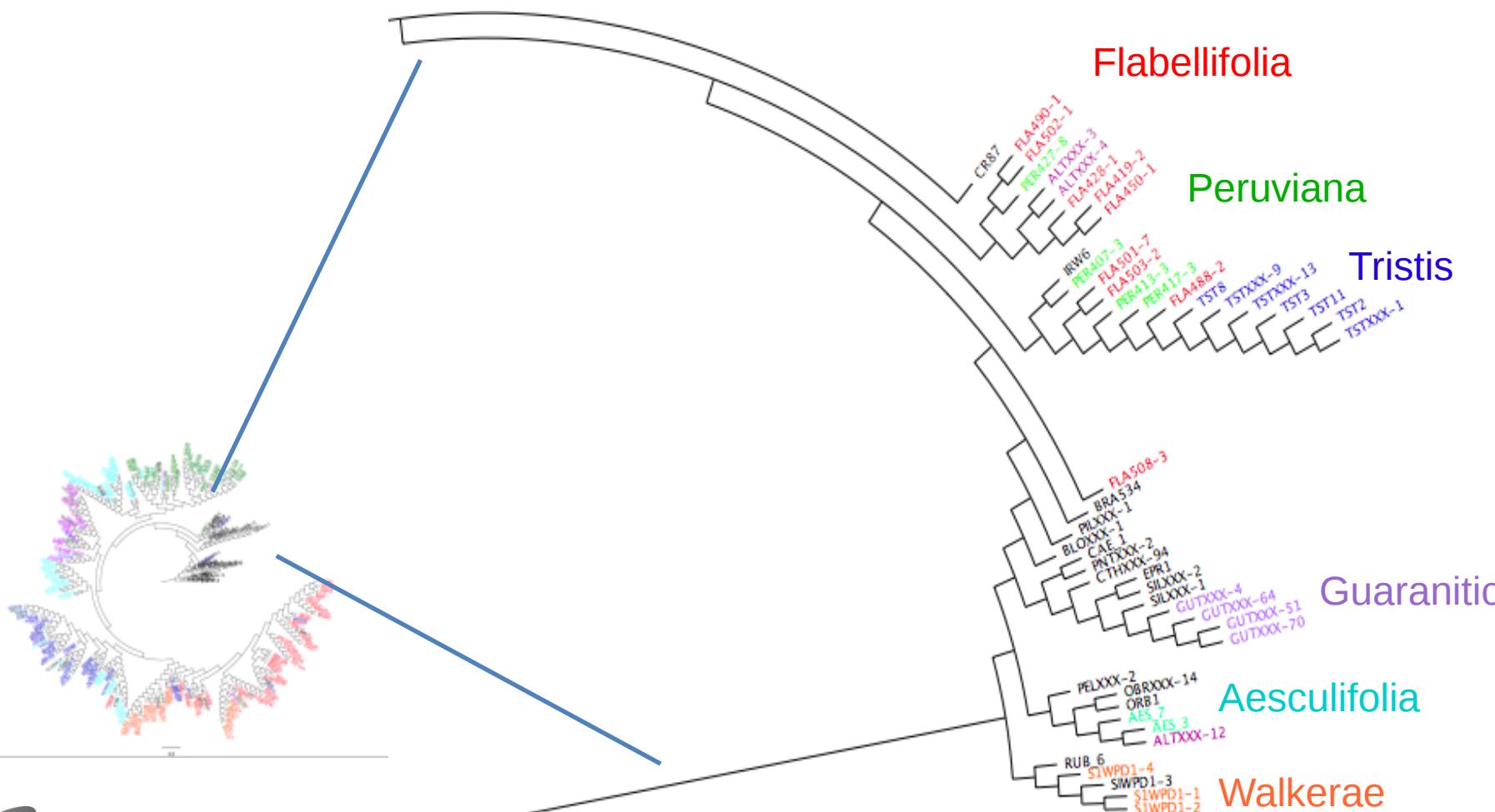
	Landraces	Wild
Sequenced Accessions	292	54
Reads per Accession	2.23 – 27M (8.57M)	3 - 17.7 (9.4M)
Base Pairs (Mbp)	191 – 2,353 (737)	255 – 1524 (810)
Average initial coverage	0.25 – 3.1x (0.97x)	0.34 – 2x (1.07x)
% Mapped Reads	75.5% - 86.3% (82.9%)	33.8% - 87.7% (71.2%)
% Mapped Reads in unique sites	26.4% - 55.6% (45.2%)	12.9% - 82.7% (49.6%)

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE. 3(10):e3376. Epub Oct 13.

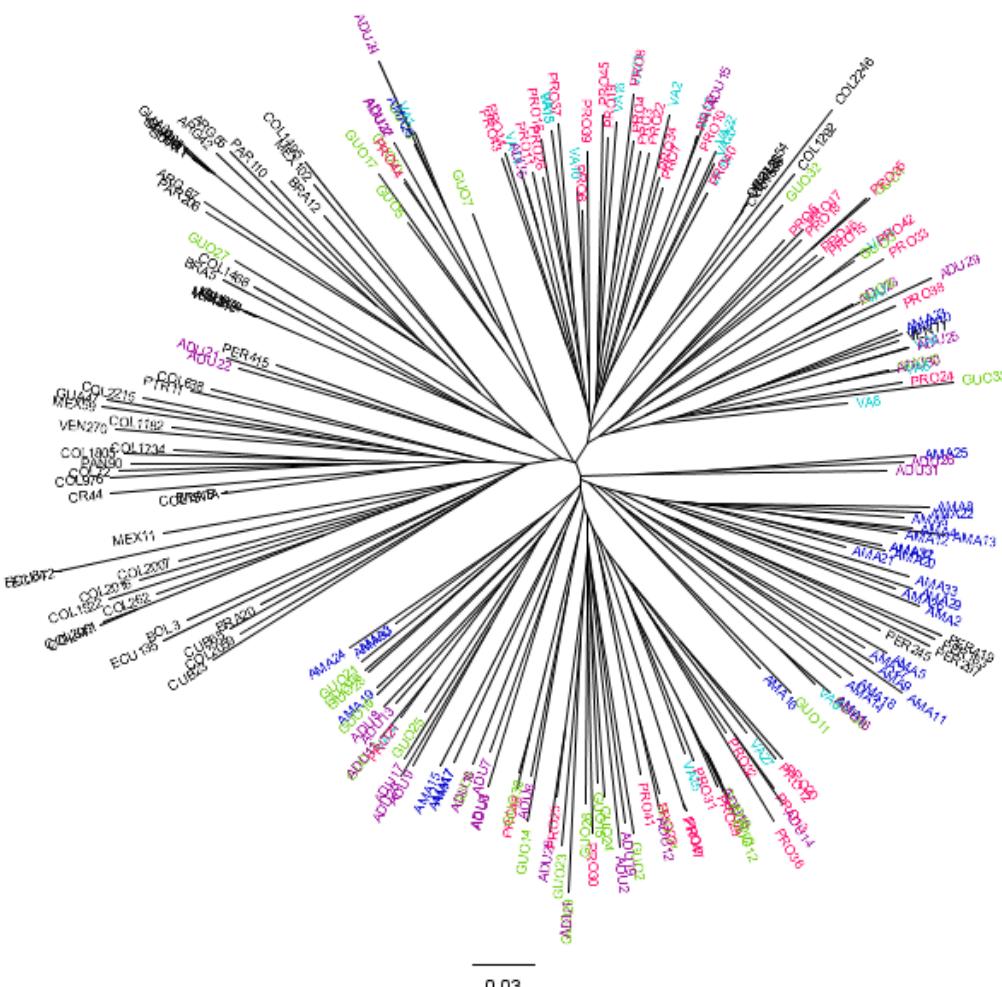
Estructura poblacional de Yuca



Variedades silvestres de Manihot



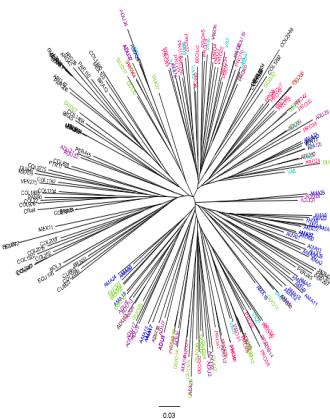
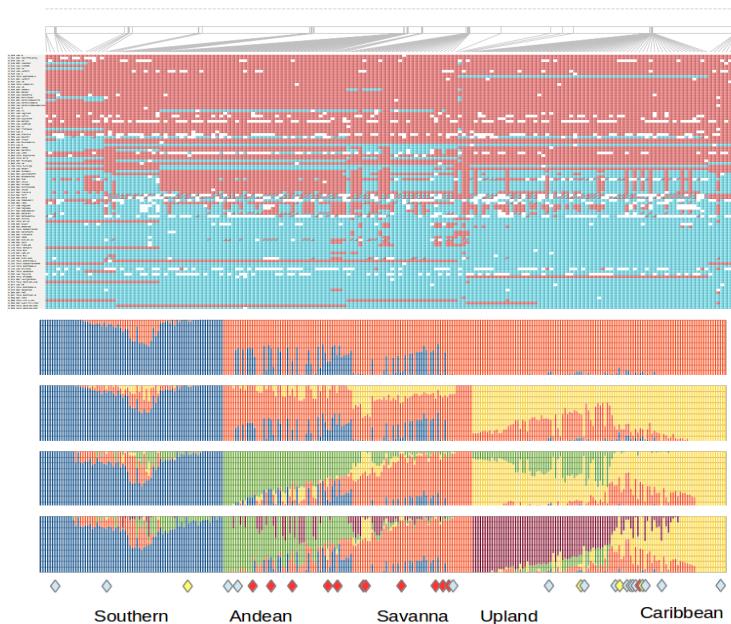
SNPs de Identificación de Muestras



Conclusiones

- La bioinformática es un componente fundamental dentro del proceso de mejoramiento de plantas utilizando técnicas moleculares
- El desarrollo de NGSEP facilitó el análisis de grandes cantidades de datos de secuenciación entregando información valiosa para:
 - Entender la estructura de nuestros genomas
 - Identificar la distribución de variabilidad de nuestro germoplasma
 - Desarrollar herramientas moleculares para apoyar el proceso de mejoramiento

El reto



<http://ricewisdom.org/>
<http://www.emergingfrontiers.com/tag/cassava/>

Agradecimientos



Estamos
buscando
practicante

- Biotecnología: Constanza Quintero, Joe Tohme
- Mejoramiento de arroz: Edgar Torres, Cesar Martinez, Mathias Lorieux
- Mejoramiento de Yuca: Luis Augusto Becerra, Tatiana Ovalle, Clair Hershey, Hernan Ceballos, Nelson Morante, Soroush Parsa
- Unidad de recursos Genéticos: Daniel Debouck, Ericsson Aranzales



Colaboradores



Patrocinio



National Science Foundation
WHERE DISCOVERIES BEGIN



Gracias

