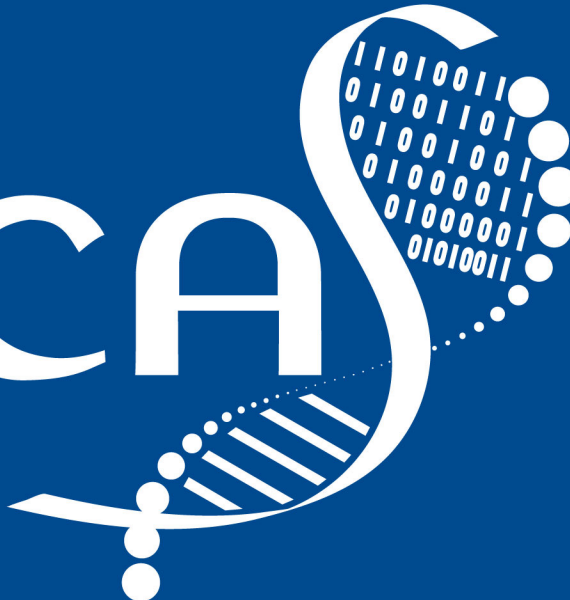


ómica

A stylized graphic on the right side of the word 'ómica'. It features a white DNA double helix structure. The top part of the helix is composed of a series of white dots, and the bottom part is composed of a series of white lines. The entire graphic is set against a dark blue background.

El futuro
es de todos

Gobierno
de Colombia



A consolidated dataset to assess gene co-expression in *Saccharum spontaneum* AP85-441

Nicolás López-Rozo, Mauricio Ramírez -Castrillón, Miguel Romero, Jorge Finke y Camilo Rocha.



PROBLEMA: MISMO CULTIVAR – 2 GENOMAS

- Genome
 - Sspon.genome.fasta
- Annotation
 - Sspon.gene.fasta
 - Sspon.cds.fasta
 - Sspon.gff3.fasta
 - Sspon.protein.fasta
 - Sspon.goslim.fasta
 - Sspon.kegg.fasta
- Molecular Maker
 - SSR
 - Primer

- Gene Expression Level(FPKM) Data
 - Leafsection
 - Circadian Rhythms
 - Growth and development

Identificadores parecidos, pero diferentes

Mejor organización de los datos en v2019 (identificadores en común para los alelos de cada gen)

Asociación de alelos y genes en la v2018 solo disponible en la Web

Expresión solo disponible en v2018

The Ming Laboratory v2019

Saccharum spontaneum AP85-441 genome

Name	Size
README-SsponAnnotation.pdf	70K
Sspon.HiC_chr_asm.fasta	2.9G
Sspon.v20190103.cds.fasta.gz	28M
Sspon.v20190103.gene.fasta.gz	88M
Sspon.v20190103.gff3.gz	13M
Sspon.v20190103.protein.fasta.gz	19M
Saccharum spont alleleTable-Jan 2019.csv	5.8M

Total visits to this site
0007737

PROBLEMA: MISMO CULTIVAR – 2 GENOMAS

v2019

Gene 1	Allele G1.1	Allele G1.2	Allele G1.3	Allele G1.4
Gene 2	Allele G2.1	Allele G2.2	Allele G2.3	Allele G2.4
Gene 3	Allele G3.1	Allele G3.2		Allele G3.4
Gene 4	Allele G4.1	Allele G4.2		

v2018

Gene x	Allele 1	Allele 2	Allele 3	Allele 4	Allele 17
Gene xx	Allele 5	Allele 6	Allele 7	Allele 8	
Gene xxx	Allele 9	Allele 10		Allele 11	
Gene xxxx	Allele 12	Allele 13	Allele 14	Allele 15	Allele 16

PROBLEMA: GENERAR REDES COEXPRESIÓN ALELOS – MUCHO ESFUERZO COMPUTACIONAL

1. CONCATENAR DATOS EXPRESIÓN

112788 filas → 112787 alelos
97 columnas → 96 experimentos

~37% de la matriz son ceros

53623 líneas (47,5%) contienen ceros.

2. TENEMOS ALELOS → ES SUFICIENTE CON GENES

Datos RNA-seq permiten obtener información de expresión específica de alelos.

Anotaciones de ontología son redundantes para cada alelo, en su mayoría.

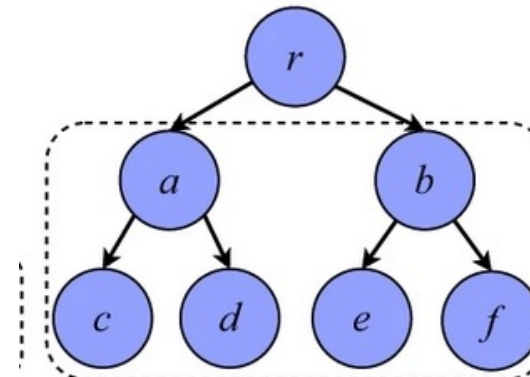
% ontologías anotadas es bajo

3. REDES CO-EXPRESIÓN

Esfuerzo computacional
Consumo tiempo excesivo
(varias semanas)

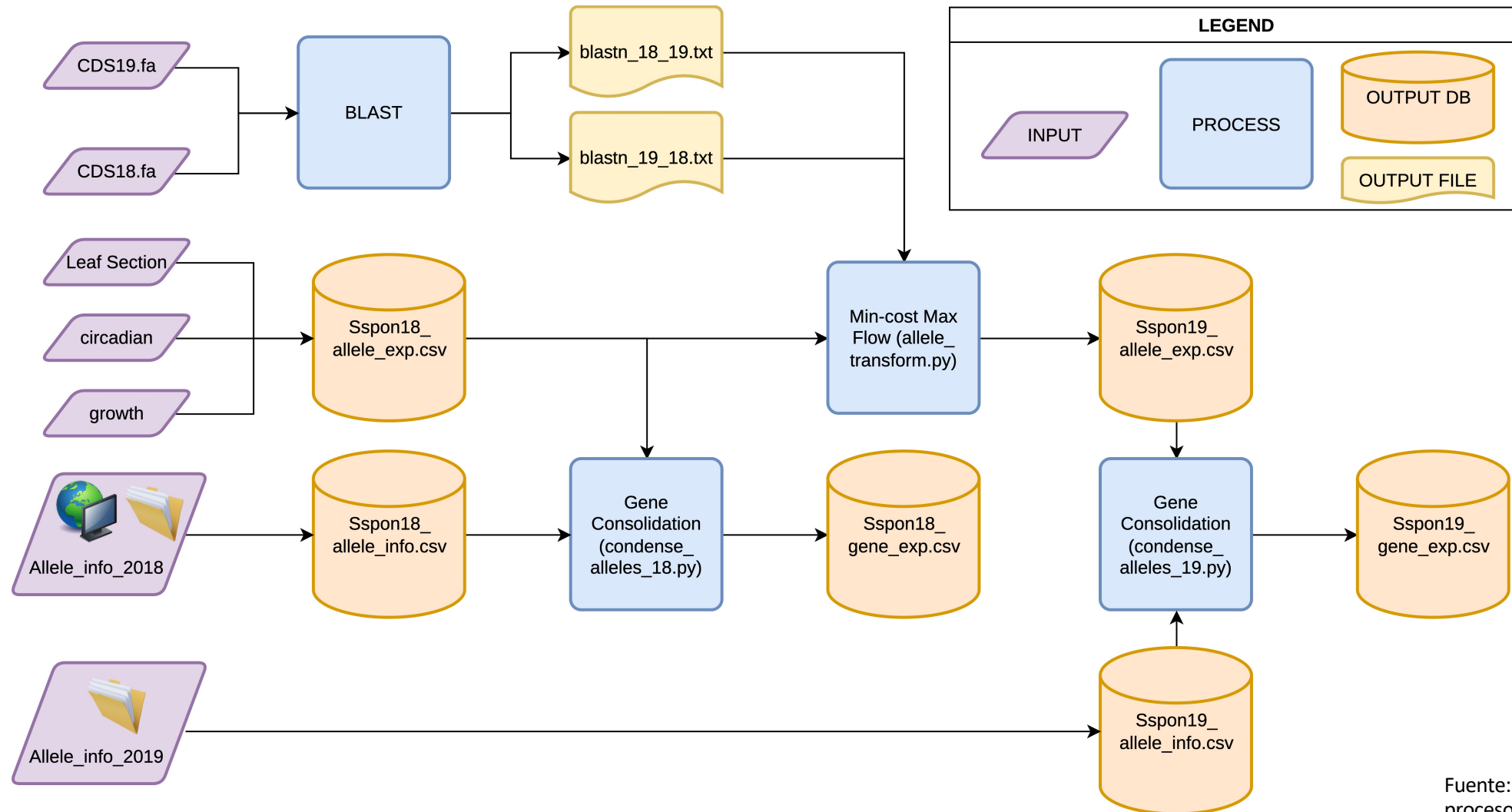


4. DIFÍCIL PREDECIR NUEVAS ANOTACIONES



Romero et al. 2022

CONTRIBUCIÓN: UN ALGORITMO PARA GENERAR MATRICES DE EXPRESIÓN A NIVEL DE GENES



Fuente: propia. Manuscrito en proceso de revisión

RESULTADOS: BLASTn

v2019

Gene 1	Allele G1.1	Allele G1.2	Allele G1.3	Allele G1.4
Gene 2	Allele G2.1	Allele G2.2	Allele G2.3	Allele G2.4
Gene 3	Allele G3.1	Allele G3.2		Allele G3.4
Gene 4	Allele G4.1	Allele G4.2		

Blastn 2019 --> 2018

1	Sspon.03G0012650-1P	Sspon.001A0000010	100.000	2232
2	Sspon.03G0012650-1P	Sspon.ctg0272480	99.507	2232
3	Sspon.03G0012650-1P	Sspon.003A0015630	98.746	2232
4	Sspon.01G0000010-1A	Sspon.001A0000020	100.000	420
5	Sspon.01G0000010-1A	Sspon.001D00000770	90.271	442
6	Sspon.01G0000020-1A	Sspon.001A0000030	100.000	549
7	Sspon.01G0000030-1A	Sspon.001A0000040	100.000	948
8	Sspon.01G0000030-1A	Sspon.001B0035610	97.904	954
9	Sspon.01G0000030-1A	Sspon.001A0000280	95.487	975
10	Sspon.01G0000030-1A	Sspon.001D00000560	94.785	978

182839 alineamientos

v2018

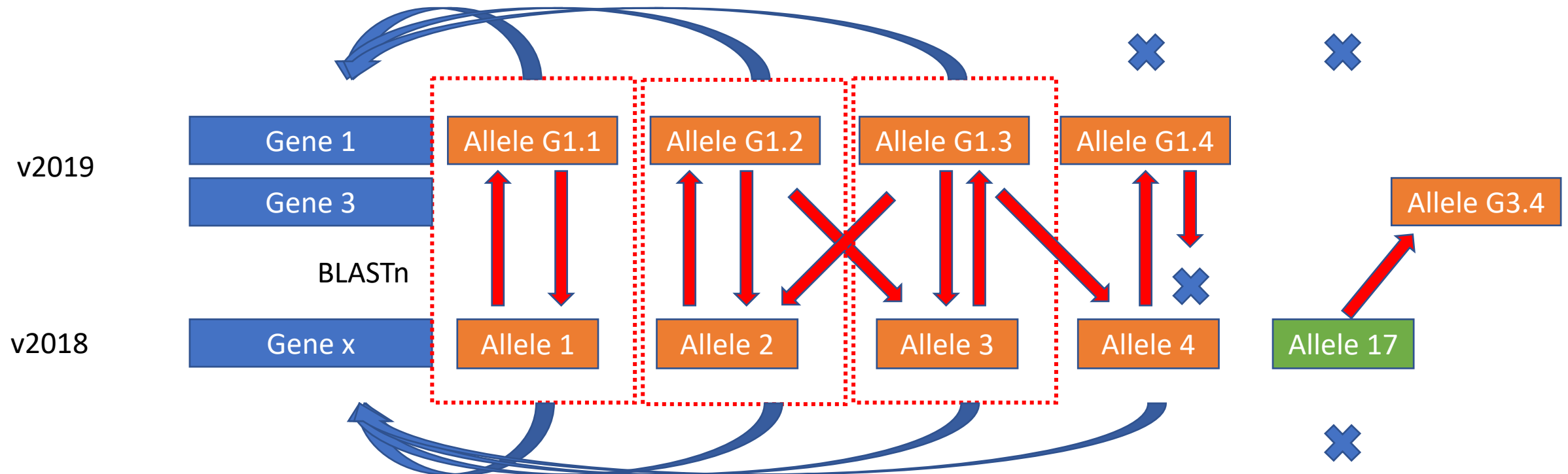
Gene x	Allele 1	Allele 2	Allele 3	Allele 4	Allele 17
Gene xx	Allele 5	Allele 6	Allele 7	Allele 8	
Gene xxx	Allele 9	Allele 10		Allele 11	
Gene xxxx	Allele 12	Allele 13	Allele 14	Allele 15	Allele 16

Blastn 2018 --> 2019

1	Sspon.001A0000010	Sspon.03G0012650-1P	100.000	2232
2	Sspon.001A0000010	Sspon.03G0012650-1A	98.746	2232
3	Sspon.001A0000020	Sspon.01G0000010-1A	100.000	420
4	Sspon.001A0000020	Sspon.01G0000010-3D	90.271	442
5	Sspon.001A0000030	Sspon.01G0000020-1A	100.000	549
6	Sspon.001A0000040	Sspon.01G0000030-1A	100.000	948
7	Sspon.001A0000040	Sspon.01G0000030-2B	97.904	954
8	Sspon.001A0000040	Sspon.01G0000030-4D	94.785	978
9	Sspon.001A0000052	Sspon.01G0000040-1A	100.000	1440
10	Sspon.001A0000052	Sspon.01G0000040-2B	99.512	1433

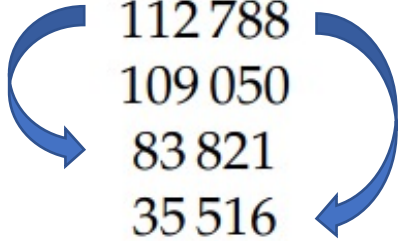
183401 alineamientos

APLICANDO EL ALGORITMO



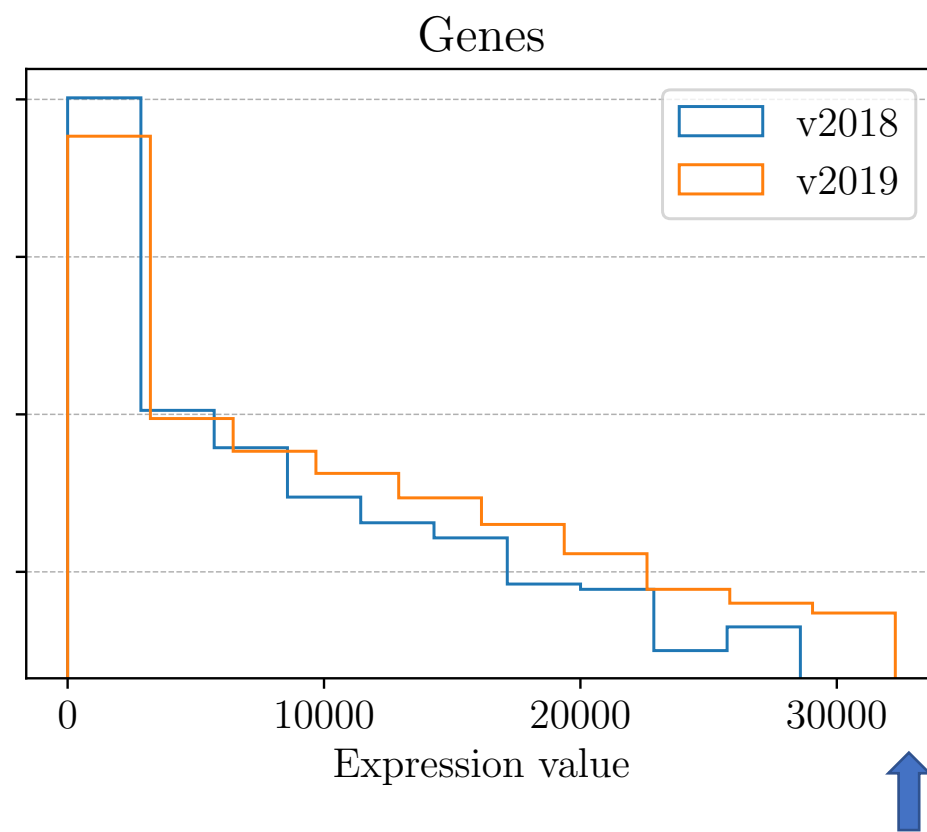
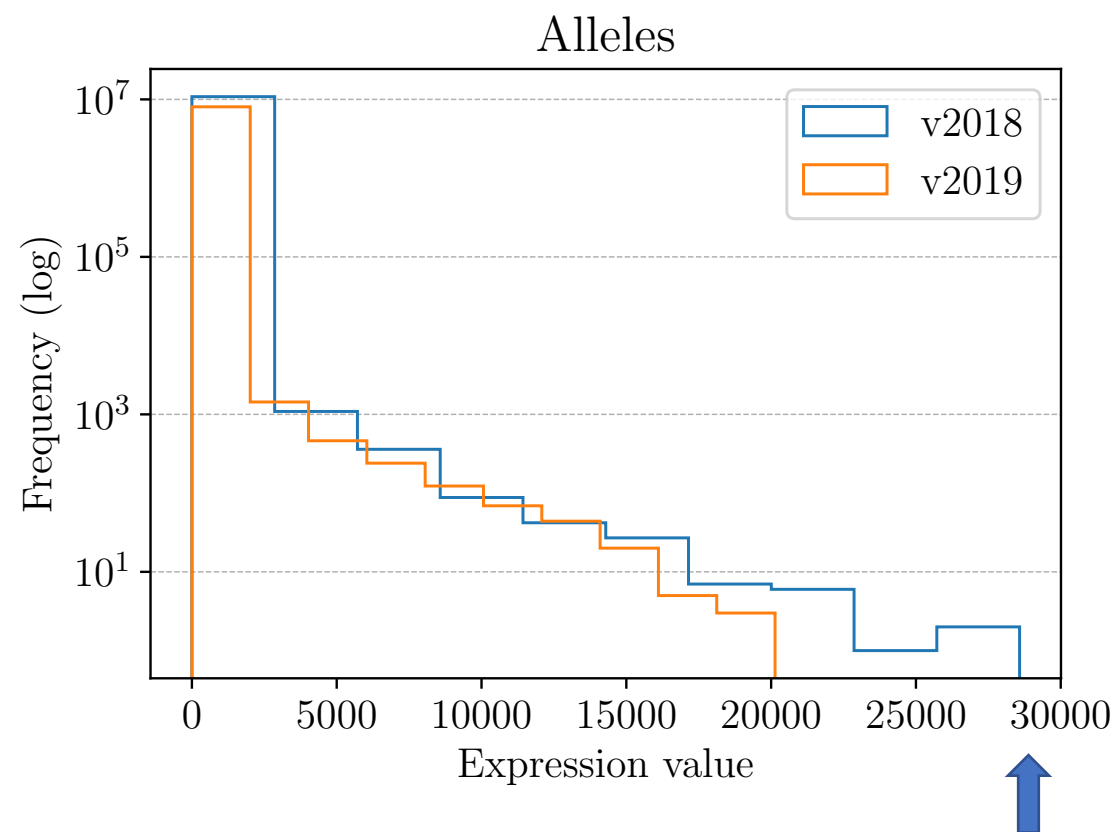
RESULTADOS: DOS MATRICES CONSOLIDADAS v2019

Table 1. Composition of expression matrices for the two *Saccharum spontaneum* AP85-441 genomes in v2018 and v2019. The matrices comprise 96 columns of experimental data.

Matrix name	Number of alleles/genes	Reference
Alleles v2018	 112 788 109 050 83 821 35 516	[3]
Genes v2018		This data descriptor
Alleles v2019		This data descriptor
Genes v2019		This data descriptor

Fuente: propia. Manuscrito en proceso de revisión

LOS VALORES TOTALES DE EXPRESIÓN SON MÁS ALTOS EN LOS GENES v2019



Fuente: propia. Manuscrito en proceso de revisión

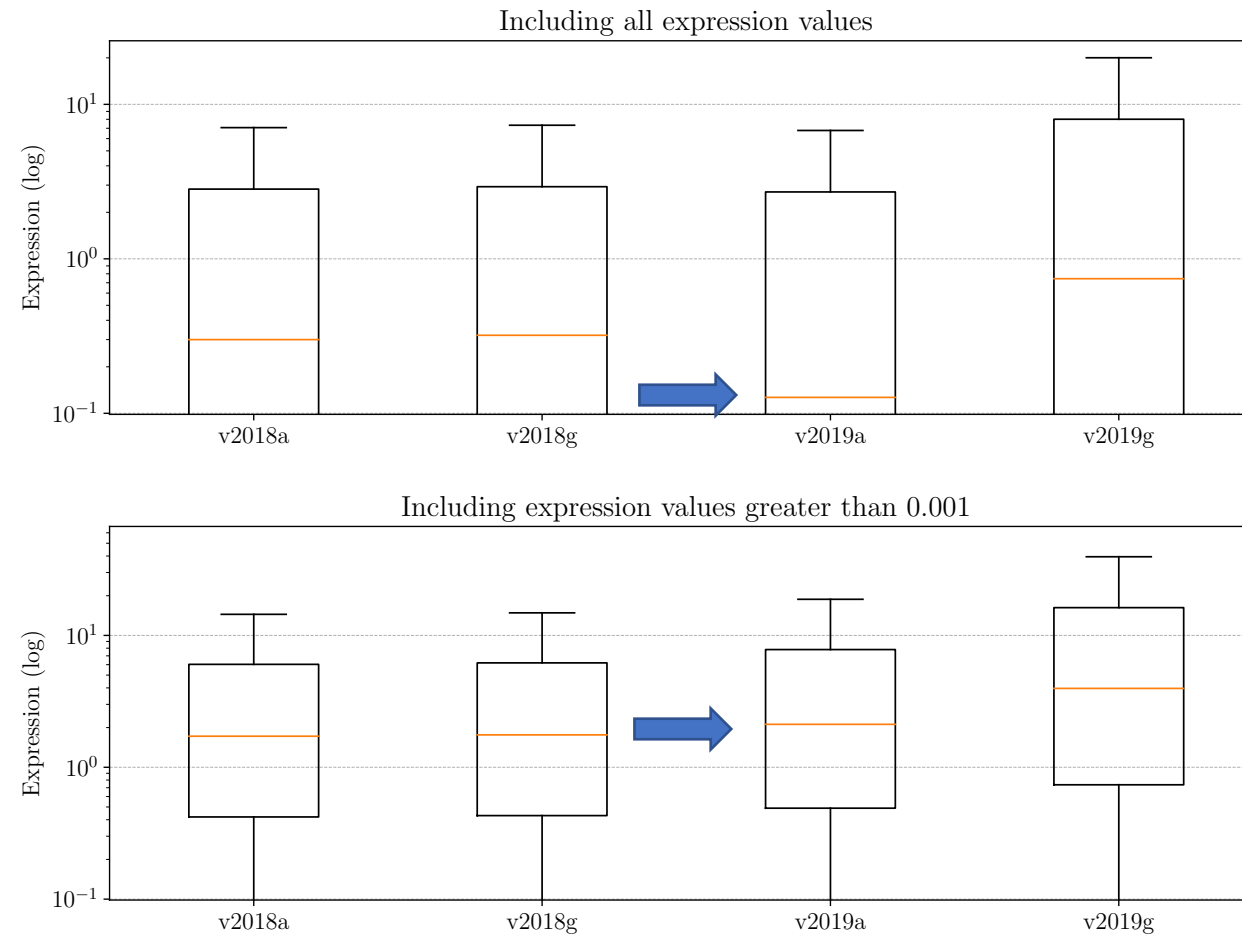
LA CONSOLIDACIÓN DE DATOS FUNCIONA CORRECTAMENTE PARA LA v2019

Table 2. Description of expression values for the two *Saccharum spontaneum* AP85-441 genomes.

Matrix name	Max value	All values		Values > 0,001	
		Count	Median	Count	Median
Alleles v2018	28 579	10 827 650	0,300	6 809 283	1,720
Genes v2018	28 579	10 468 800	0,320	6 621 587	1,761
Alleles v2019	20 135	8 046 816	0,127	4 460 747	2,117
Genes v2019	32 282	3 409 536	0,744	2 277 266	3,963

~33% = 0.0

LA CANTIDAD DE CEROS AFECTA LAS ESTADÍSTICAS DE ALELOS v2019



Fuente: propia. Manuscrito en proceso de revisión

REPOSITORIO – Github.com

 **mauriciogeteg** / **sugarcane-gene-expression** Public

  1  0  Star

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

 main  1 branch  0 tags

[Go to file](#) [Add file](#) [Code](#)

 **mauriciogeteg** Delete figure1.eps 60f9344 6 hours ago 18 commits

	Codes	Update Readme.md	2 months ago
	blast_results	Files were added	2 months ago
	figures	Delete figure1.eps	6 hours ago
	inputs	Input files added	2 months ago
	LICENSE	Update LICENSE	2 months ago
	README.md	Update README.md	2 months ago
	Sspon18_allele_.7z	Files were added	2 months ago
	Sspon18_gene_.7z	Files were added	2 months ago
	Sspon19_.7z	Files were added	2 months ago

README.md 

Sugarcane dataset for gene-expression analysis

About

Dataset to assess gene expression for sugarcane

 Readme

 CC-BY-4.0 license

 0 stars

 1 watching

 0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors

2

APLICANDO EL DATASET PARA CONSTRUIR REDES DE COEXPRESIÓN

Gene	Expression Values					
G ₁	■	■	■	■	■	■
G ₂	■	■	■	■	■	■
G ₃	■	■	■	■	■	■
G ₄	■	■	■	■	■	■
G ₅	■	■	■	■	■	■
G ₆	■	■	■	■	■	■

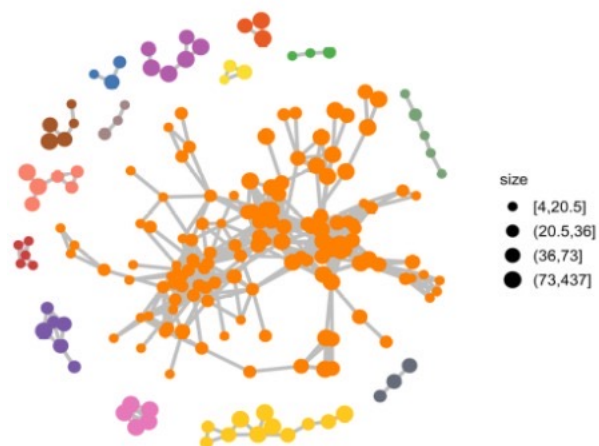
Metric
Computation

Metric	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆
G ₁	■	■	■	■	■	■
G ₂	■	■	■	■	■	■
G ₃	■	■	■	■	■	■
G ₄	■	■	■	■	■	■
G ₅	■	■	■	■	■	■
G ₆	■	■	■	■	■	■

Threshold
Procedure

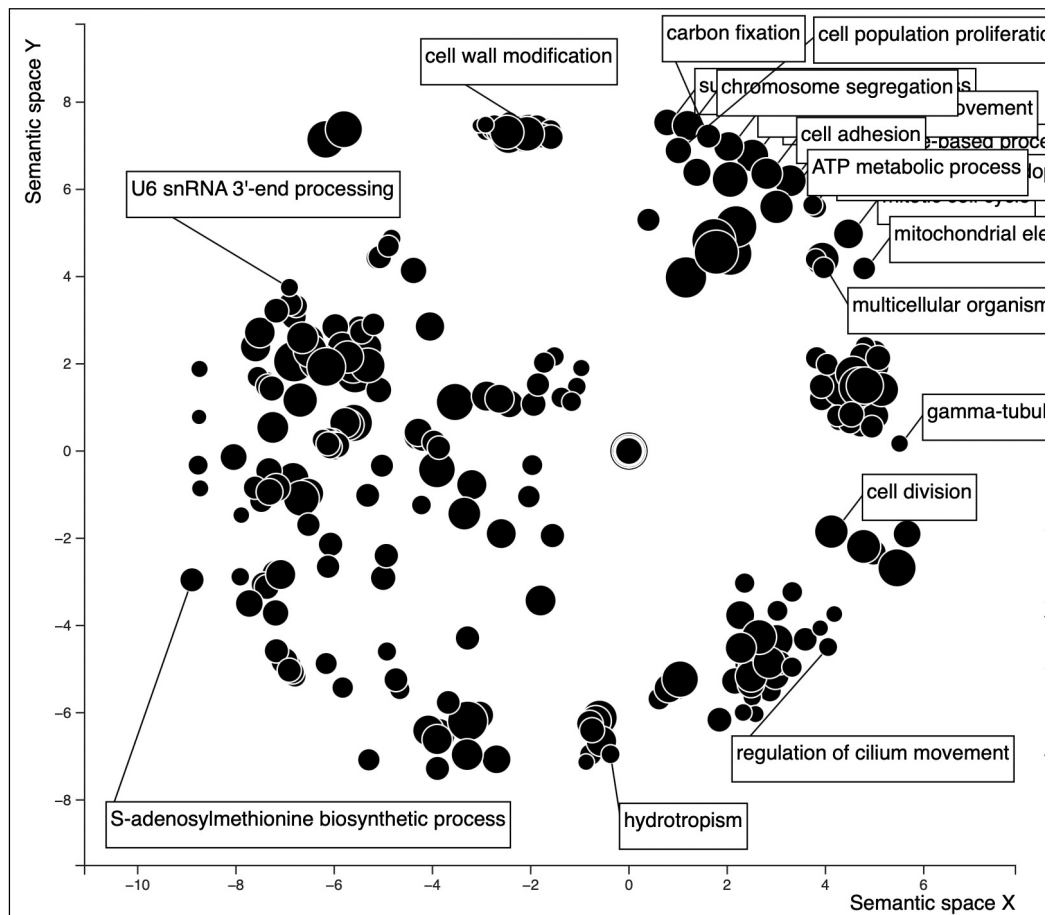
Threshold	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆
G ₁	■	■	■	■	■	■
G ₂	■	■	■	■	■	■
G ₃	■	■	■	■	■	■
G ₄	■	■	■	■	■	■
G ₅	■	■	■	■	■	■
G ₆	■	■	■	■	■	■

Gene
Enrichment

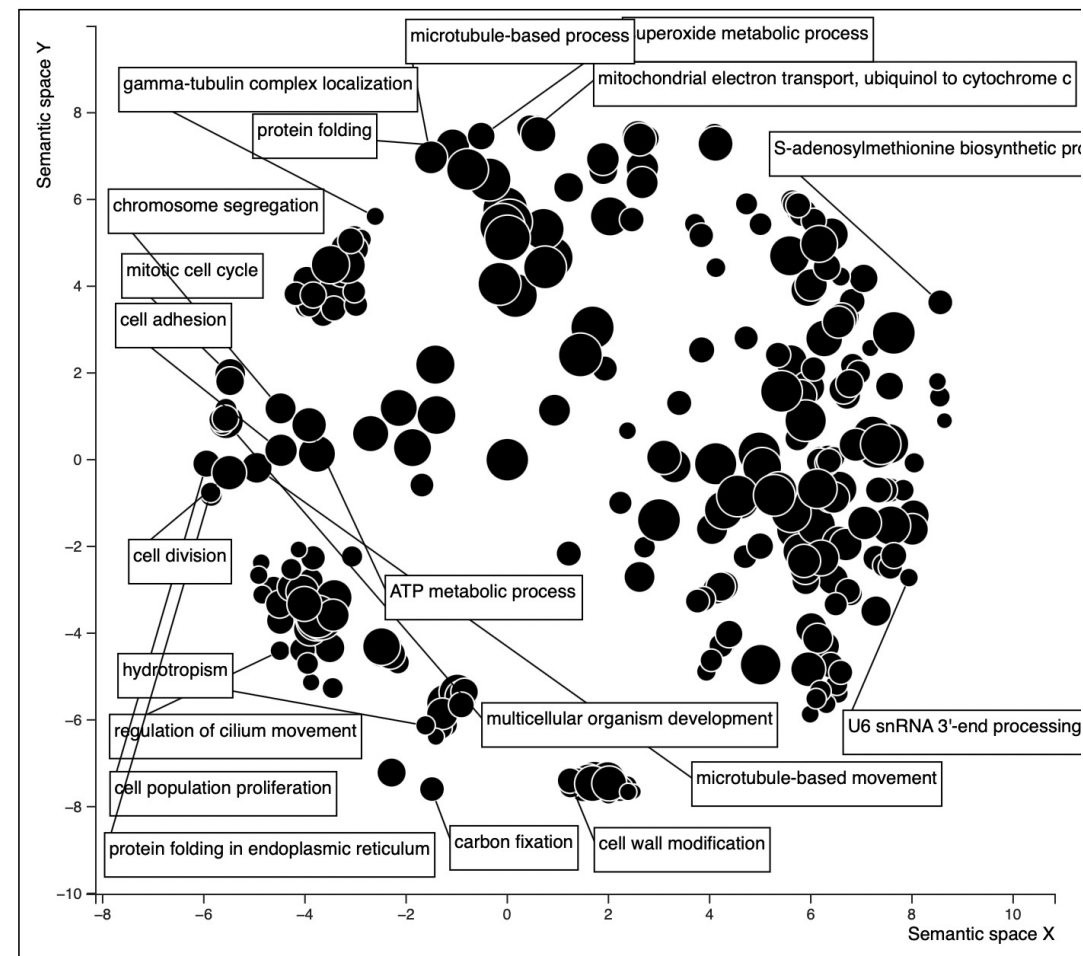


¿QUÉ PODEMOS HACER CON ESTOS DATOS? PREDICCIÓN DE NUEVAS ANOTACIONES FUNCIONALES

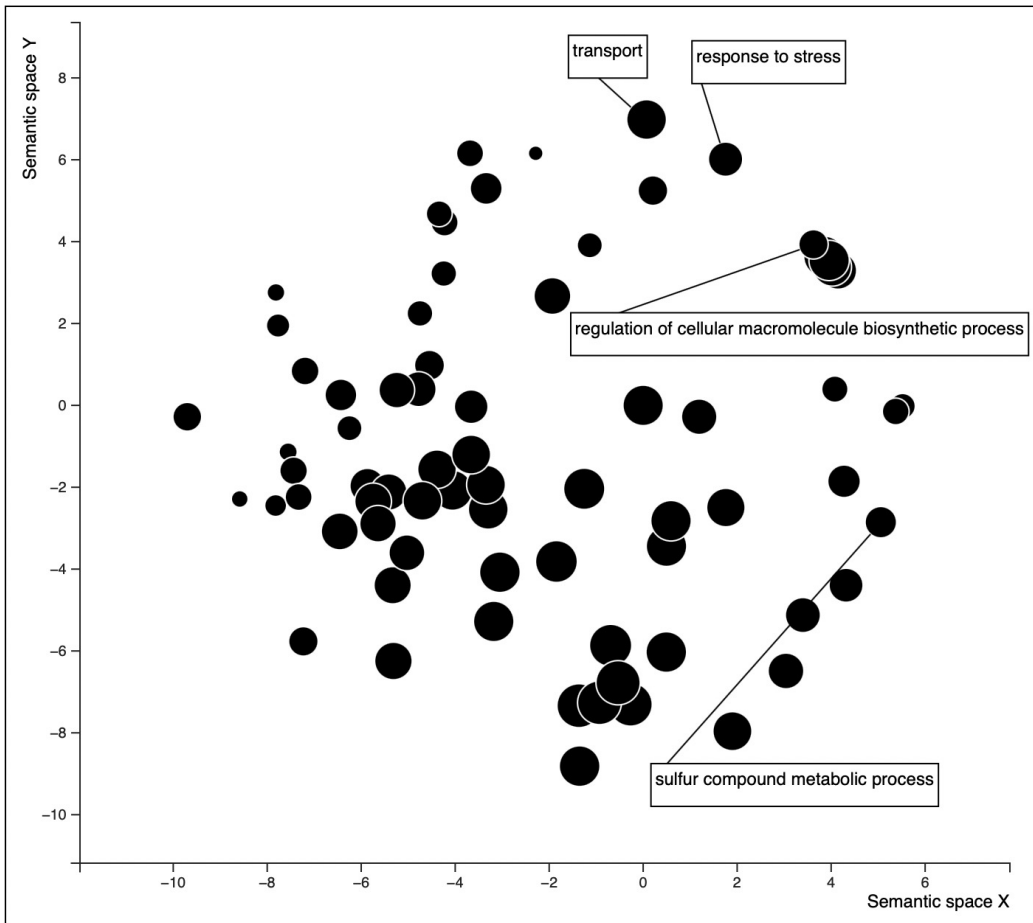
ANTES: Gene Ontology – Procesos biológicos: 449



DESPUÉS (umbral >0,9): Gene Ontology PB: 515 (En proceso)



¿QUÉ PODEMOS HACER CON ESTOS DATOS? PREDICCIÓN DE NUEVAS ANOTACIONES FUNCIONALES



Ontologías con mayor unicidad

	GO:0006950	response to stress
—	GO:0006810	transport
	GO:0051234	establishment of localization
	GO:0055085	transmembrane transport
	GO:0044281	small molecule metabolic process
	GO:0006790	sulfur compound metabolic process
	GO:1901135	carbohydrate derivative metabolic process
	GO:0009058	biosynthetic process
	GO:0006091	generation of precursor metabolites and energy
	GO:0006629	lipid metabolic process
	GO:0005975	carbohydrate metabolic process
	GO:0006643	membrane lipid metabolic process
—	GO:0006013	mannose metabolic process
	GO:0006006	glucose metabolic process
	GO:2000112	regulation of cellular macromolecule biosynthetic process
—	GO:0046486	glycerolipid metabolic process
	GO:0006644	phospholipid metabolic process
	GO:0019222	regulation of metabolic process
	GO:0044262	cellular carbohydrate metabolic process
	GO:0006807	nitrogen compound metabolic process
	GO:0050789	regulation of biological process
	GO:0044238	primary metabolic process
	GO:1901361	organic cyclic compound catabolic process
	GO:0044237	cellular metabolic process
	GO:0050794	regulation of cellular process
	GO:1901360	organic cyclic compound metabolic process
	GO:0071704	organic substance metabolic process
	GO:0006793	phosphorus metabolic process
	GO:0034968	histone lysine methylation
	GO:0043170	macromolecule metabolic process
	GO:0006650	glycerophospholipid metabolic process

CONCLUSIÓN

El algoritmo permite la consolidación de datos de expresión desde nivel de alelos a genes.

La reducción significativa de filas permite la generación de redes de co-expresión de forma más rápida y con menor esfuerzo computacional, sin la pérdida significativa de datos.

Las matrices compartidas, permite el análisis directo de datos de expresión en diferentes versiones de un mismo genoma, a nivel de alelos o de genes.

El código, así como los conjuntos de datos se encuentran disponibles para todo el público en www.github.com/mauriciogeteg/sugarcane-gene-expression





El futuro
es de todos

Gobierno
de Colombia



ómica



Aliados

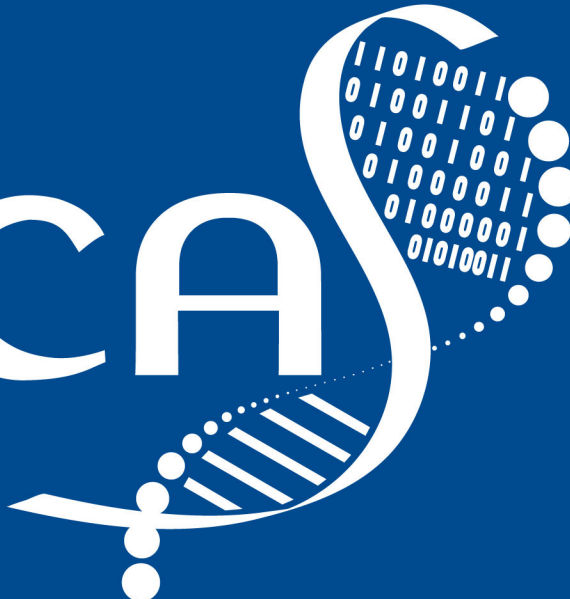


Apoyan



MUCHAS GRACIAS POR
SU ATENCIÓN

ómica

A stylized graphic element on the right side of the word 'ómica'. It features a white DNA double helix structure. The top strand of the helix is composed of a series of white dots, each followed by a short sequence of binary code (0s and 1s). The bottom strand is a solid white line. The helix curves upwards and to the right.

El futuro
es de todos

Gobierno
de Colombia

