

**“De la secuenciación
global a la solución local:
innovando en el
epigenotipado de cultivos”.**

Carla Valeria Filippi

Laboratorio de Bioquímica

FAGRO, Udelar

“Epigenética del estrés ambiental”

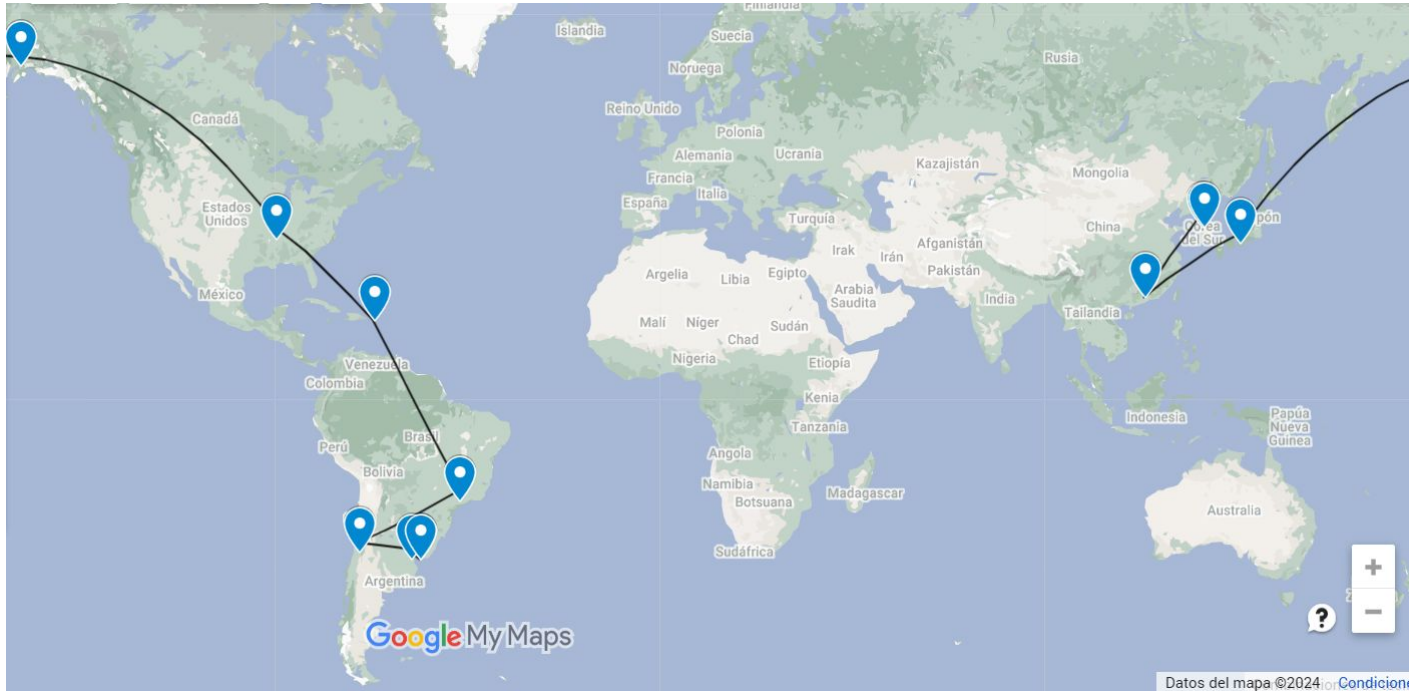


UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY





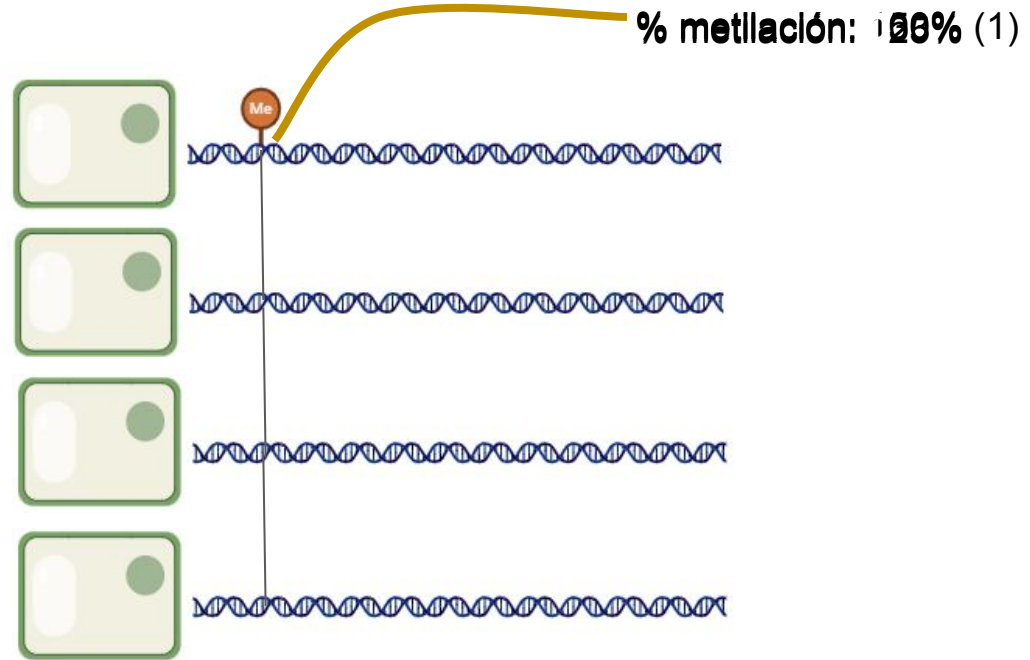
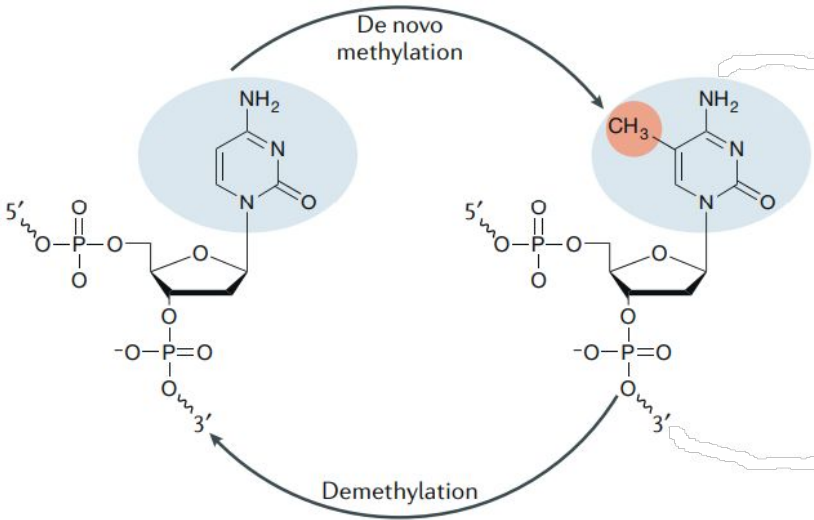
Ómicas y sequencing facilities



Qué podemos hacer para “apropiarnos” de estas tecnologías, a nivel local?



Metilación del ADN (5mC)

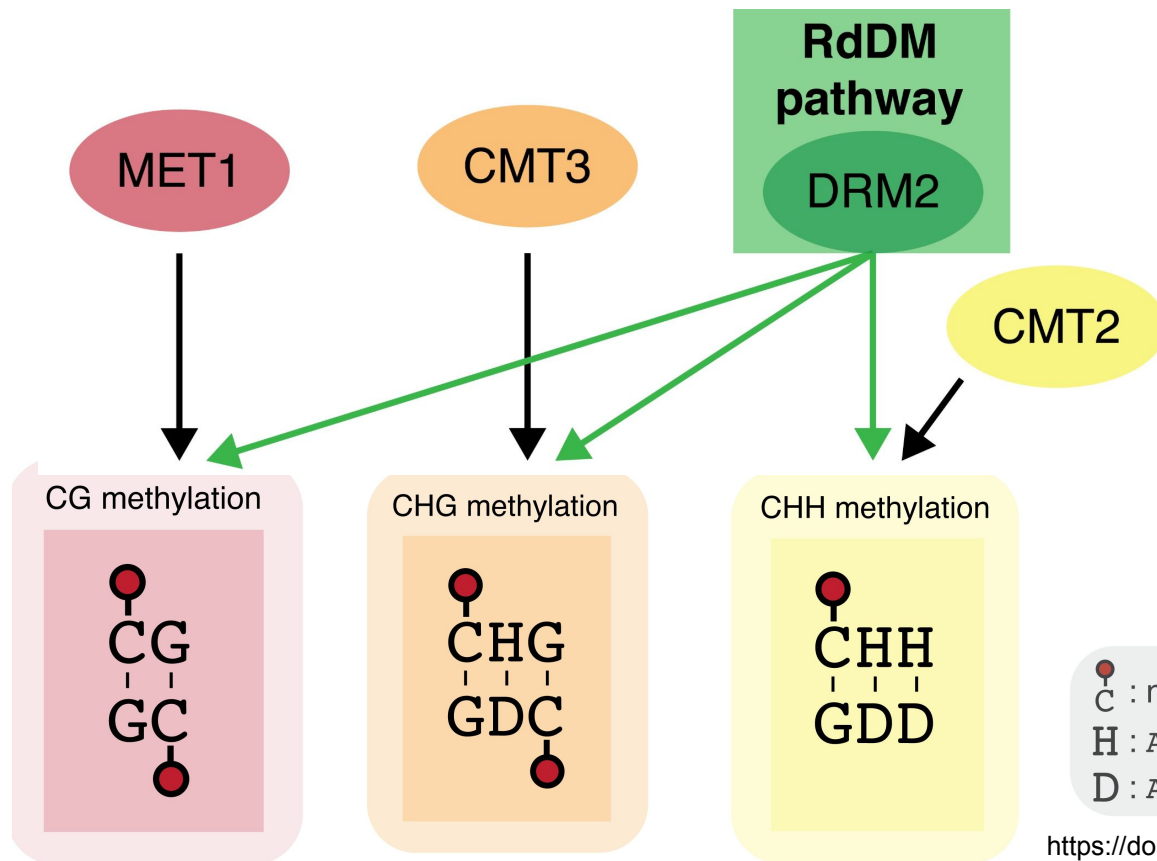


Estudios de metilación a nivel genoma (metiloma) requieren **réplicas (3)** y buena **profundidad** de secuenciación (**~25-30X**)



<https://doi.org/10.1038/s41580-018-0016-z>

Metilación del ADN (5mC) en plantas



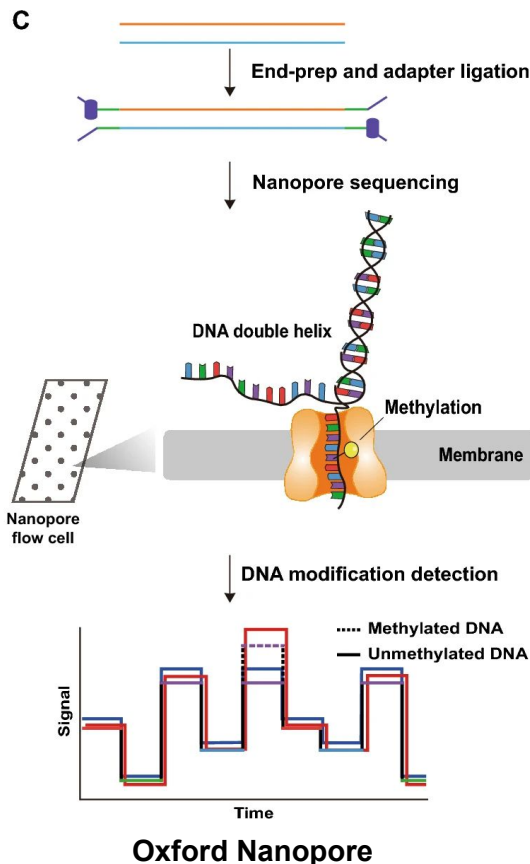
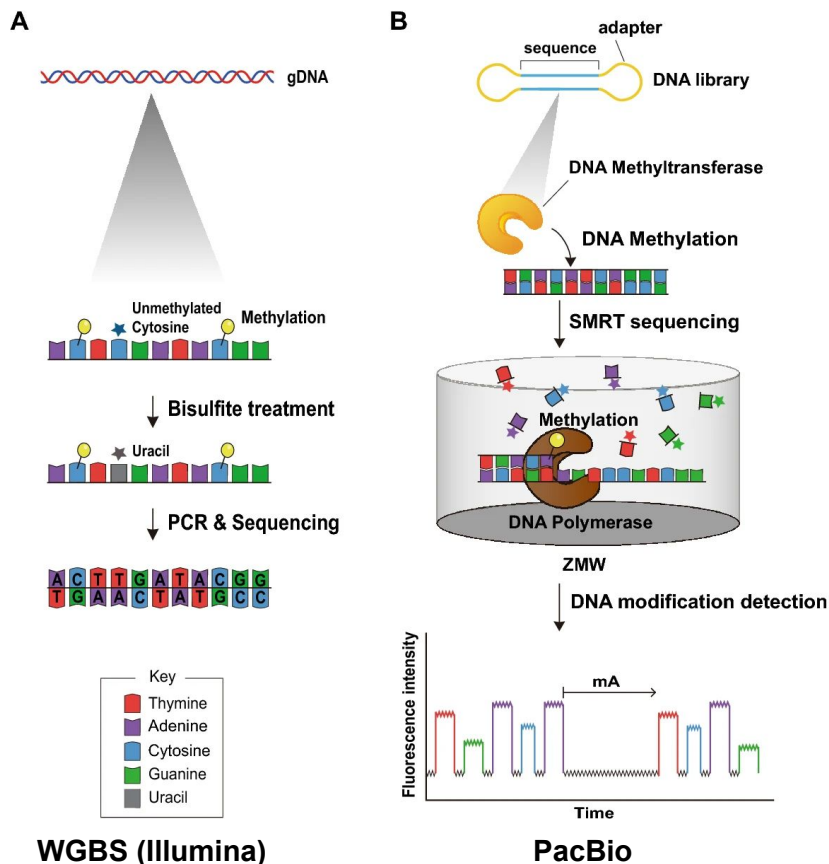
Metilación en contexto asimétrico requiere mayor profundidad de secuenciación.

C^{\bullet} : methylated cytosine (C)
H : A, T or C (not G)
D : A, T or G (not C)





Estrategias para la identificación de 5mC (*Genome wide*)



Molécula de ADN nativo (**no tratado**) en PacBio/ Nanopore

Análisis de metilación en contextos CHG y CHH **no optimizados** en Pacbio/ Nanopore

Costos: Para 3 réplicas cobertura 30X, soja (genoma 1 Gb):
~USD1200 WGBS,
~USD3200 PacBIO,
~USD3300 Oxford Nanopore

<https://doi.org/10.1186/s13068-024-02529-x>





Estrategias para la identificación de 5mC (RRBS)

Variando la(s) enzima(s) y/o el rango de size selection, cambian las regiones muestreadas del genoma.

→ Especie específico

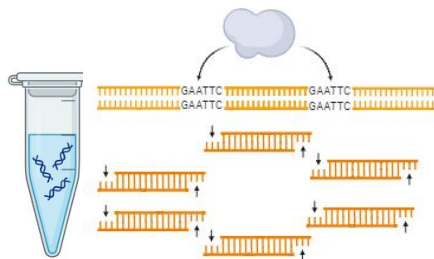
diagenode
Innovating Epigenetics Solutions

IGATech

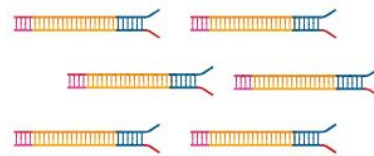
BsaWI
(W|CCGGW),
ApeKI (G|CWGC)
DpnII (^GATC).

TaqI-v2
(T|CGA)

Protocolos estandarizados, puestos a punto para unas pocas especies y aplicadas a otras por defecto → sub-óptimo



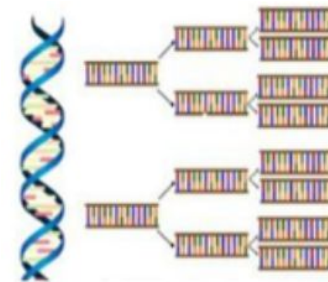
Digestión con enzima(s) de restricción



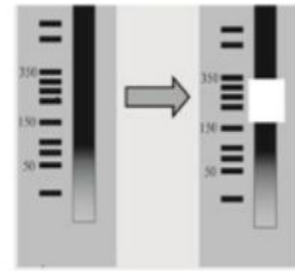
Ligación de adaptadores



**Tratamiento con bisulfito
(5mC → C y C → U)**



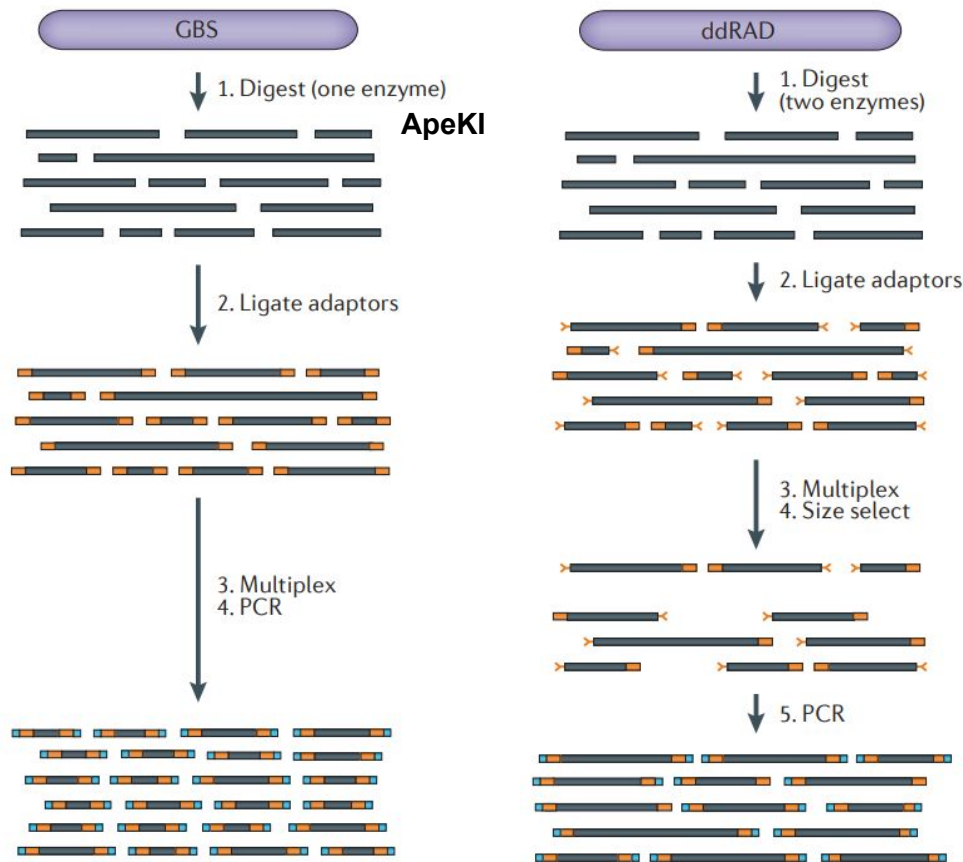
PCR (pocos ciclos)



Size selection (fragmentos con un rango de tamaño)



Optimización y puesta a punto de epi-ddRADseq



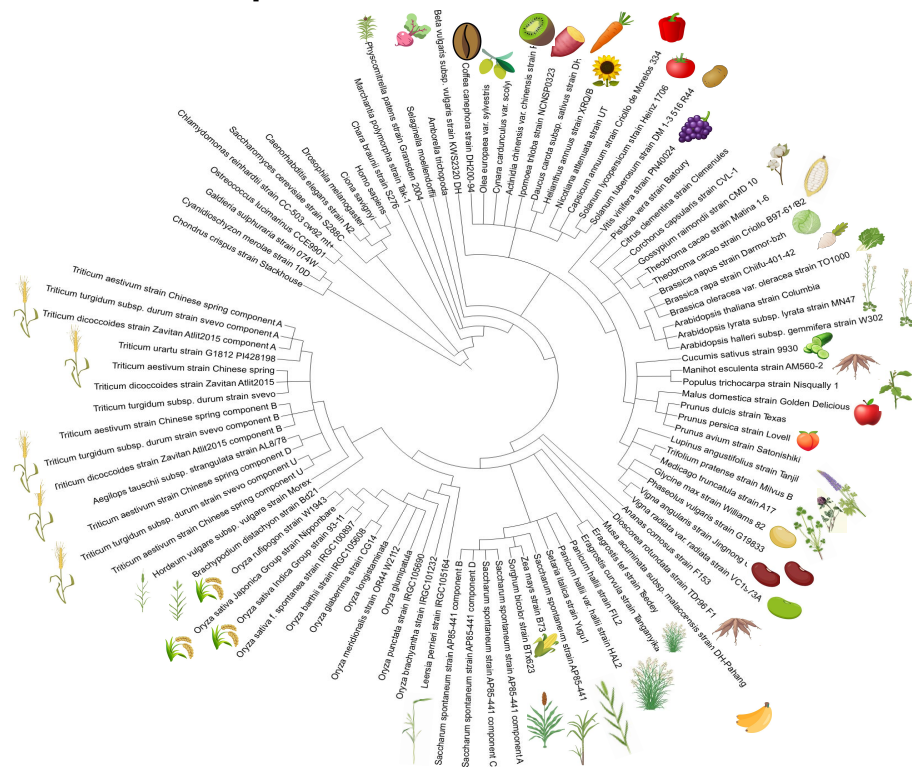
- Digerir con dos enzimas (A y B, en lugar de una sola como en GBS);
- Seleccionar, usando adaptadores, solo fragmentos de digestión doble (AB o BA)
- Size selection preciso en gel (y no con Ampure beads como en GBS)

→ Reduce duplicados de las regiones a secuenciar, genera profundidad de lectura similar entre individuos en cada región del genoma y disminuye el % de datos perdidos.





Parte I: predicciones *in silico*



75 especies

- Genomas variables (120 Mb (*A. thaliana*) - 14,5 Gb (*T. aestivum*))
- Mono/ dicotiledóneas
- Di/ poliploides
- Cultivos, especies de interés evolutivo, etc

Poaceae	24	Asparagaceae	1	Juglandaceae	1
Fabaceae	9	Amaranthaceae	1	Euphorbiaceae	1
Brassicaceae	7	Cannabaceae	1	Nymphaeaceae	1
Solanaceae	4	Rubiaceae	1	Oleaceae	1
Rosaceae	3	Betulaceae	1	Papaveraceae	1
Cucurbitaceae	2	Myrtaceae	1	Funariaceae	1
Malvaceae	2	Apiaceae	1	Salicaceae	1
Asteraceae	2	Dioscoreaceae	1	Fagaceae	1
Actinidiaceae	1	Moraceae	1	Pedaliaceae	1
Amborellaceae	1	Convolvulaceae	1	Vitaceae	1



Enzimas testeadas

AseI

+

NsiI

5'... AT▼TA AT ... 3'
3'... TA AT▲TA ... 5'

5'... ATGCA▼T ... 3'
3'... T▲ACGTA ... 5'

Csp6I

+

NsiI

5'... G▼TAC ... 3'
3'... CATG▲ ... 5'

5'... ATGCA▼T ... 3'
3'... T▲ACGTA ... 5'

MspI

+

ApeKI

5'... C▼CGG ... 3'
3'... GGCC▲ ... 5'

5'... G▼CWGC ... 3'
3'... CGWC▲G ... 5'

MspI

+

DpnII

5'... C▼CGG ... 3'
3'... GGCC▲ ... 5'

5'... G▼ATC ... 3'
3'... CTAG▲ ... 5'

PstI

+

MspI

5'... CTGCA▼G ... 3'
3'... G▲ACGTC ... 5'

5'... C▼CGG ... 3'
3'... GGCC▲ ... 5'

PstI

+

DpnII

5'... CTGCA▼G ... 3'
3'... G▲ACGTC ... 5'

5'... G▼ATC ... 3'
3'... CTAG▲ ... 5'

PstI

+

MspI

5'... GCATG▼C ... 3'
3'... CGTACG ... 5'

5'... C▼CGG ... 3'
3'... GGCC▲ ... 5'

SphI

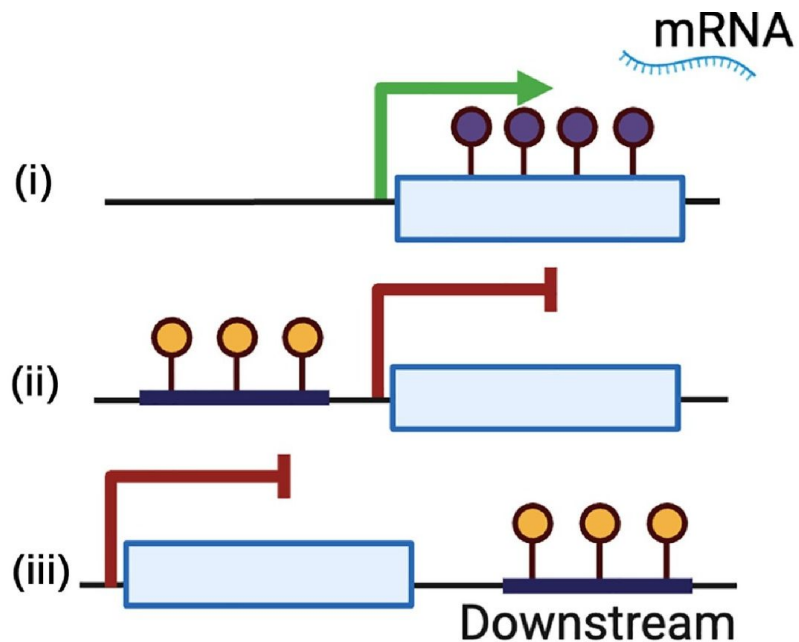
+

DpnII

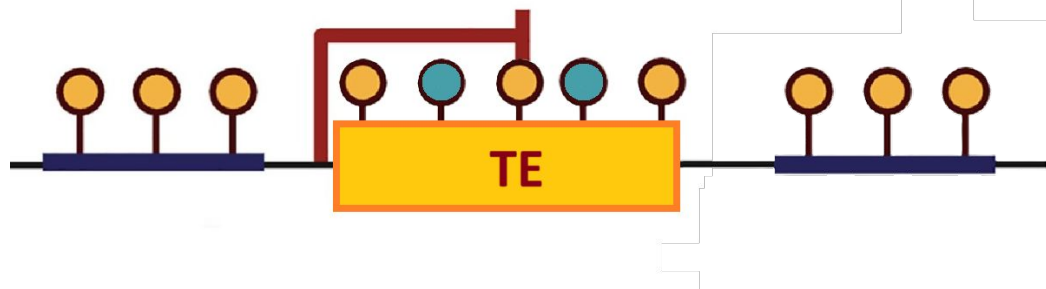
5'... GCATG▼C ... 3'
3'... CGTACG ... 5'

5'... G▼ATC ... 3'
3'... CTAG▲ ... 5'

Rangos de *size selection* y regiones evaluadas



Any context CG CHH

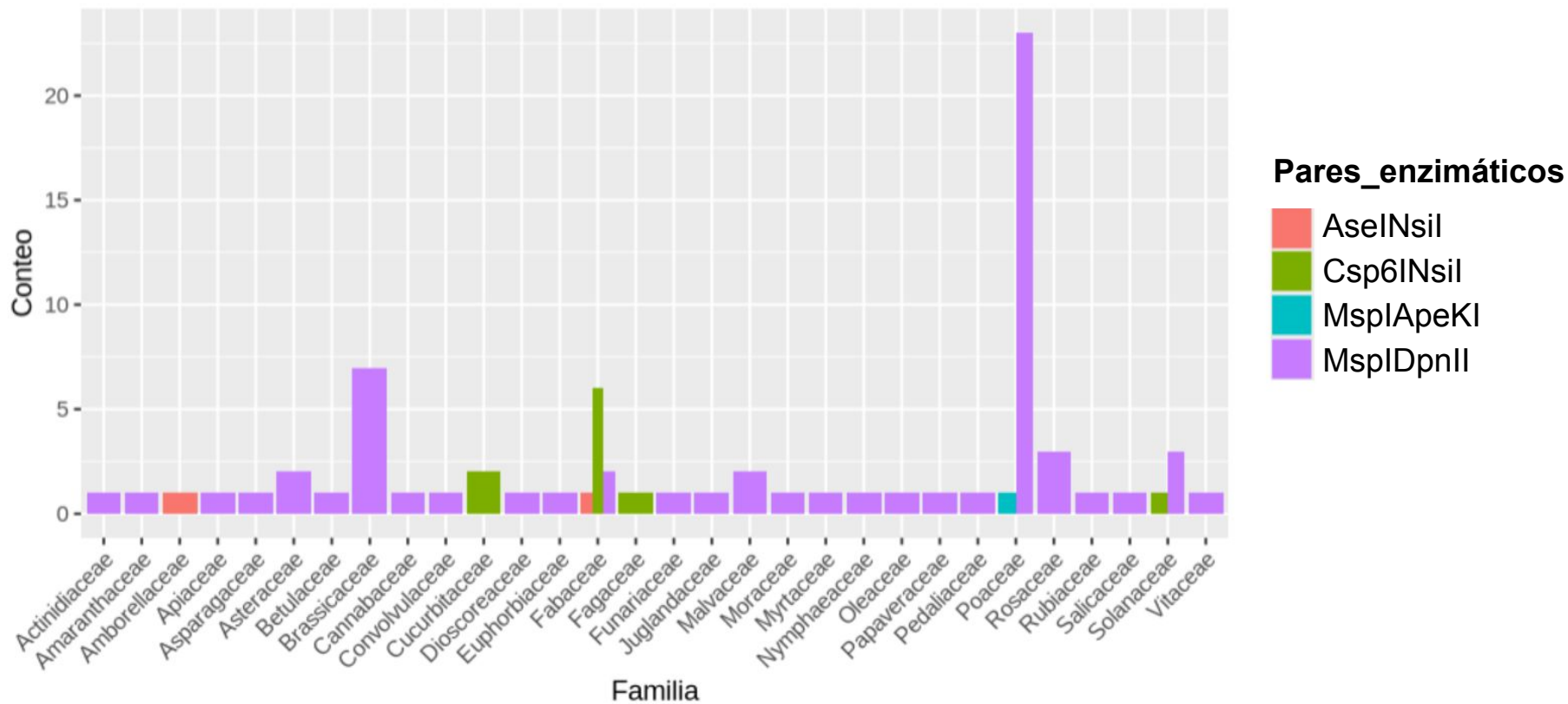


Genes, TE, sus elementos rio arriba y rio abajo (definidos como 1500 pb a cada lado)





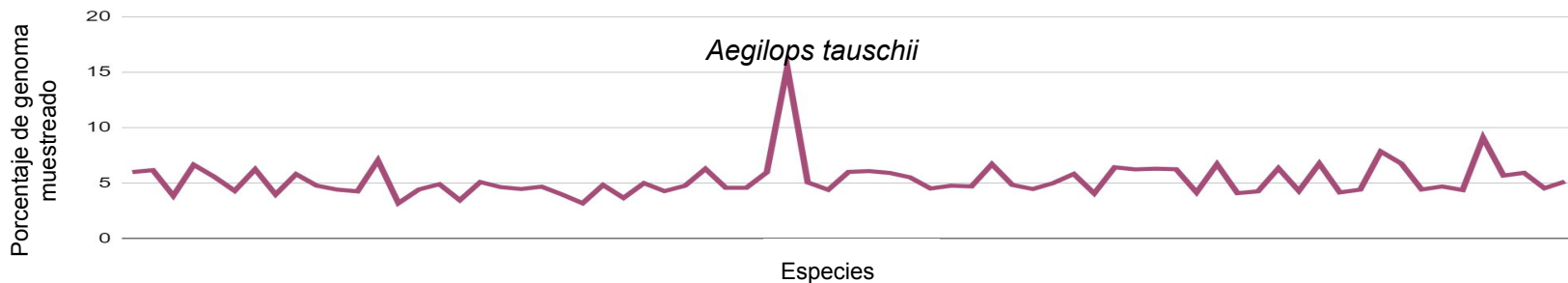
Selección del mejor par enzimático y rango - general



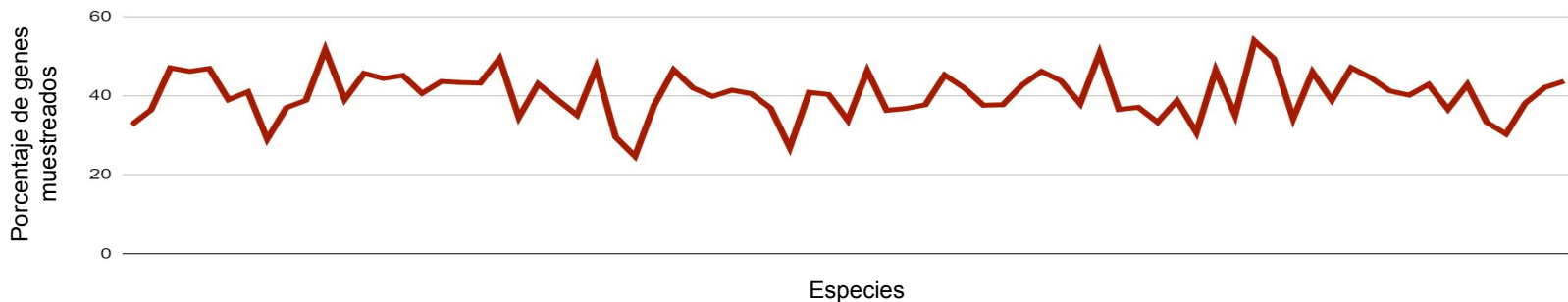


Selección del mejor par enzimático y rango - general

En todos los casos (menos uno), el mejor par enzimático seleccionado por especie muestrea alrededor del 5% del genoma) → Acorde a lo esperado para estrategias de RRBS.

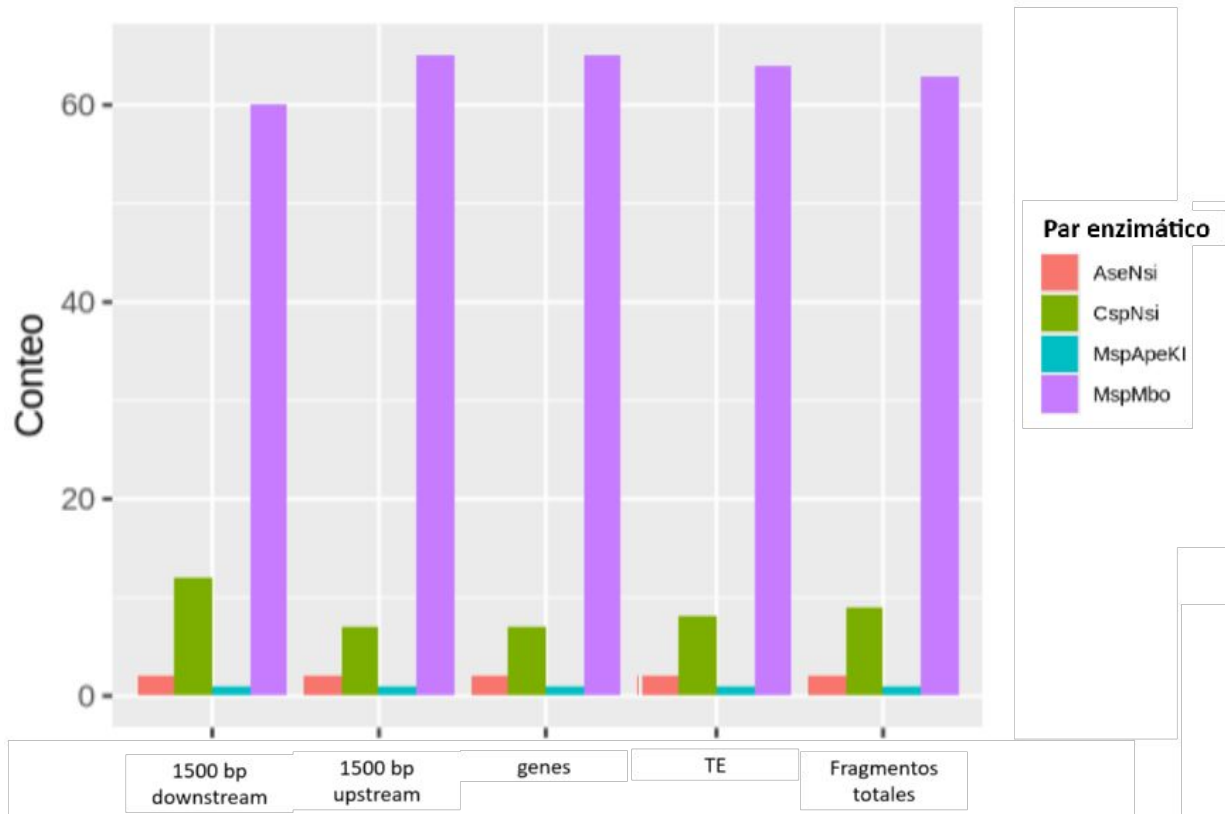


Pero el porcentaje de elementos de interés muestreado es más alto (ejemplo, genes ronda el 40%, regiones *upstream* el 25%)





Selección del mejor par enzimático y rango - por región



Dependiendo no solo de la **especie**, sino también de la **pregunta biológica** de interés (me interesa estudiar la metilación global? O focalizada en una región en particular? Ejemplo, *gene body methylation*) **varía** en par enzimático **óptimo**.

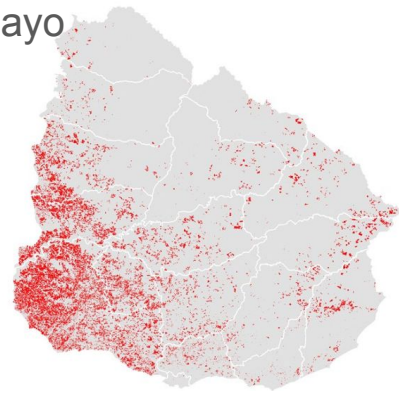


Parte II: validación - *case study* en soja (*Glycine max*)

Para soja (Genoma: 1 Gb) el mejor par enzimático predicho es Csp6INsil (Rango de Size selection de 300-450 pb)

∞ Comparar datos obtenidos por secuenciación a genoma completo tratado con bisulfito (WGBS) con los obtenidos por epi-ddRADseq con Csp6INsil y otras combinaciones.

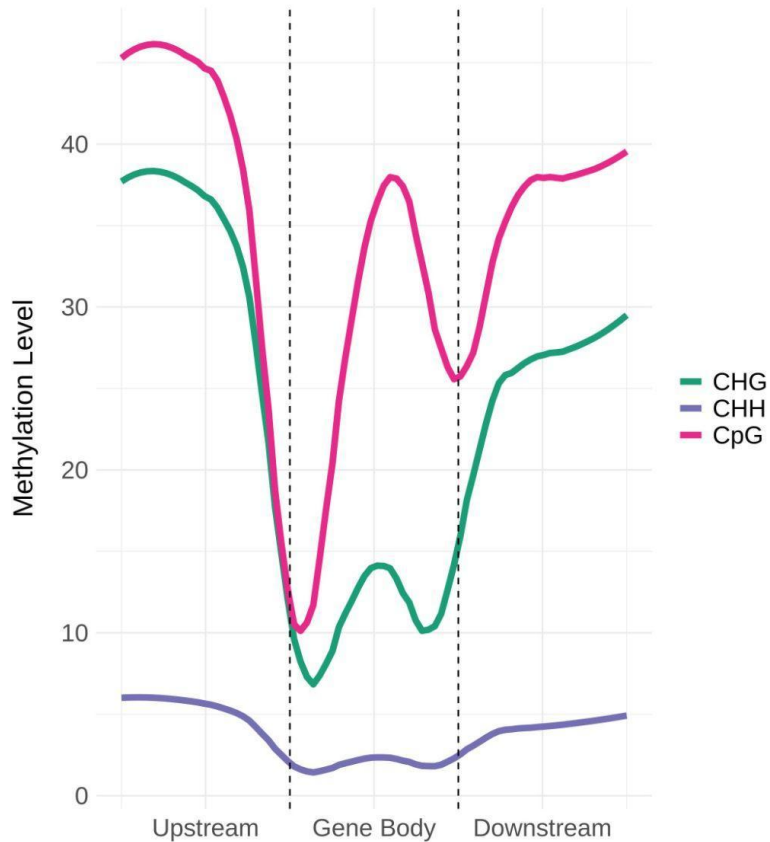
∞ Los datos WGBS fueron generados en condiciones de estrés por sequia y control. Un genotipo de soja, tres réplicas, 25 X. Datos de dos generaciones (G1 y G2, ensayo intergeneracional)



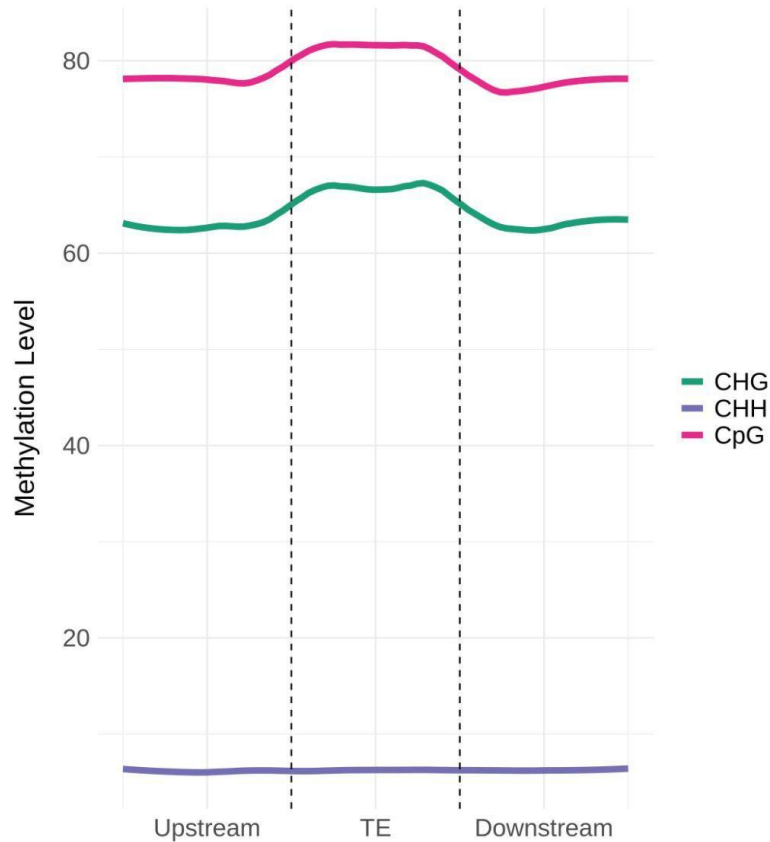


Perfiles de metilación a **genoma completo** en soja, por región

Methylation Profile Across Regions - Genes



Methylation Profile Across Regions - TE



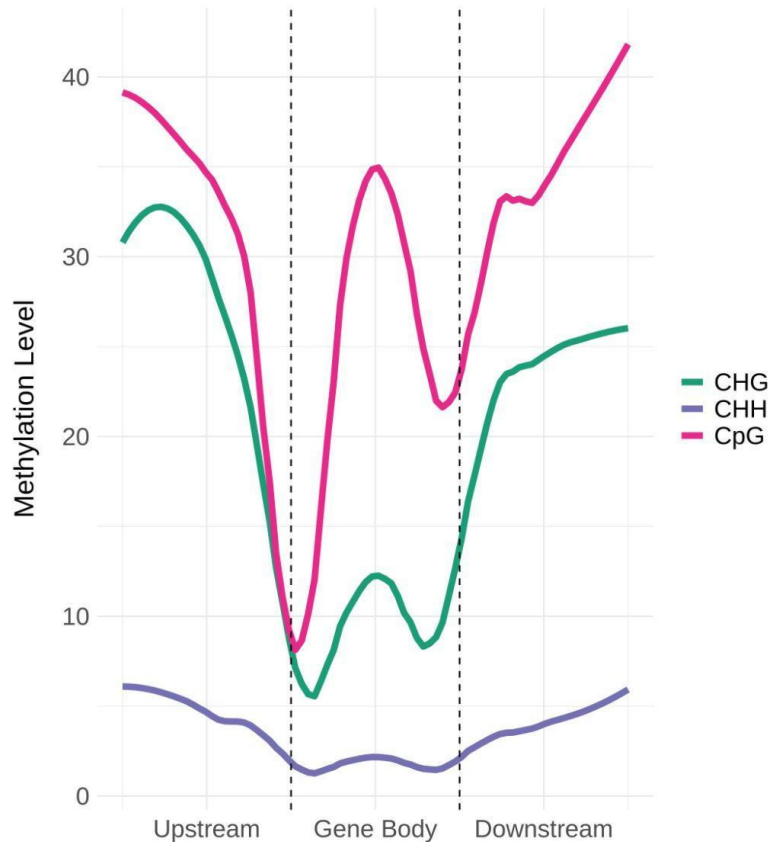
Metilación difiere
entre regiones
(genes, TE, y sus
regiones
upstream y
downstream)

Datos de WGBS

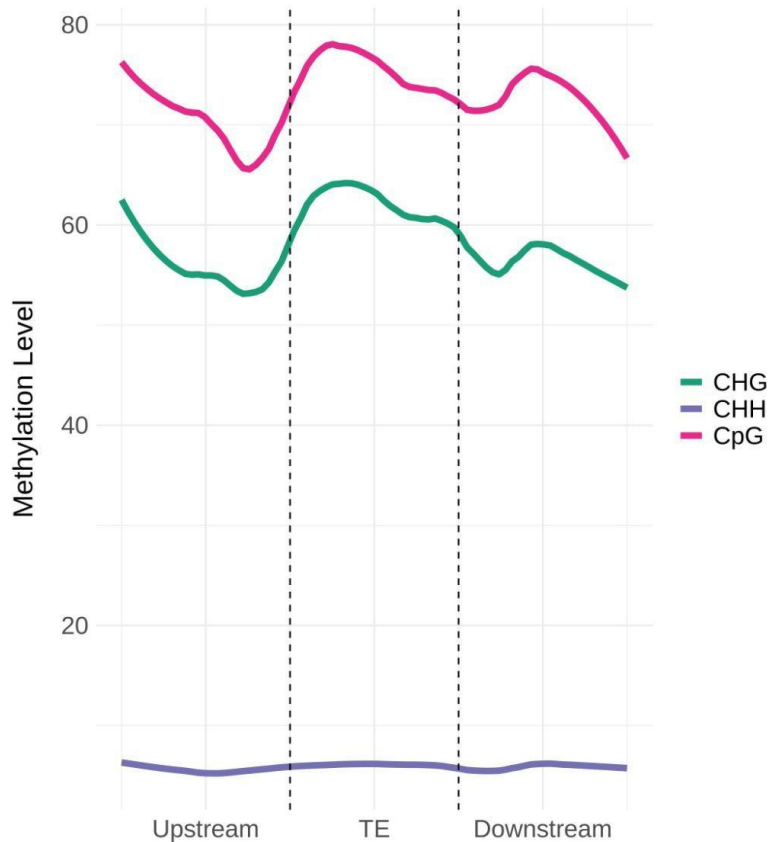


Perfiles de metilación epiddRADseq (**Csp6INsil**), por región

Methylation Profile Across Regions - Genes



Methylation Profile Across Regions - TE



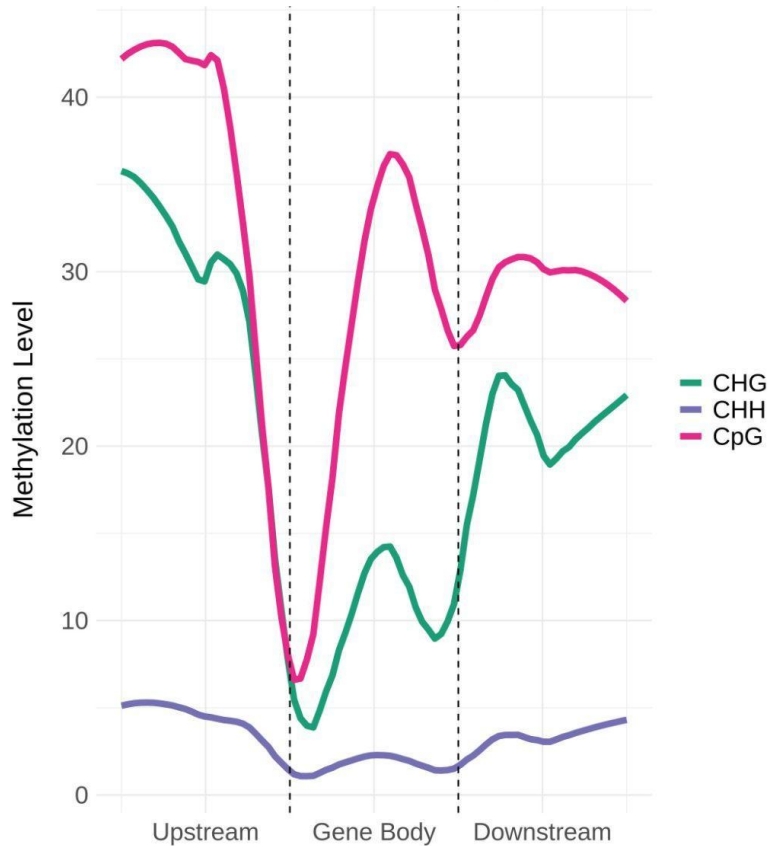
Csp6INsil fue el
mejor par
enzimático soja

Por KS en genes
no hay diferencias
entre los valores
de metilación en
genes y sus
regiones, si en
TE.

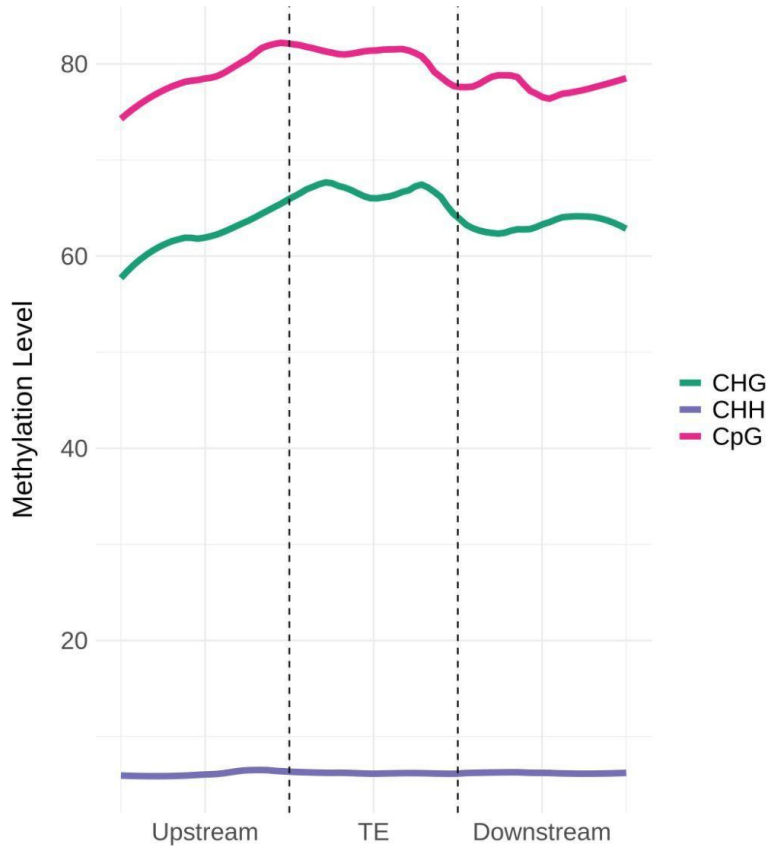


Perfiles de metilación epiddRADseq (**MspIDnplI**), por región

Methylation Profile Across Regions - Genes



Methylation Profile Across Regions - TE

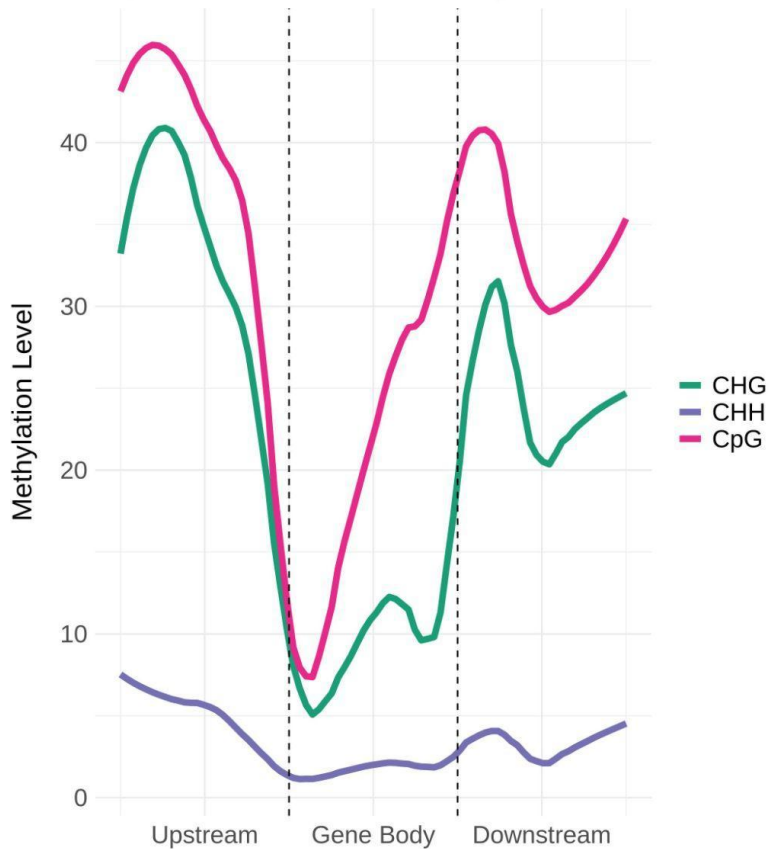


MspIDnplI fue el mejor par enzimático en la mayor parte de las poaceas y brassicaceae.

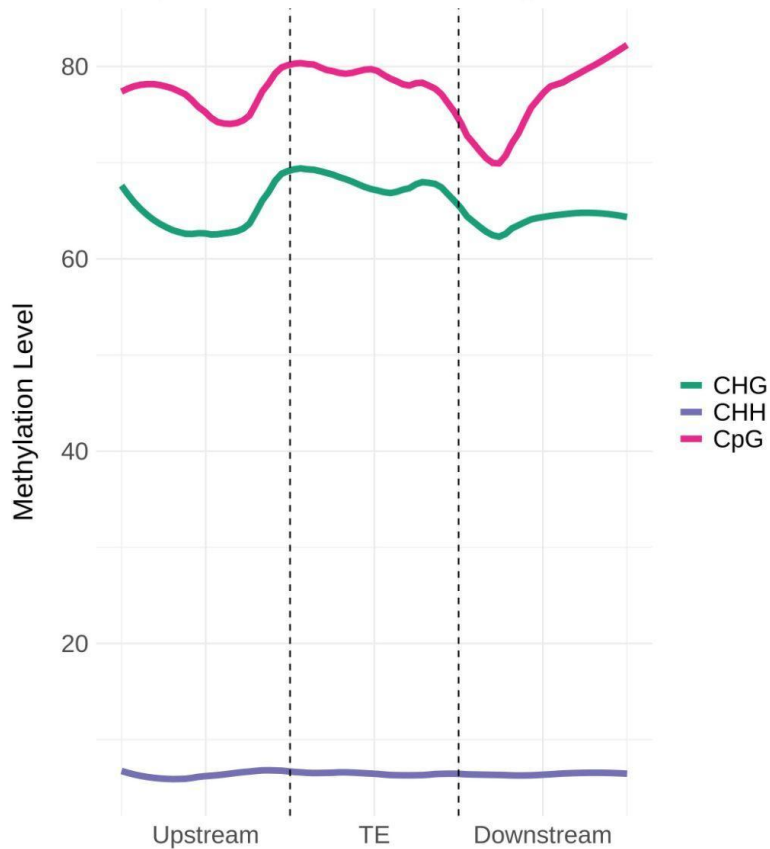


Perfiles de metilación epiddRADseq (**SphIMspl**), por región

Methylation Profile Across Regions - Genes



Methylation Profile Across Regions - TE





Cerrando la charla...

- ⌚ La estandarización en el uso de enzimas en RRBS (idem GBS, ofrecida por las *sequencing facilities*) puede generar resultados sub-óptimos → tener en cuenta especie y pregunta!
 - ☑ La epigenética es un campo que puede favorecer el desarrollo y apropiación de tecnologías ómicas locales, mediante la optimización y puesta a punto de estrategias para exploración del epigenoma → Ojo, las librerías de epiddRAD luego se secuencian en el exterior.
 - ⌚ Recordar: estamos secuenciando ADN (tratado con bisulfito) → además de marcas de metilación se pueden identificar SNPs entre individuos (estudios genéticos + epigenéticos)
 - ⌚ Aplicaciones? Cientos: diversidad (epi)genómica, EWAS (como GWAS pero epi), metilación diferencial en respuesta a estrés, herencia...
- ➡ Nuestros datos de WGBS pertenecen un estudio de epigenetica del estres ambiental y memoria intergeneracional. Los resultados apuntan a que existen más DMR conservadas a nivel de promotores. Por ende para nuestra pregunta biológica, seleccionaremos enzimas que favorezcan el muestreo de promotores.



Agradecimientos



- Natalia Aguirre
- Norma Paniego
- Verónica Lia
- Andrea Puebla



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



- Mateo Bertolotti
- Omar Borsani
- Mariana Sotelo
- Martha Sainz
- Selene Piriz
- Mauro Martinez



UNIVERSITÀ
DEGLI STUDI
DI UDINE

- Emanuele De Paoli
- Giusi Zaina



- Sarah Dyer



AGENCIA NACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN