

The plant retrovirome and its regulation by small RNA.

CSHL Advanced Sequencing Course
November 15th 2022



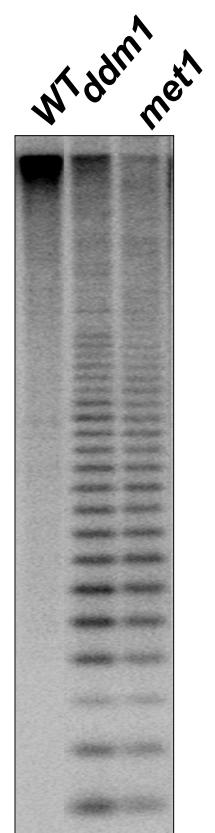
COLD SPRING HARBOR
LABORATORY



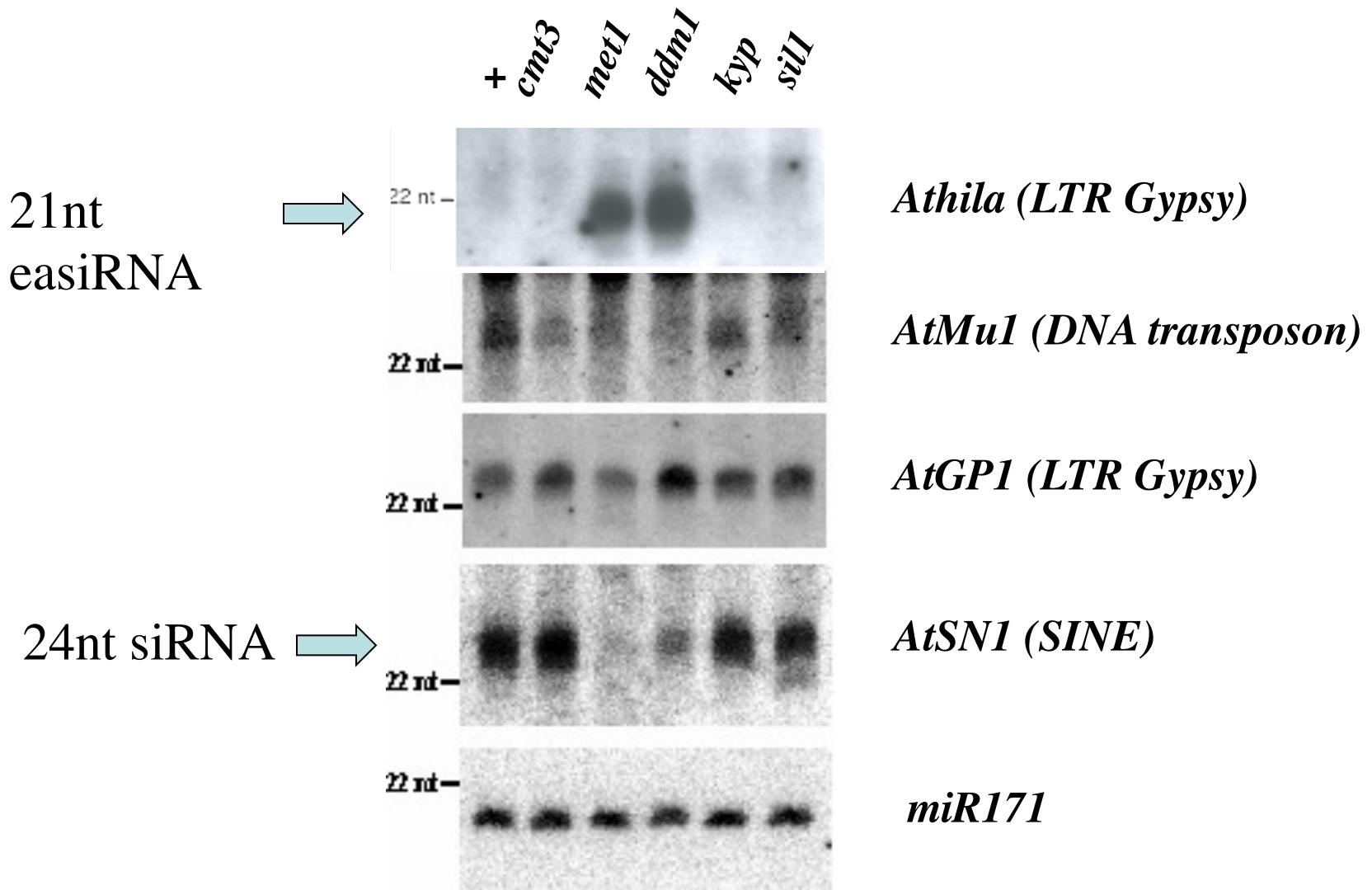
Howard Hughes
Medical Institute

DDM1 and *MET1* maintain DNA methylation in plants (and mammals)

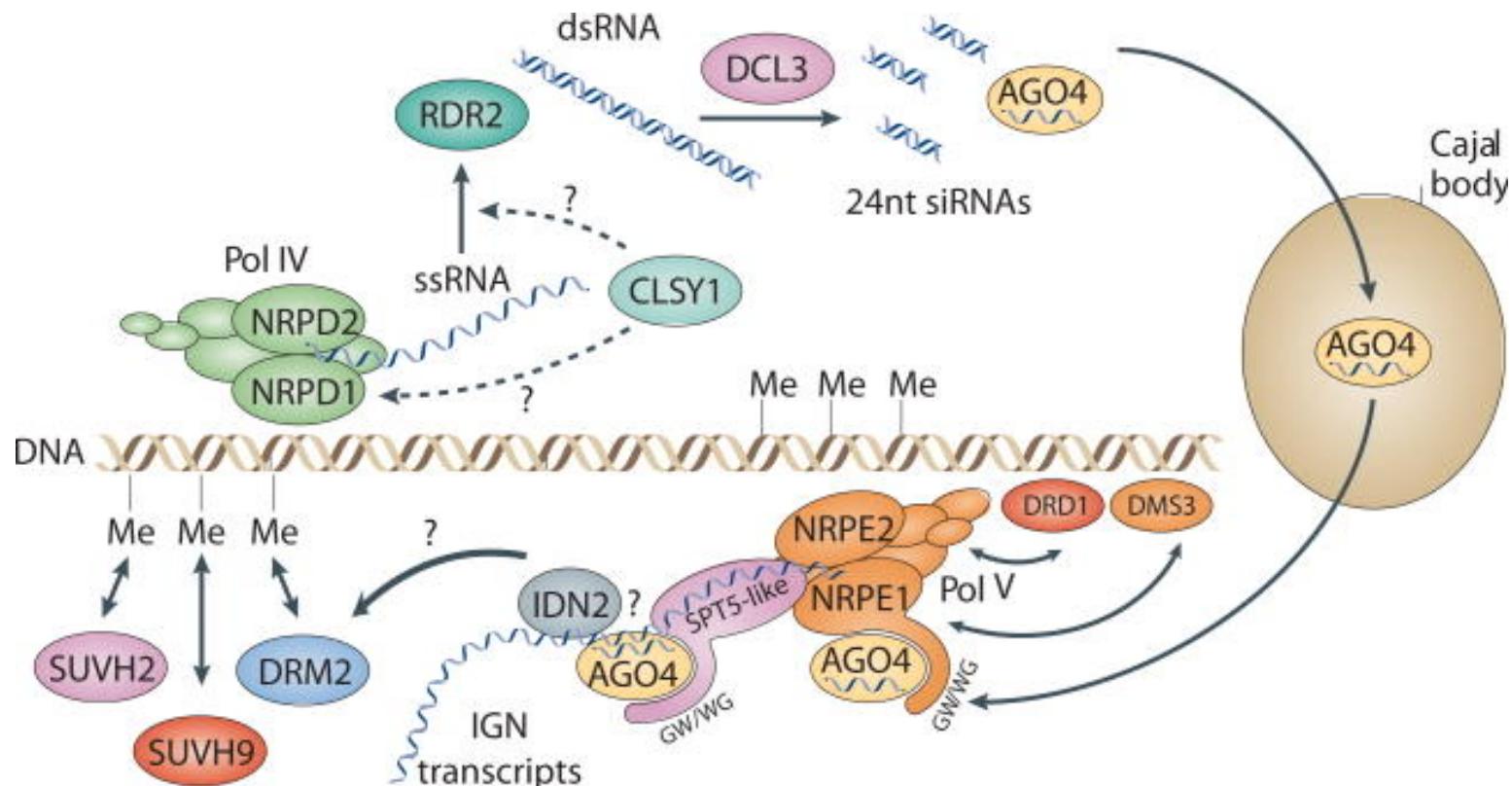
- *DDM1* (*decrease in DNA methylation1*) encodes a SWI2/SNF2 ATPase (Lsh1)
- *MET1* (*ddm2*) encodes DNA methyltransferase (Dnmt1)
- Both strongly effect expression of retrotransposons (endogenous retroviruses)



21nt “epigenetically activated” easiRNA accumulate in *metl* and *ddm1*



24nt small RNA guides CHH methylation to plant transposons

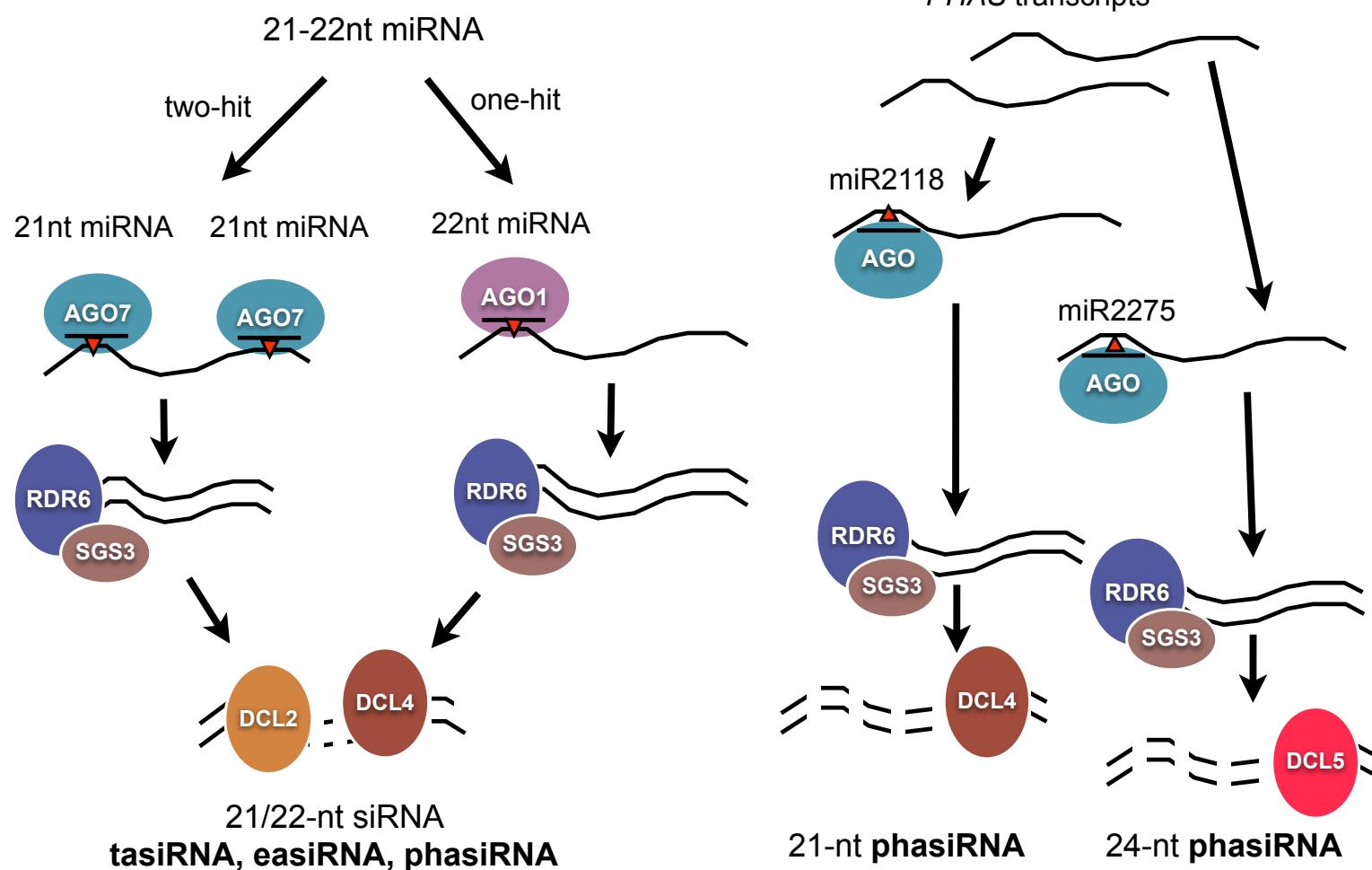


Nat Rev Genet. 2010 March ; 11(3): 204–220. doi:10.1038/nrg2719.

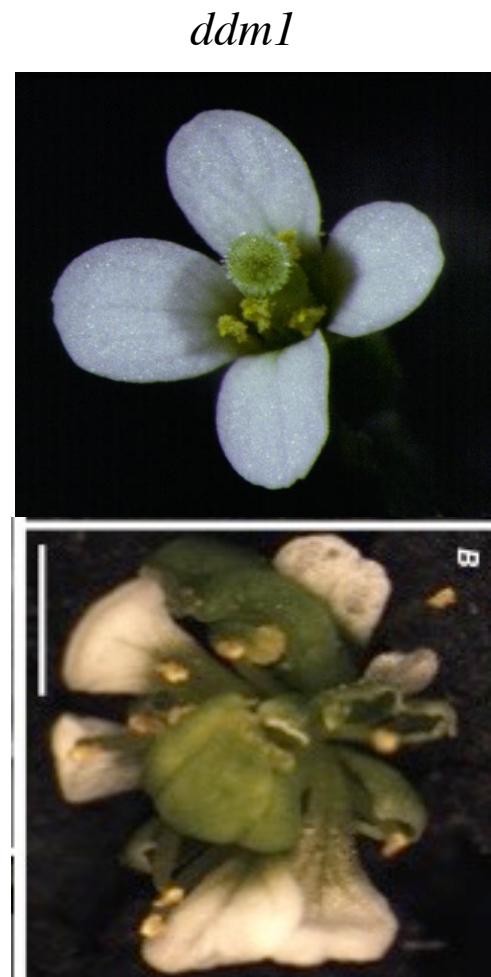
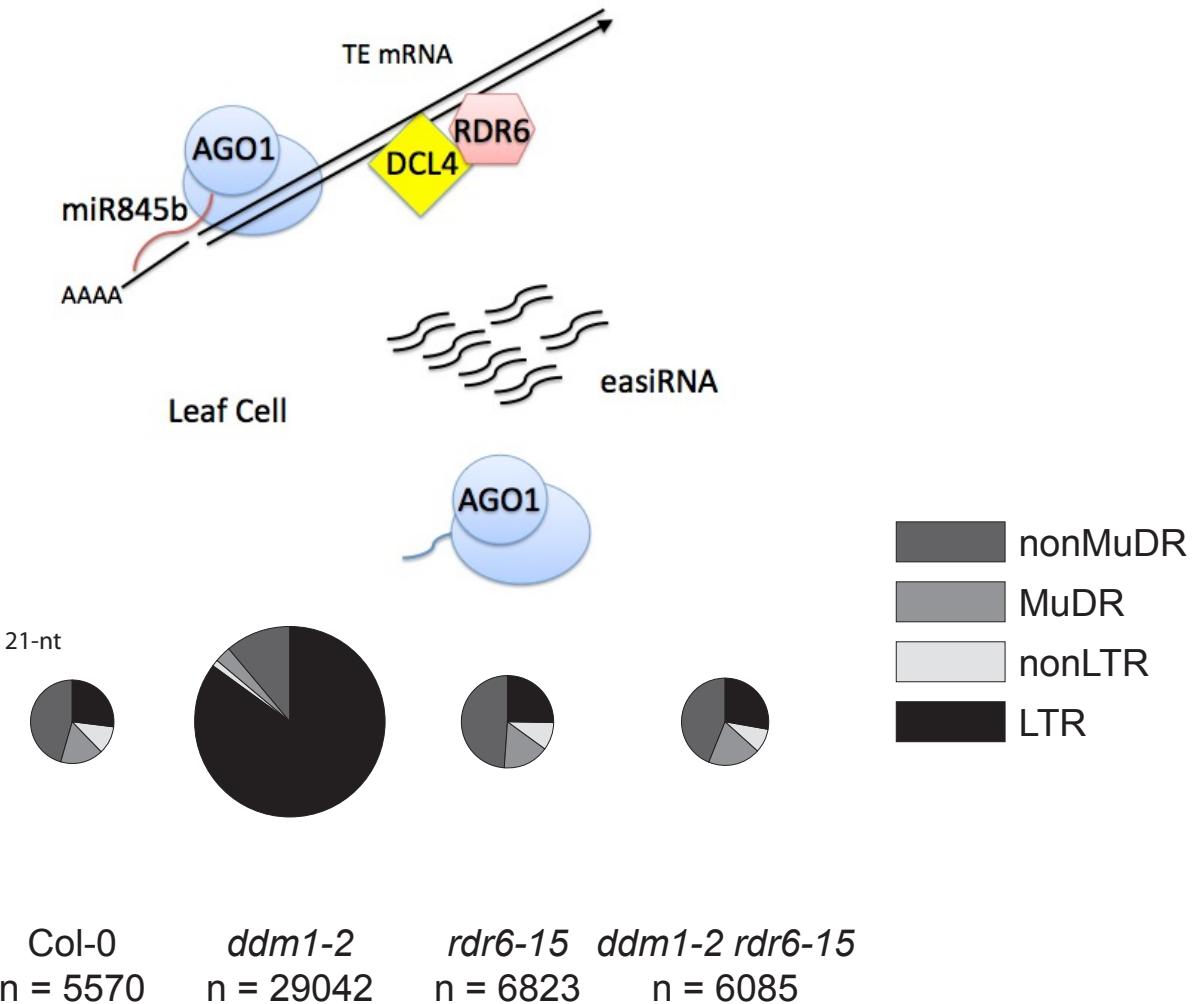
Establishing, maintaining and modifying DNA methylation patterns in plants and animals

Julie A. Law¹ and Steven E. Jacobsen^{1,2,*}

miRNA triggers secondary siRNA biogenesis in plants

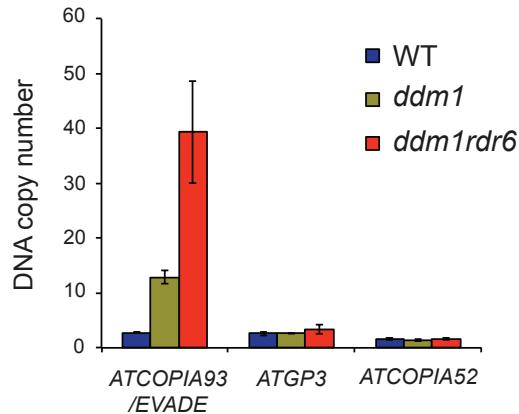


21nt easiRNA biogenesis is triggered by miRNA and is required for fertility

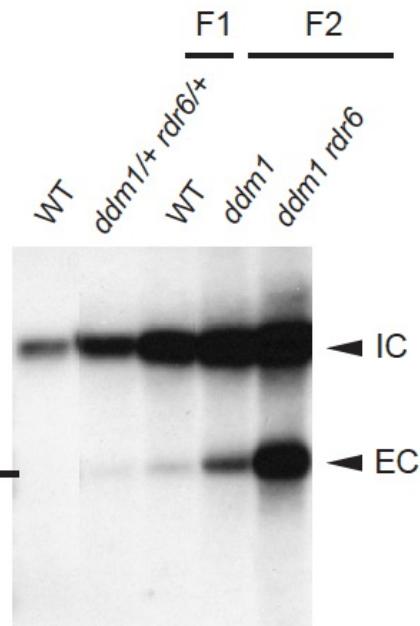
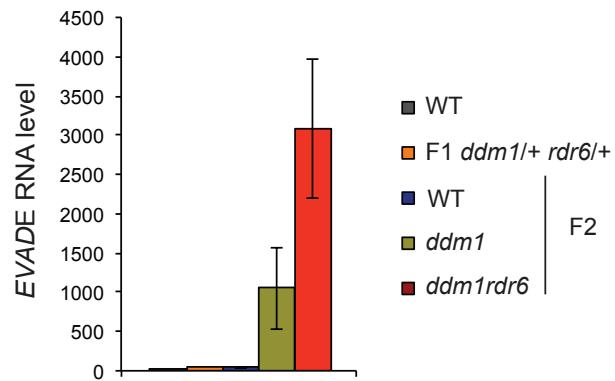


easiRNA inhibit Copia RNA accumulation (RNAi)

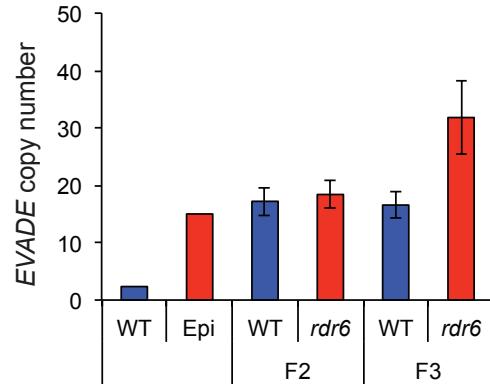
A



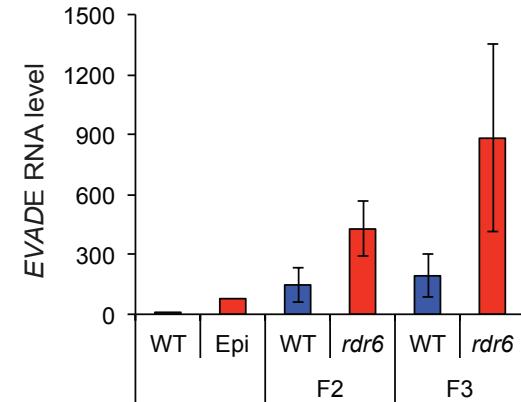
B



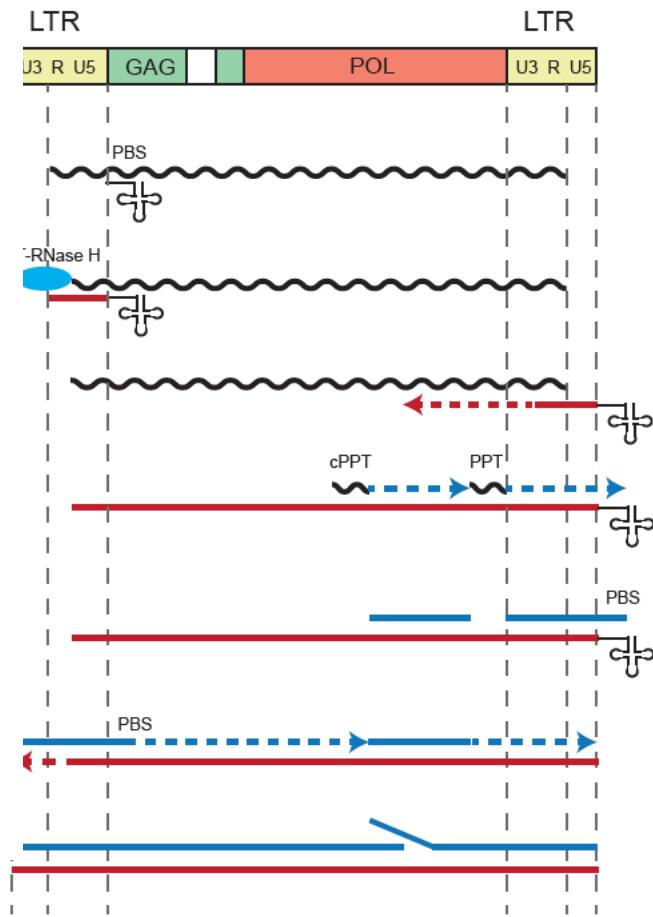
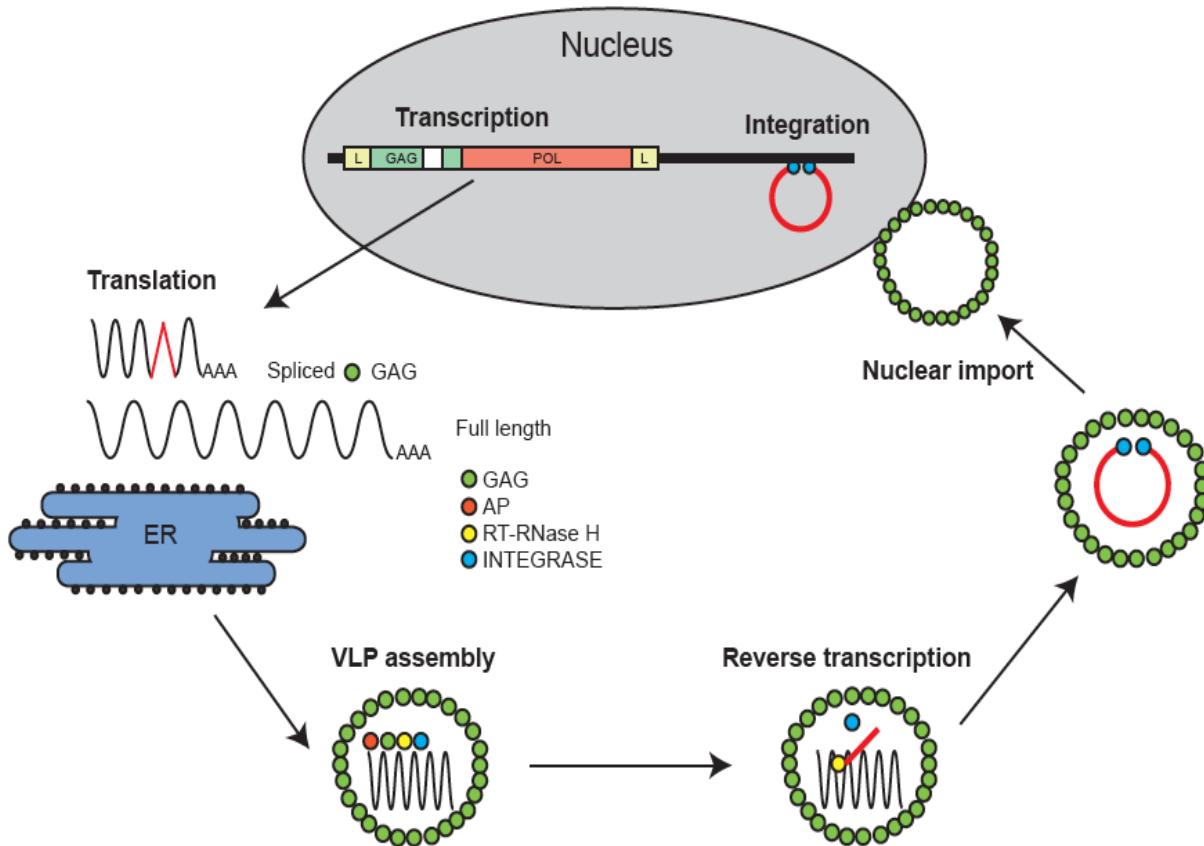
C



D

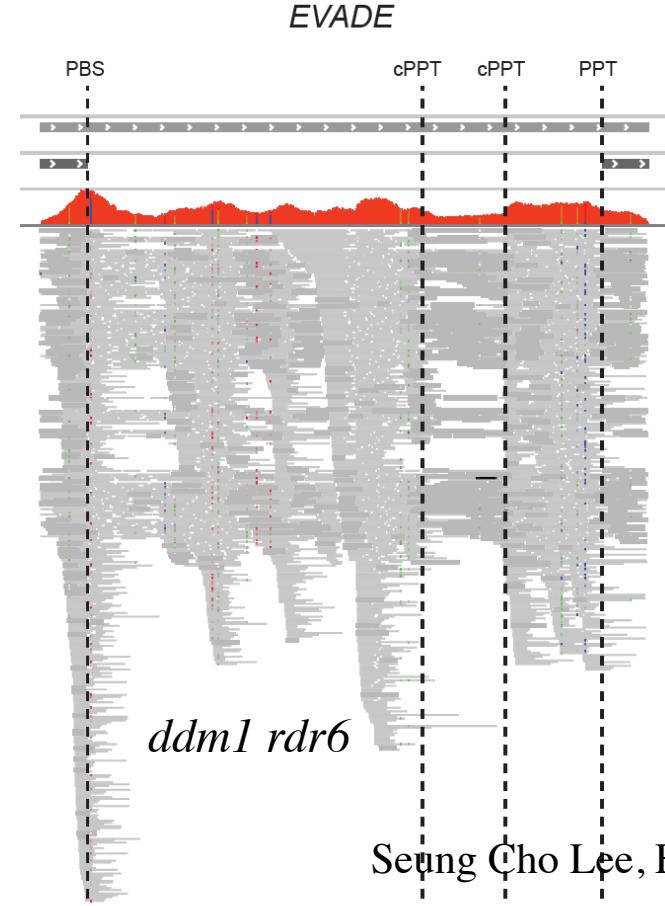
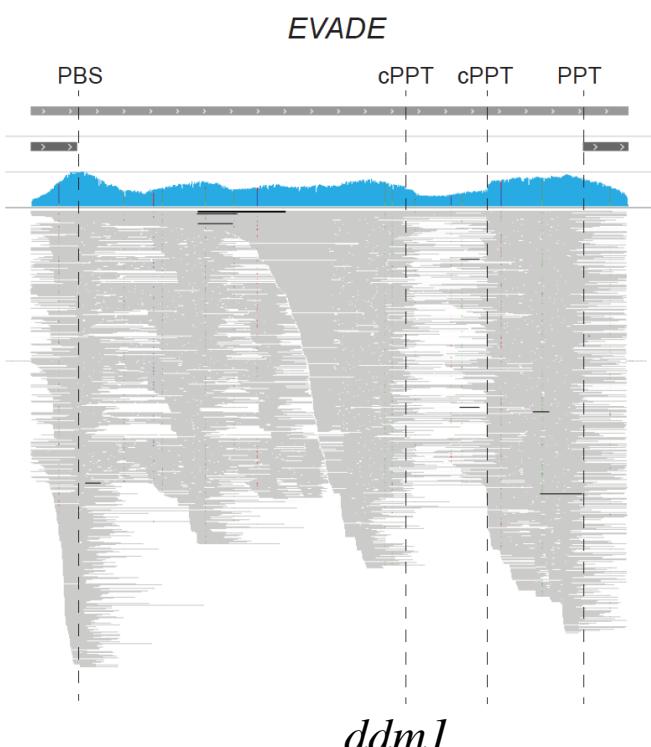


Virus-like particle (VLP) sequencing of retrotransposition intermediates

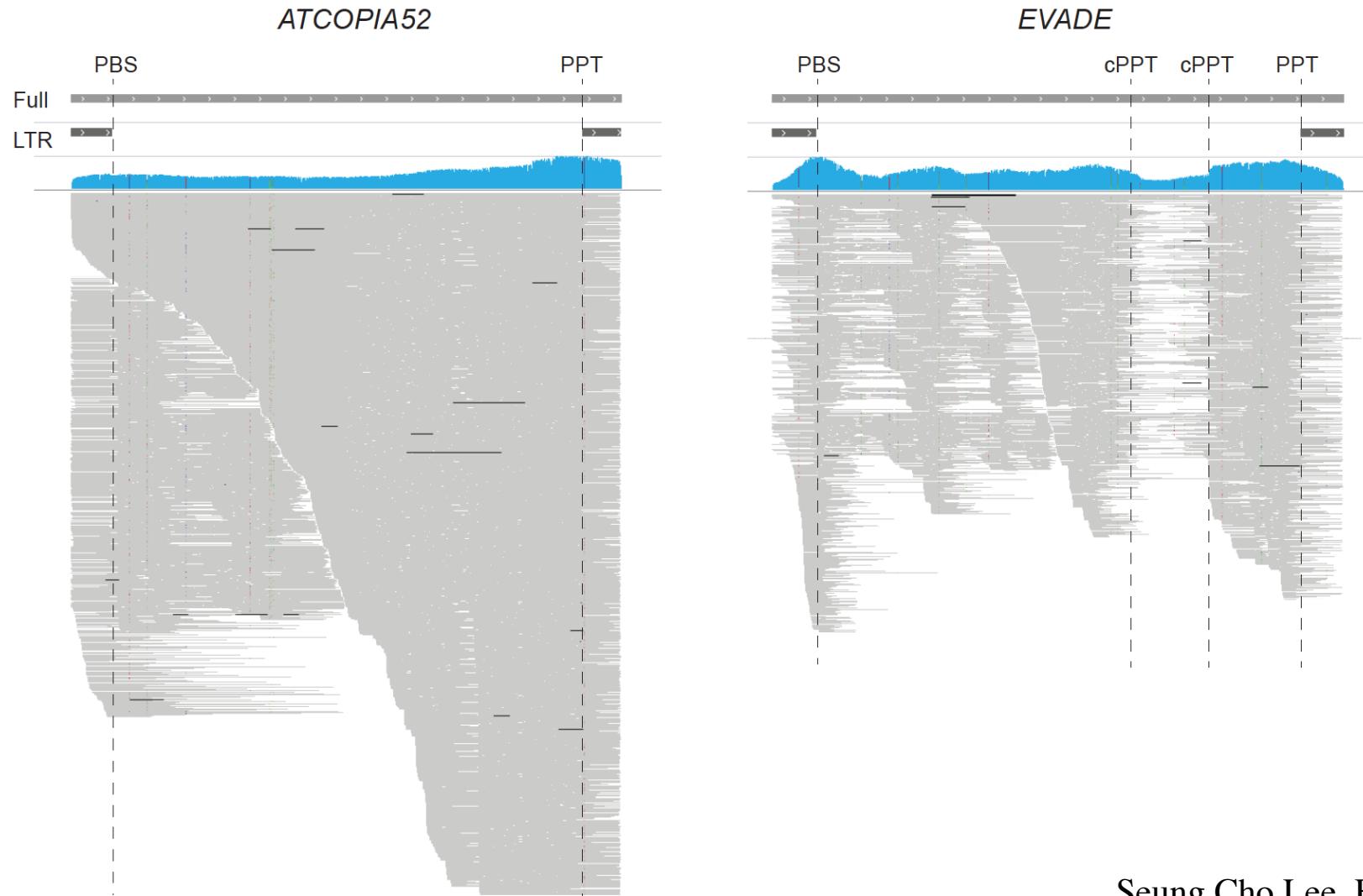


Nanopore VLP sequencing of retrotransposition intermediates

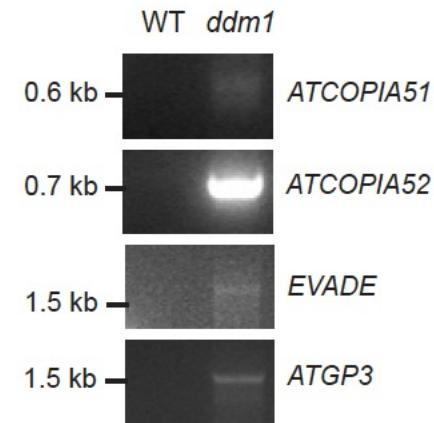
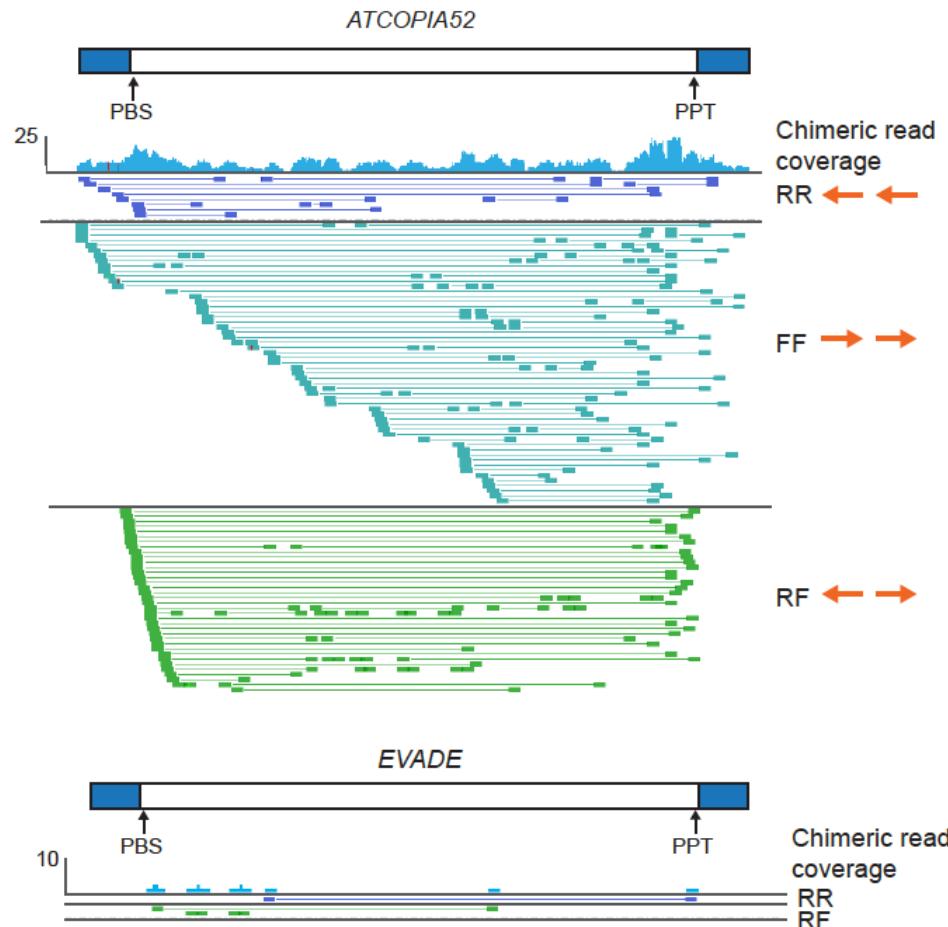
- Virus like particles (VLP) purified from *ddm1* and *ddm1rdr6*
- dsDNA subjected to long-read Oxford Nanopore sequencing



COPIA52 preferentially forms autointegration circles

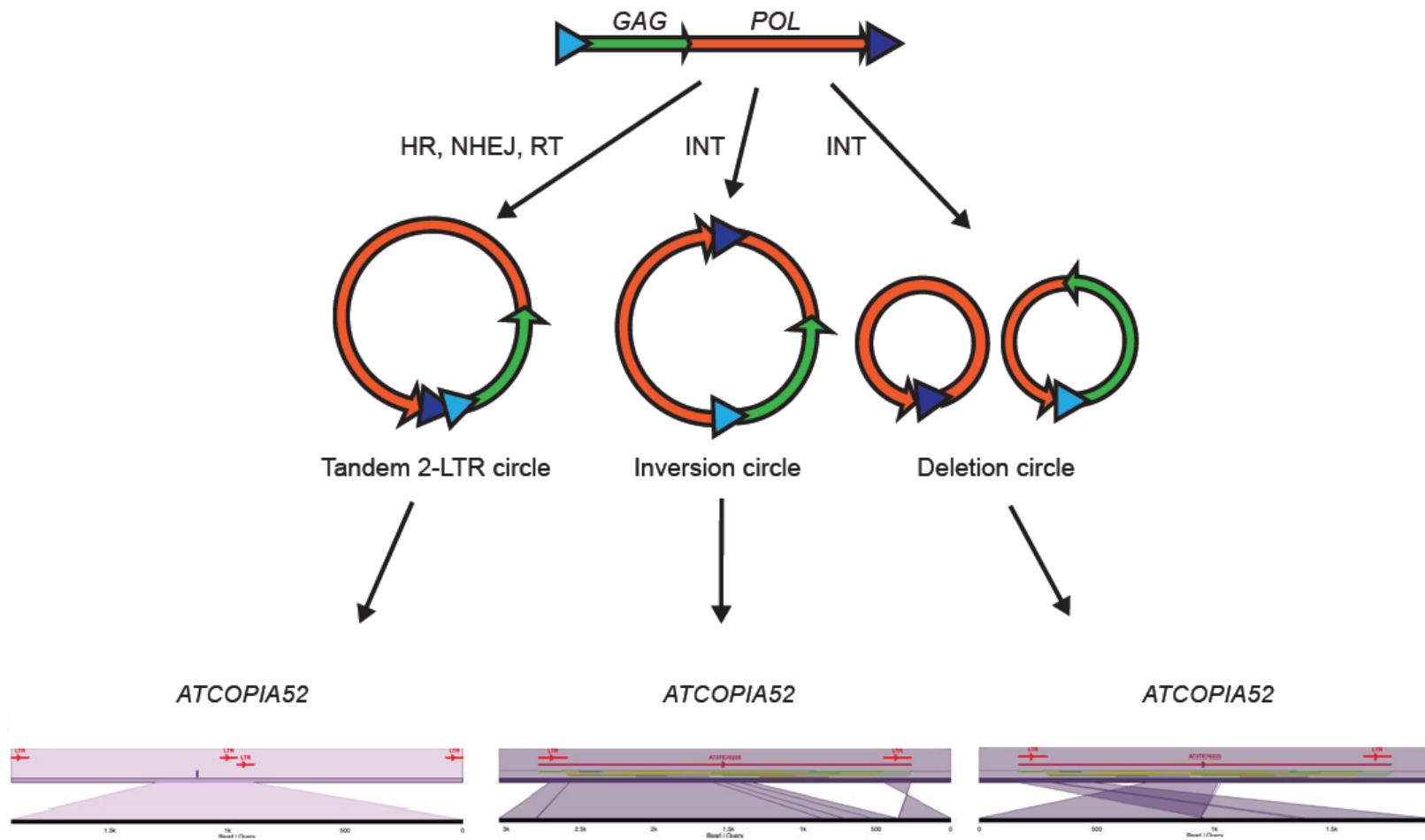


COPIA52 preferentially forms autointegration circles



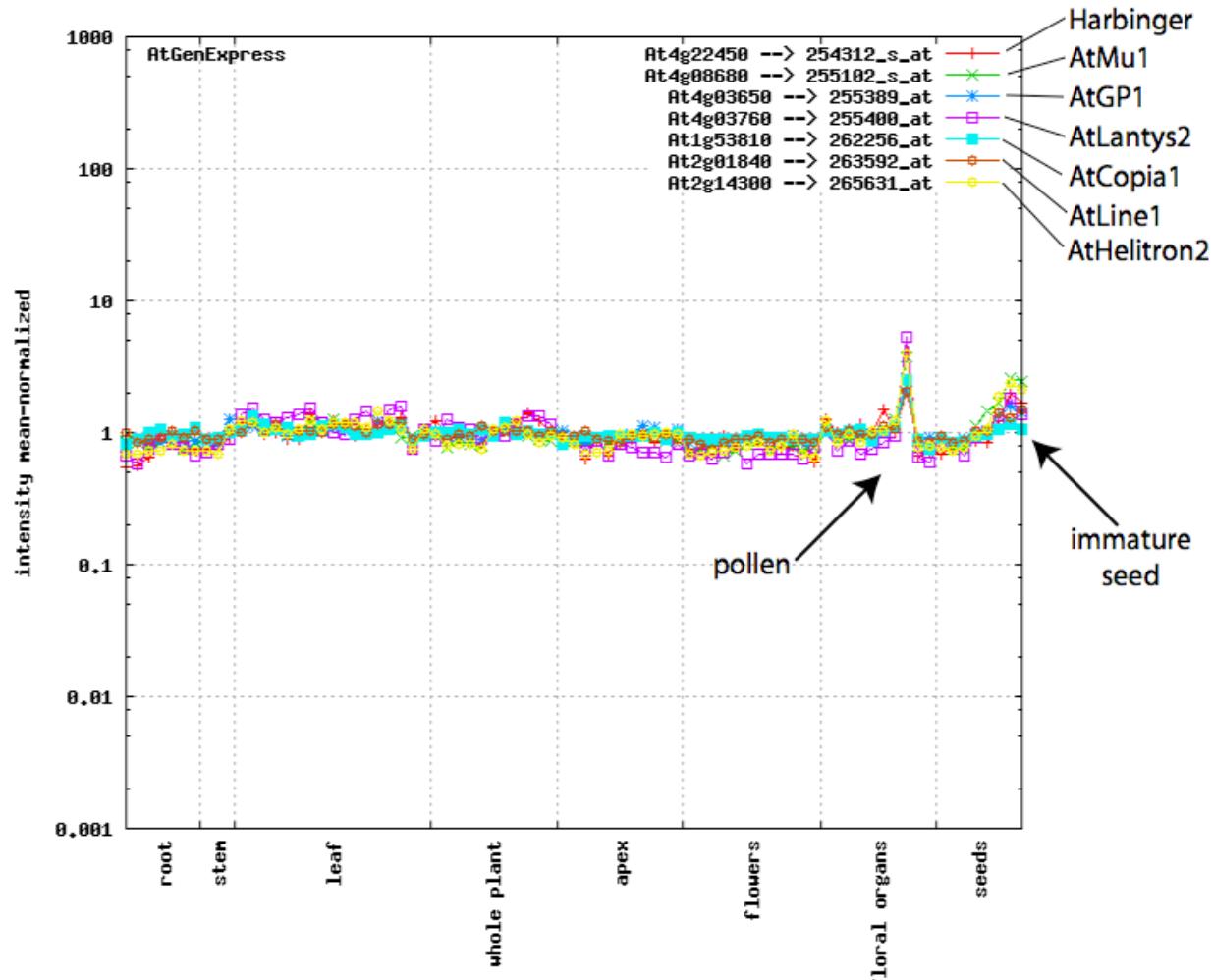
- VLP dsDNA sequenced by paired end Illumina short reads
- reveals non-concordant read orientations (autointegration circles)

Retrotransposon autointegration

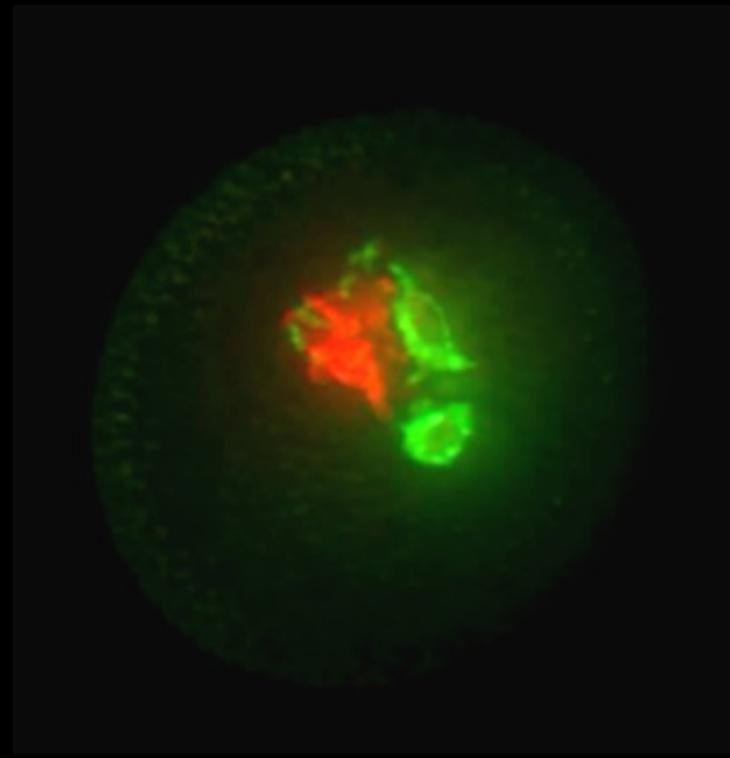


Transposons are expressed in wild-type pollen

Microarray
expression
profiles

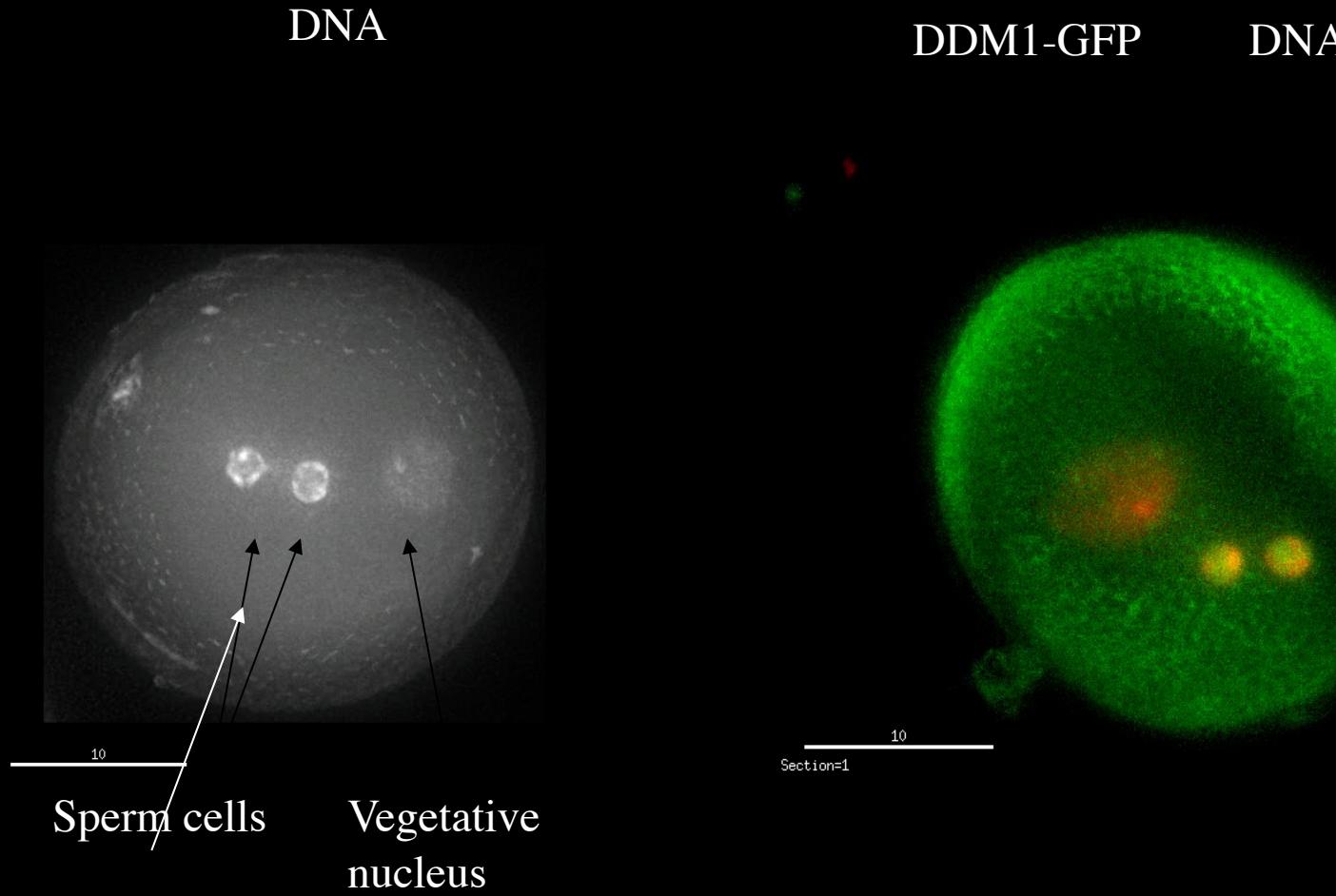


Germline reprogramming in plants

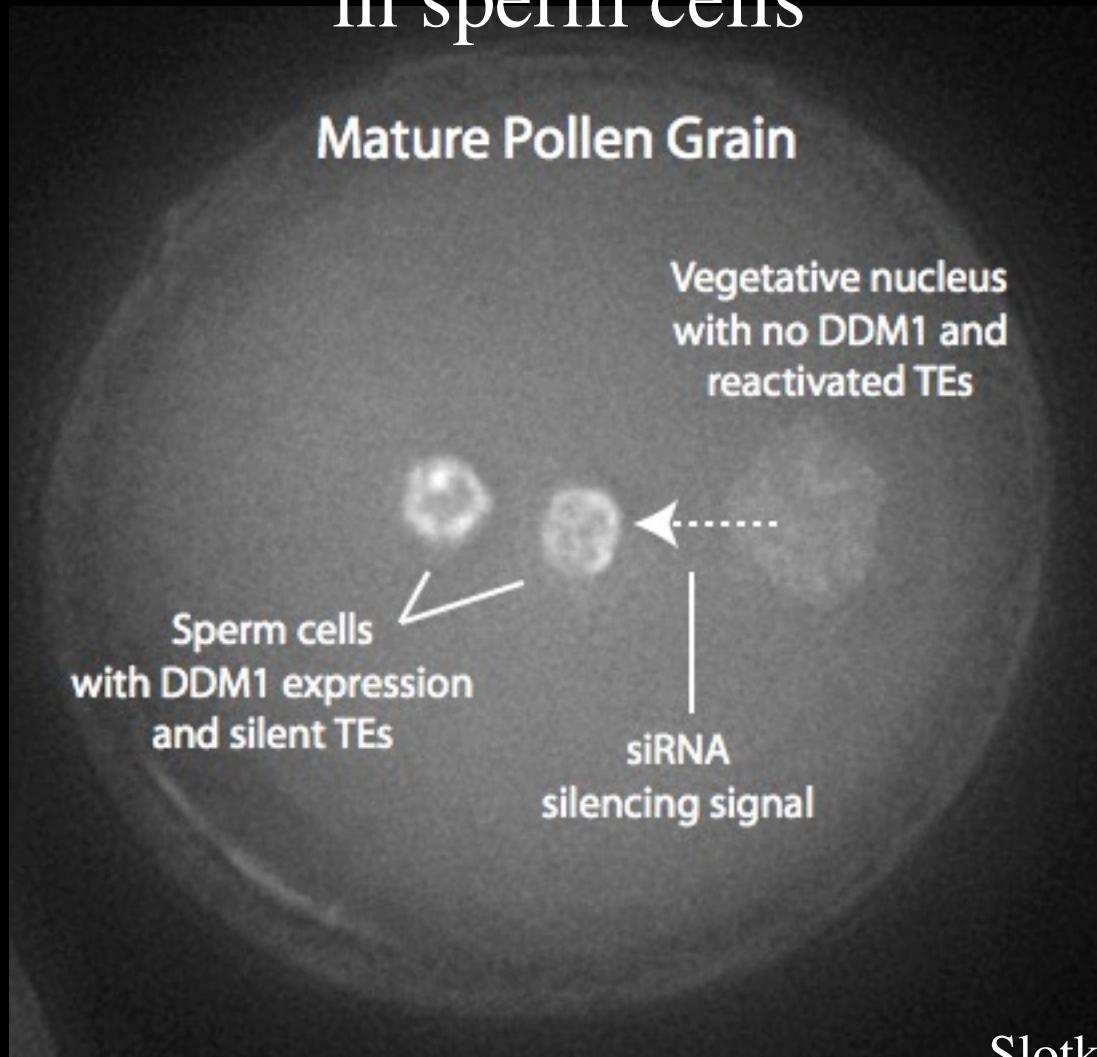


Filipe Borges

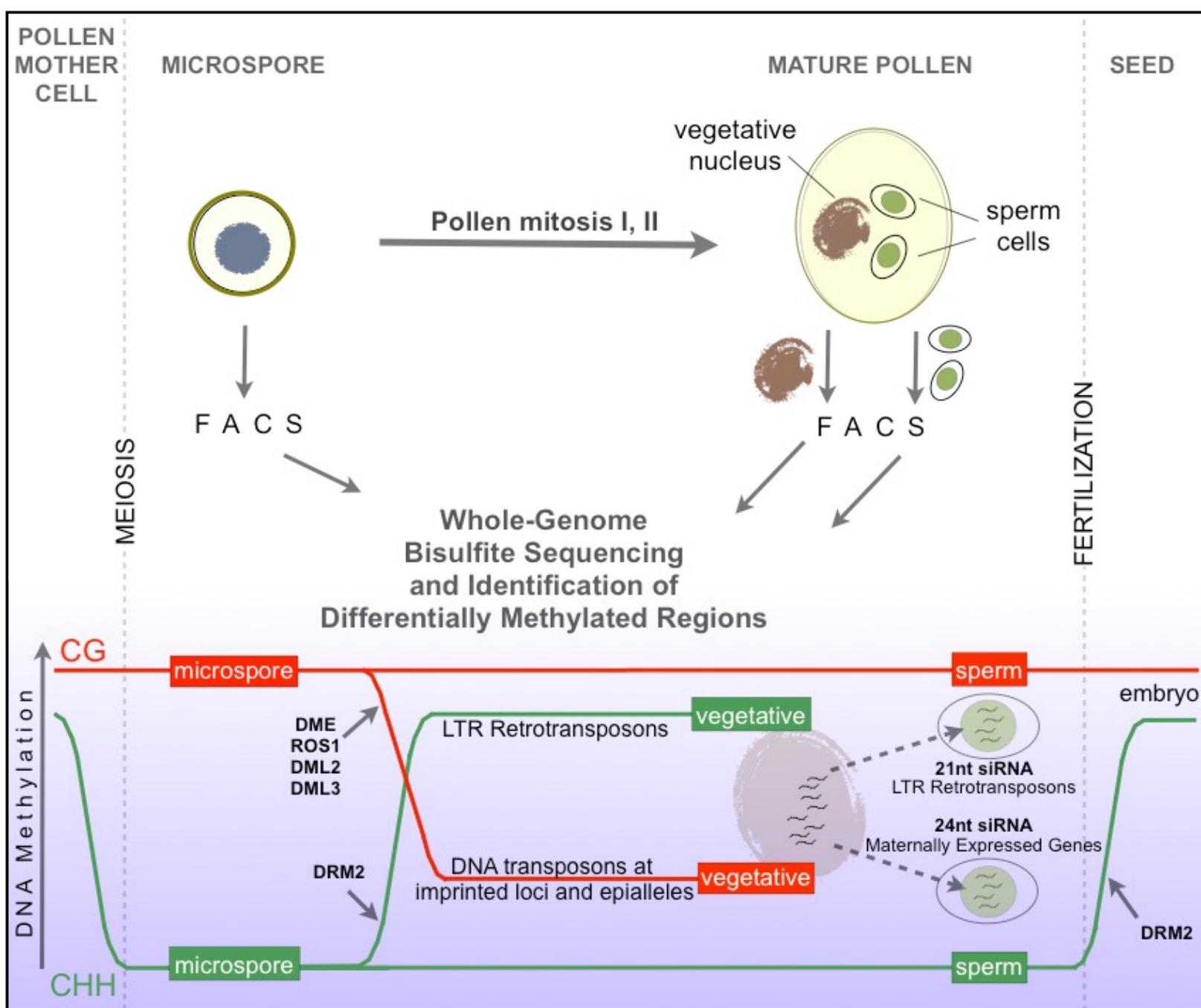
DDM1 is expressed in sperm cells, but not in the vegetative nucleus



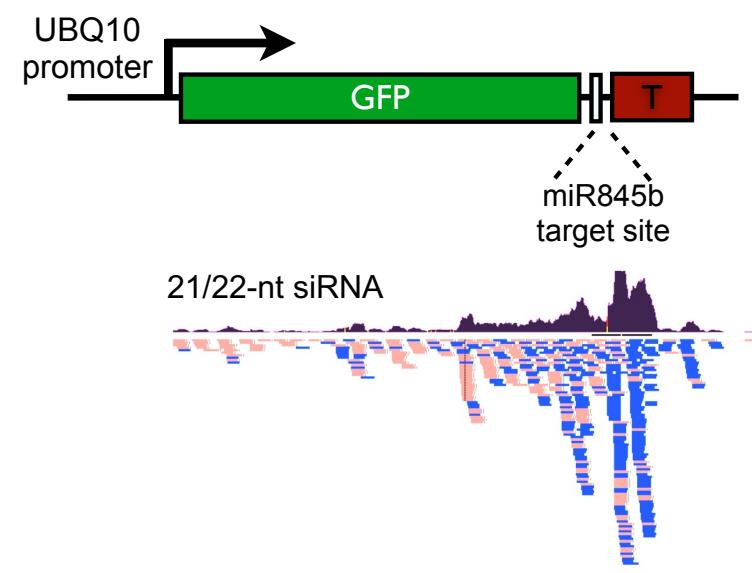
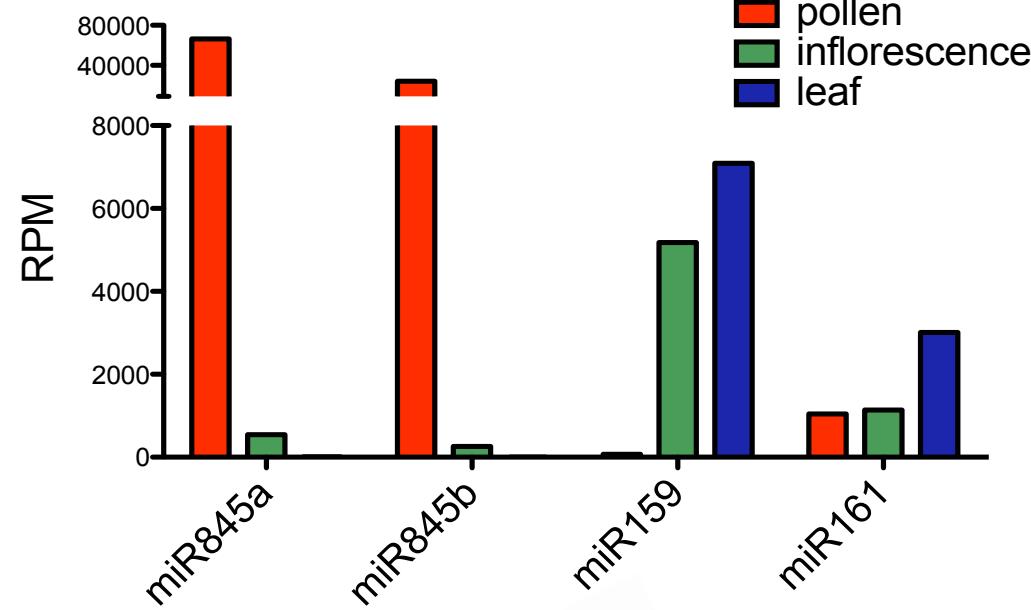
Small RNA generated in pollen accumulates in sperm cells



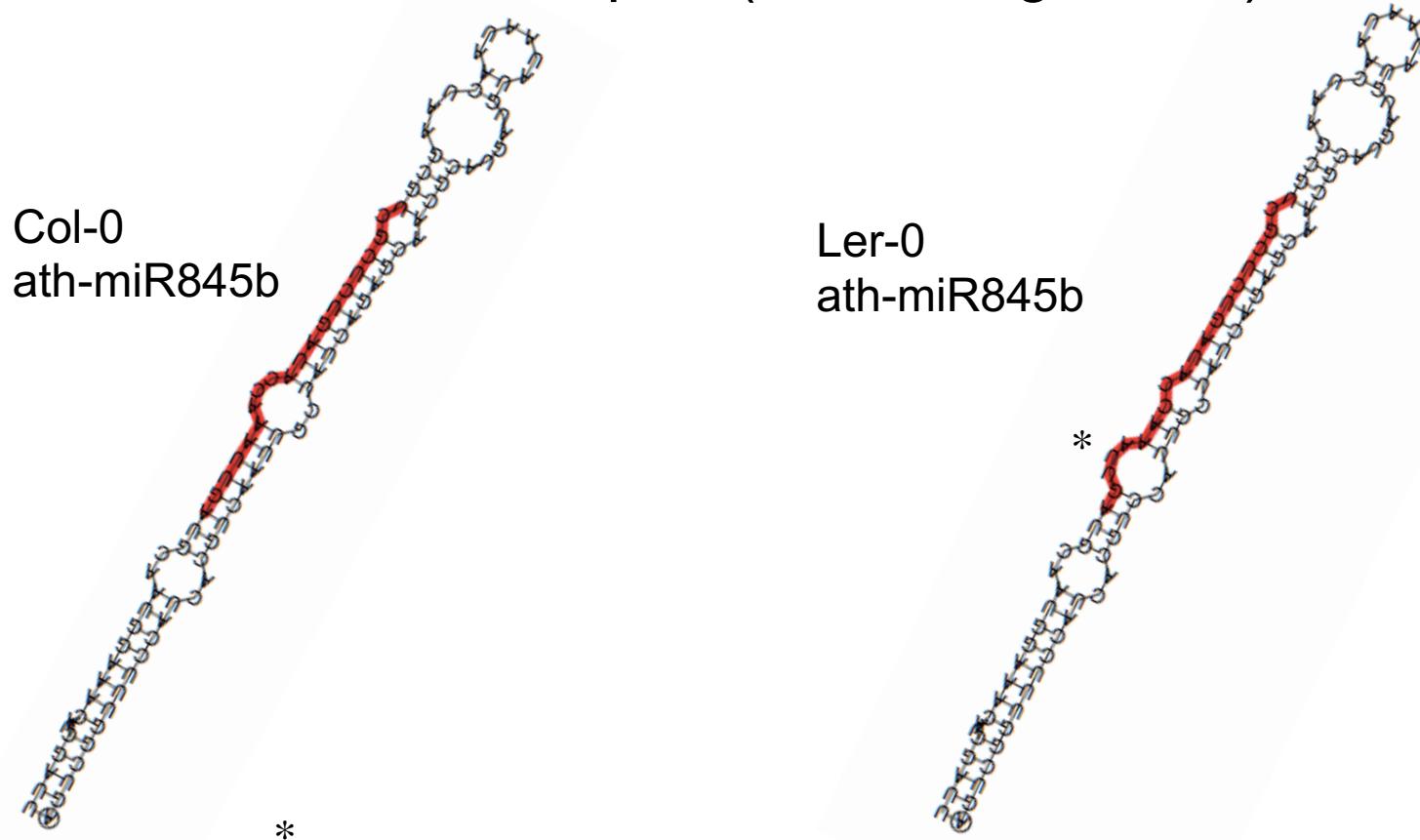
Slotkin et al., *Cell* 2009



miR845 is very abundant in pollen and triggers easiRNA

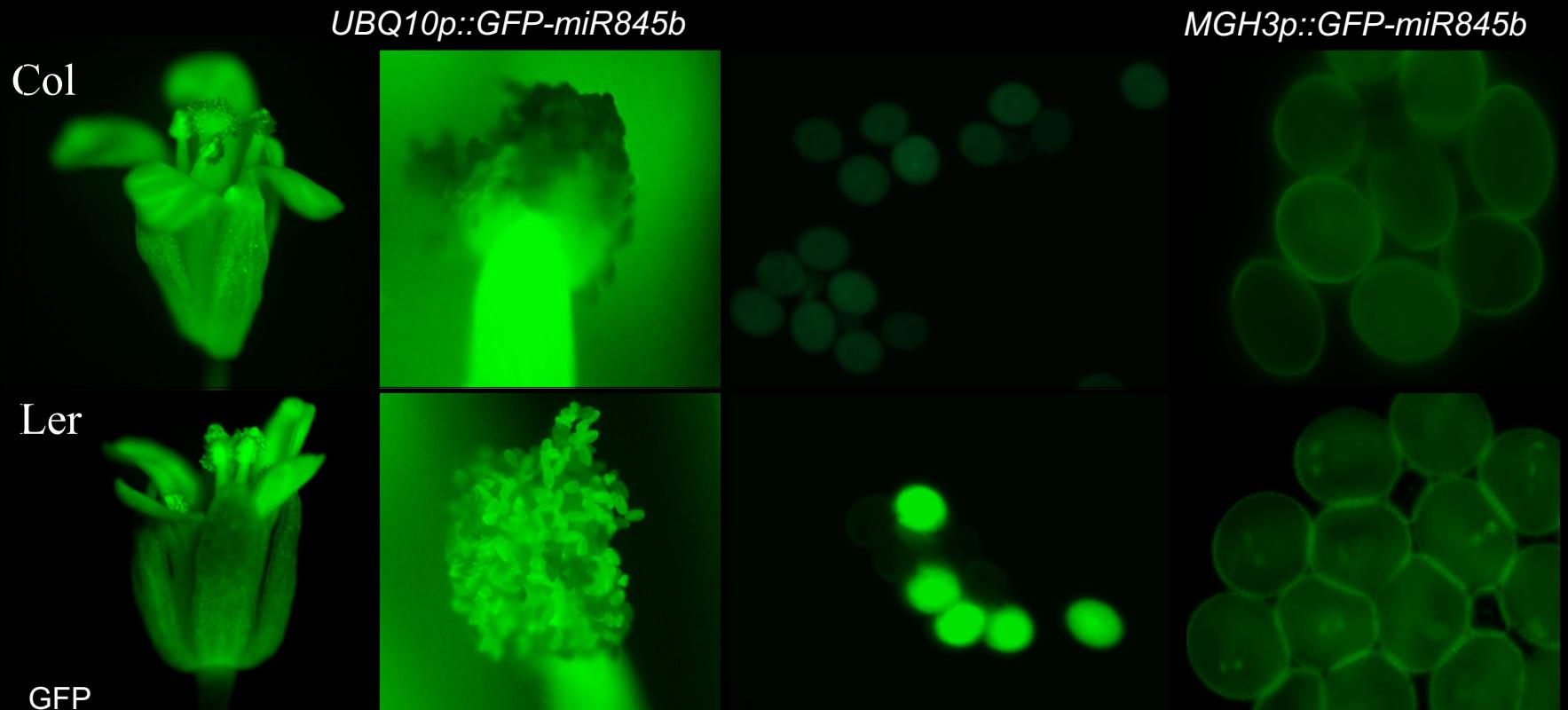


MIR845b polymorphism blocks processing in some strains of Arabidopsis (*Landsberg erecta*)



Col-0	1	AGTCGGTTCCATCACGTCAATTGCTATCAGAGCAACGCATGATGTATAAACTAAGCGTCGCTCTGATACCAAATTGATGCAATGGAAACATCGATT	99
Ler-0	1	AGTCGGTTCCATCACGTCCATTGCTATCAGAGCAACGCATGATGTATAAACTAAGCGTCGCTCTGATACCAAATTGATGCAATGGAAACATCGATT	99

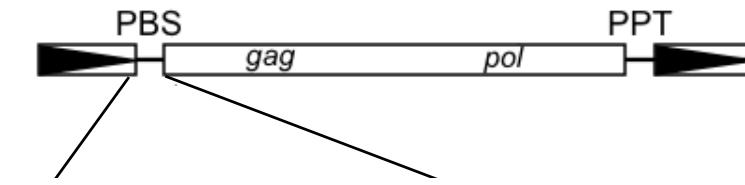
miR845 activity in Columbia (Col) but not Landsberg (Ler) pollen



miR845b
target site

Filipe Borges

PBS (primer binding site) of LTR retrotransposons is targeted by miR845 in pollen

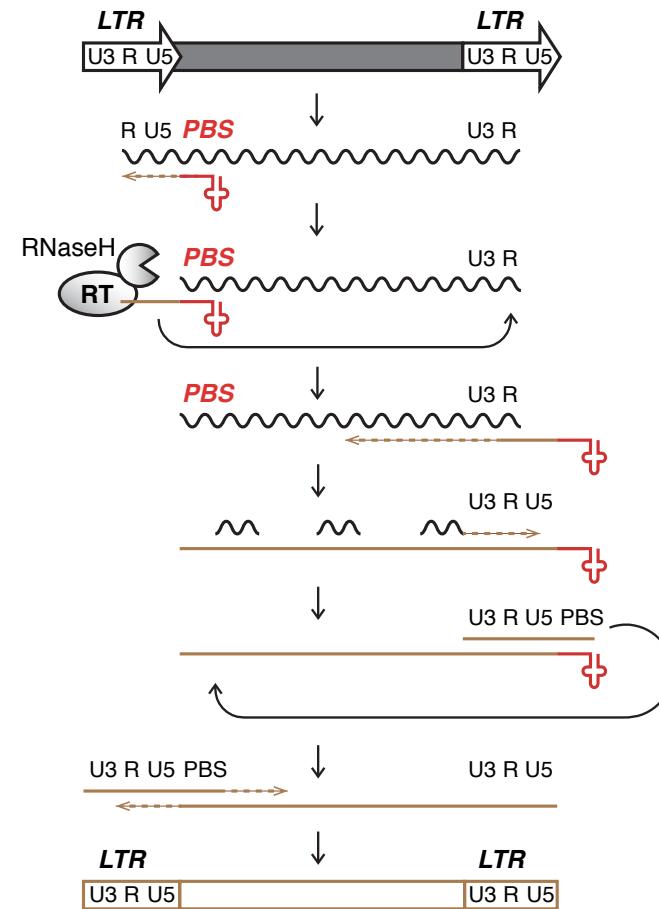
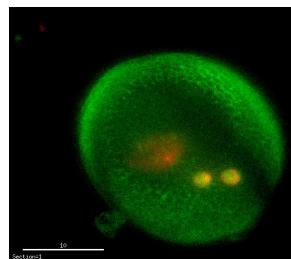


5'-TTTCAATTGGTATCAGAGCGATTACGG-3' AtGP1
|||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

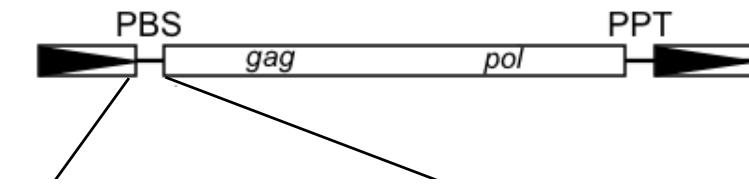
5'-TTTCAAGTTGGTATCAGAGCATGCTTGG-3' AtGP2
|||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

5'-CATTAACATGGTATCAGAGCTTATCGAT-3' AtCopia36
|||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

5'-TTATCACATGGTATCAGAGCAGTGATTCT AtCopia21
|||||
3'-GTAGTTAACCATAGTCTCGGC-5' miR845a



PBS (primer binding site) of LTR retrotransposons is targeted by miR845 in pollen

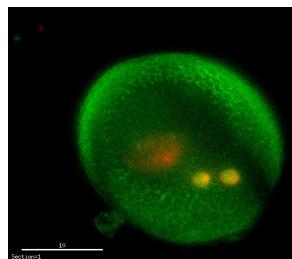


5'-TTTCAATTGGTATCAGAGCGATTACGG-3' AtGP1
||||| |||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

5'-TTTCAAGTTGGTATCAGAGCATGCTTGG-3' AtGP2
||||| |||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

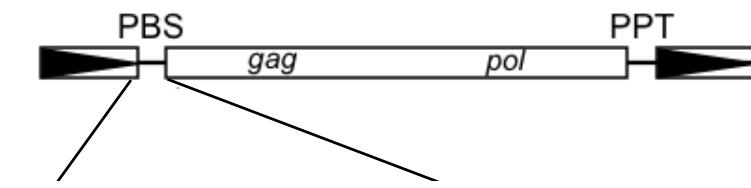
5'-CATTAACATGGTATCAGAGCTTATCGAT-3' AtCopia36
||| ||| |||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

5'-TTATCACATGGTATCAGAGCAGTGATTCT AtCopia21
||||| ||||| |||
3'-GTAGTTAACCATAGTCTCGGC-5' miR845a



- Almost all COPIA and GYSPY retrotransposons use tRNA^{iMet} primer in *Arabidopsis*
- PBS are invariable providing uniquely sensitive targets
- In mouse, PBS of ERV are targeted by 3'tRF (tRNA fragments)

PBS (primer binding site) of LTR retrotransposons is targeted by miR845 in pollen

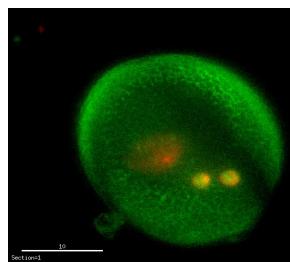
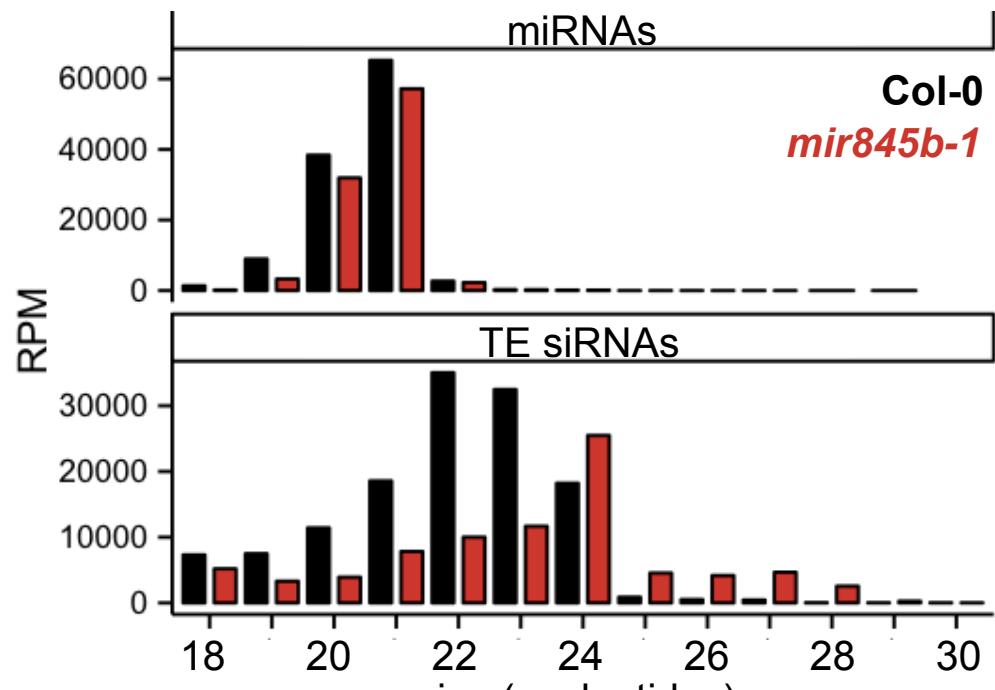


5'-TTTCAATTGGTATCAGAGCGATTACGG-3' AtGP1
|||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

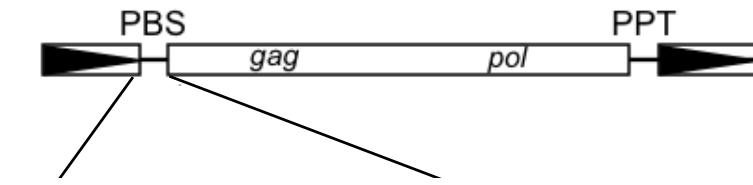
5'-TTTCAAGTTGGTATCAGAGCATGCTTGG-3' AtGP2
|||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

5'-CATTAACATGGTATCAGAGCTTATCGAT-3' AtCopia36
|||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

5'-TTATCACATGGTATCAGAGCAGTGATTCT AtCopia21
|||||
3'-GTAGTTAACCATAGTCTCGGC-5' miR845a



PBS (primer binding site) of LTR retrotransposons is targeted by miR845 in pollen

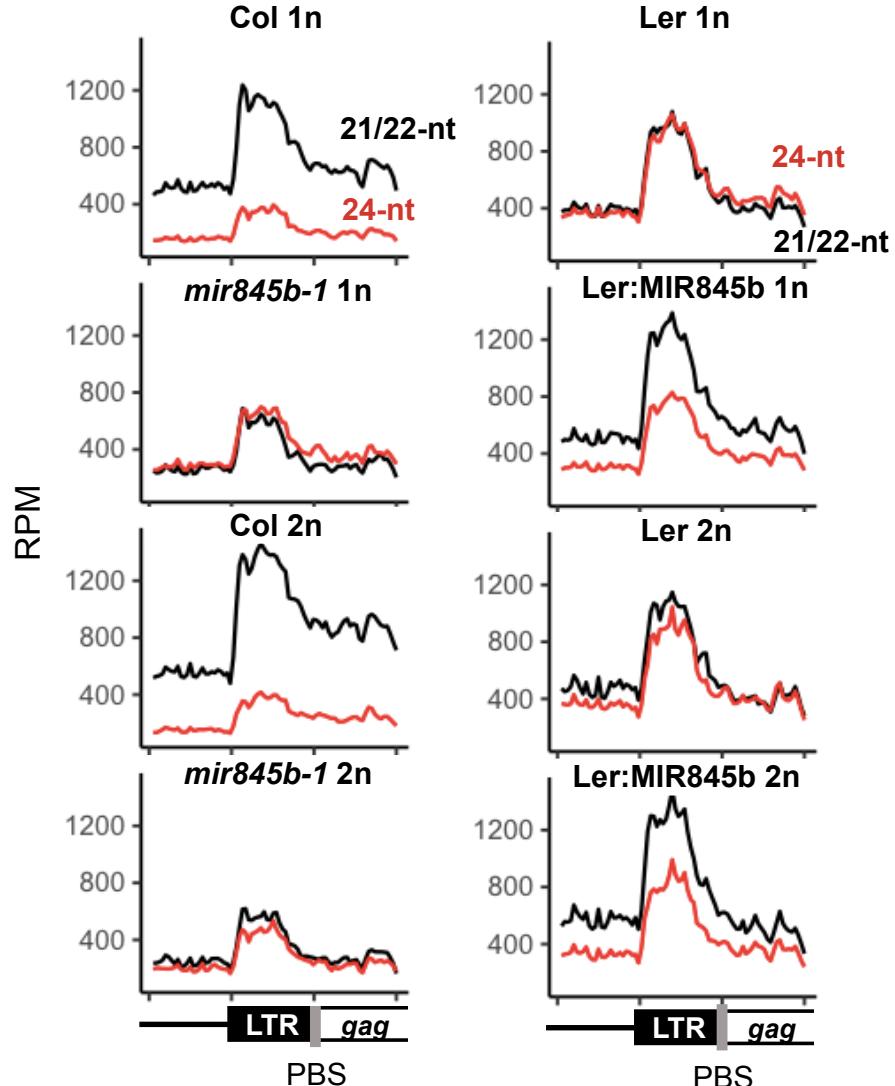
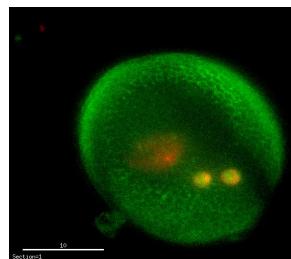


5'-TTTCAATTGGTATCAGAGCGATTACGG-3' AtGP1
|||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

5'-TTTCAAGTTGGTATCAGAGCATGCTTGG-3' AtGP2
|||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

5'-CATTAACATGGTATCAGAGCTTATCGAT-3' AtCopia36
|||||
3'-GTAGTTAACCATAGTCTCGCT-5' miR845b

5'-TTATCACATGGTATCAGAGCAGTGATTCT AtCopia21
|||||
3'-GTAGTTAACCATAGTCTCGGC-5' miR845a



Albert Blaskeeslee and the triploid block



the corresponding

CSHL
1921

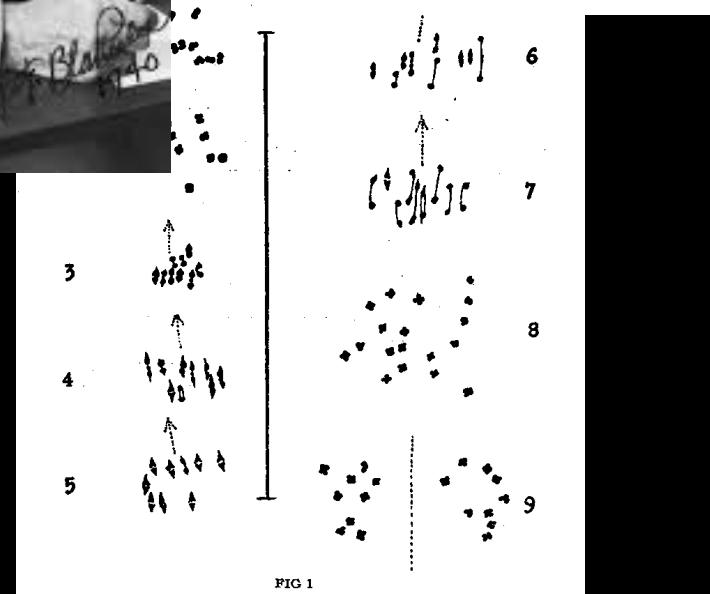


TABLE II

TYPES OF CHROMOSOMAL DUPLICATION, GAMETIC AND SOMATIC FORMULAE FOR PLANTS HETEROZYGOUS FOR FACTOR PAIR A AND a AND RATIOS OBTAINED WHEN SUCH PLANTS ARE SELFED, TOGETHER WITH DIAGRAMS ILLUSTRATING THE CHROMOSOMAL CONDITION IN SOMATIC CELLS

No. of Extra Chromosomes in Set	No. of Sets Affected	Gametic Formula	Selfed Ratios	Somatic Formula	Somatic Diagram
2	12	AA + Aa AA + 4Aa + aa Aa + aa (12 + 12)	1A : 0a 35A : 1a 3A : 1a	AAAA AAaa Aaaa (12 + 12) + (12 + 12)	R POSSIBLE ION. ¹
1	1	2A + a + AA + 2Aa A + 2a + 2Aa + aa 12, (12 + 1)	NOR. 8A : 1a MUT. 9A : 0a NOR. 5A : 4a MUT. 7A : 2a	{ AAA { Aaa (12 + 12) + 1	ION
1	12			(12 + 12) + 12	

No. of chromosomes	0	1	2	3	4	5	6	7	8	9	10	11	12
Frequencies	1	12	66	220	495	792	924	792	495	220	66	12	1

R POSSIBLE ION.¹

saw the rise of cted biological in the trend of The mutation mrose, and the ettled the date s. The studies ed a basis for investigations. ssociated with ation than was

at first imagined, the evening primrose and the theory of mutation connected with it are by many considered to furnish an example of a valuable theory founded upon incorrect interpretations. The belief is growing that most of the new forms which have appeared in cultures of the *Oenotheras* are not mutations at all and that the evening primroses, as an abnormal group of plants, are not to be seriously considered as representative of the processes of evolution in normal forms.

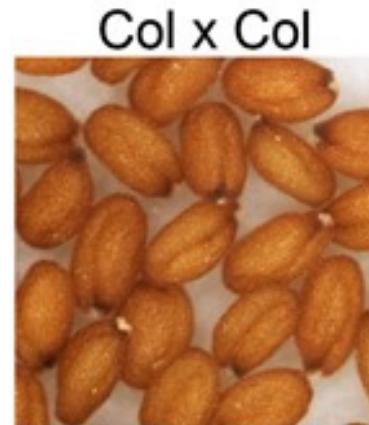
In the short time at my disposal, I wish to outline some recent findings in the jimson weed (*Datura Stramonium*) which it is hoped may throw incidentally some light on the more highly involved phenomena in the *Oenotheras*, and which may serve as a basis of a brief discussion of their possible evolutionary significance.

The jimson weed is not supplied with a wide range of obvious Mendelian characters. The early studies of

¹ A paper presented before the American Society of Naturalists at the Chicago meeting, December 30, 1920.

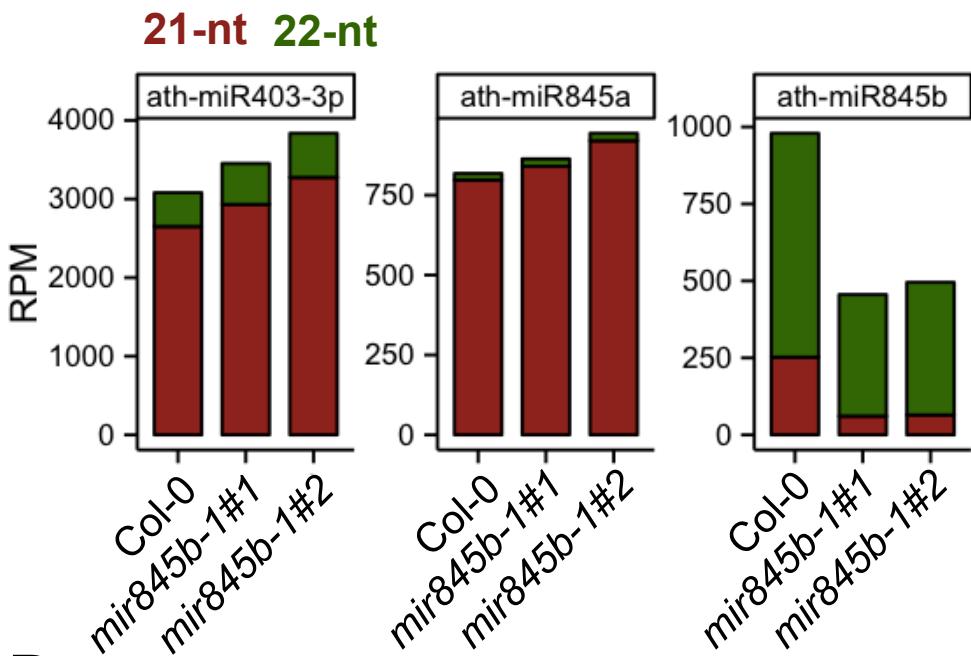
Lansberg is defective in the “triploid block”: dose-dependent lethality

- Intraspecific polyploidy common in plants, but triploid meiosis strongly detrimental
- In most flowering plants, triploid seeds collapse due to imprinting defects (“endosperm balance”)
- *osd1* mutants have diploid pollen, and *osd1* seed collapse in Columbia, but not in Landsberg.

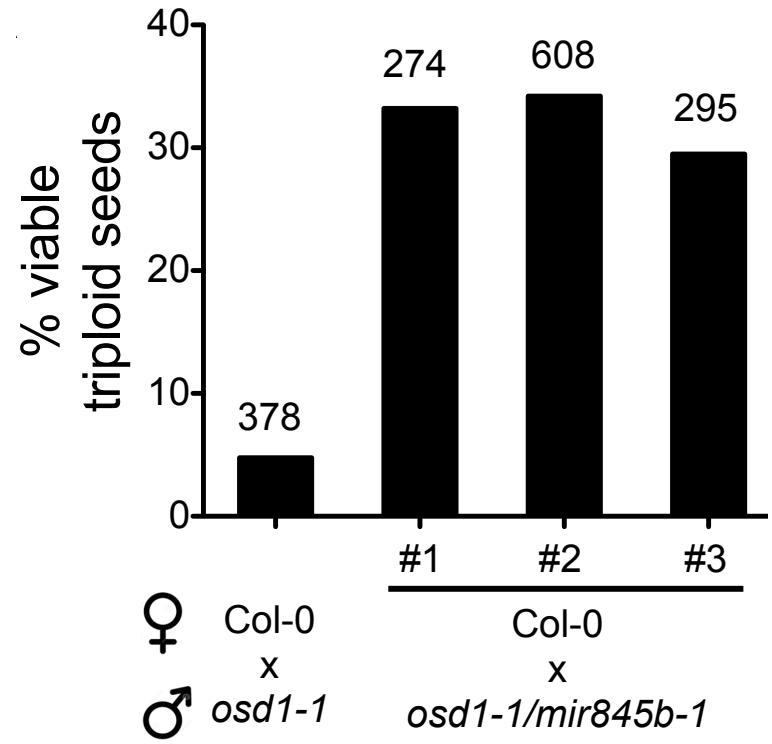


miR845b mutants relax the triploid block

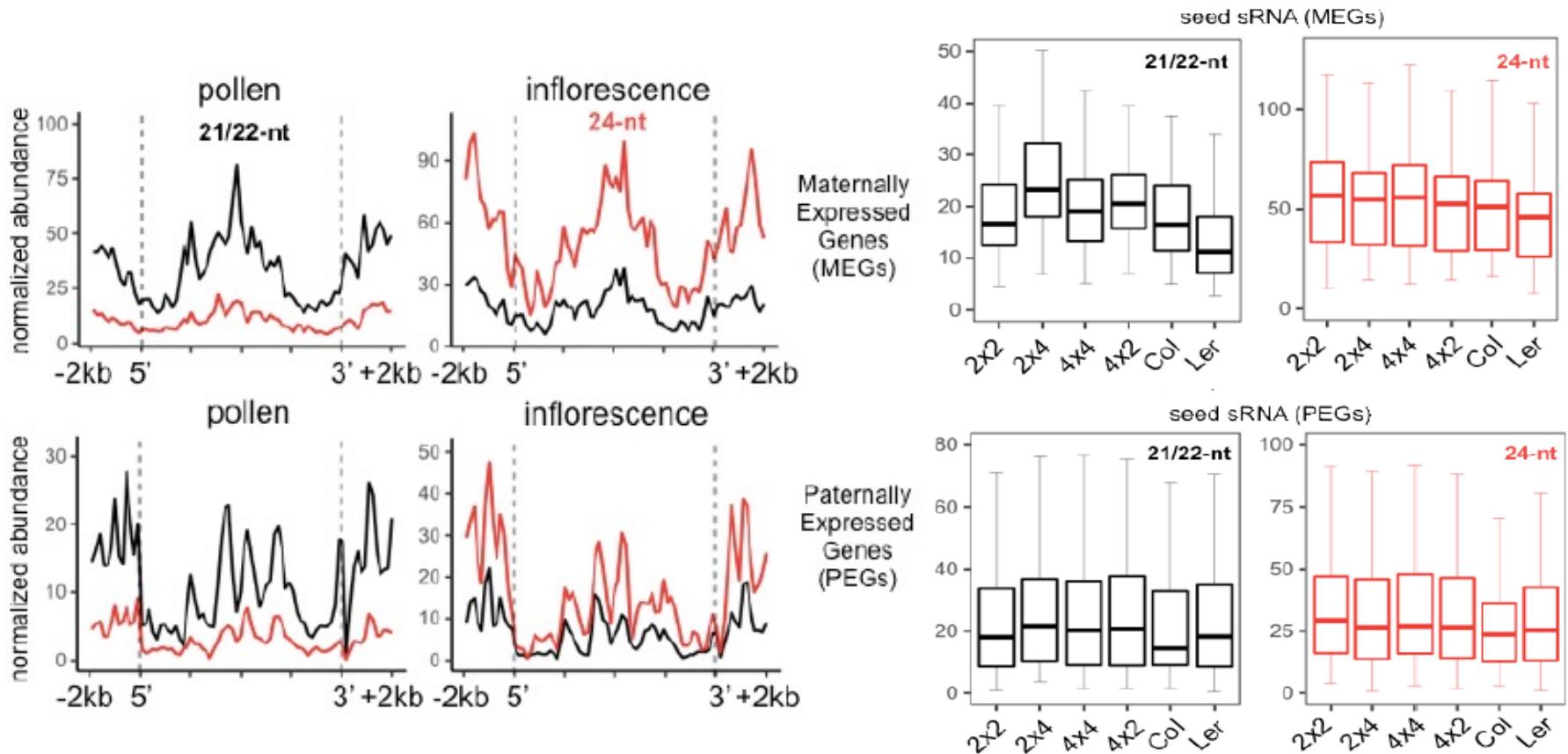
miRNA abundance



Seed viability



easiRNAs from pollen target imprinted genes in seeds



Summary (Part I)

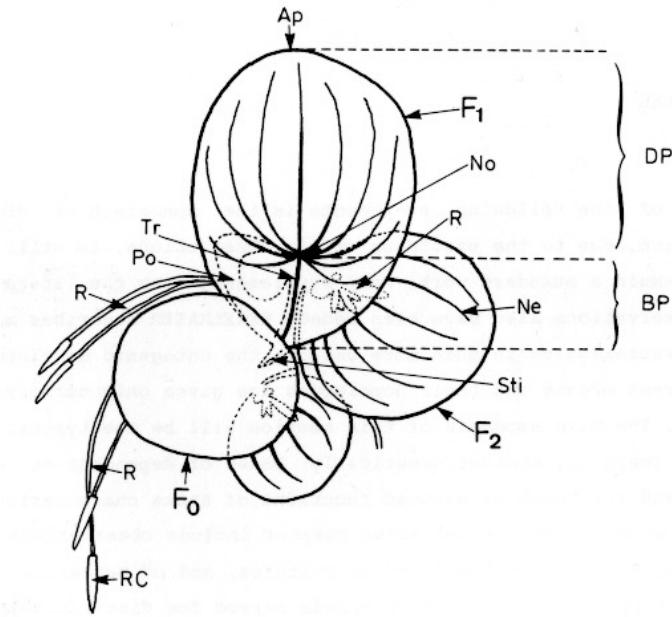
- Reprogramming activates transposons in plant and animal germlines
- Retrotransposon tRNA primer binding site (PBS) is targeted by miR845 in pollen (and tRNA fragments in mouse) that triggers easiRNA in sperm
- easiRNA control retrotransposons in DNA methylation mutants via RNAi and histone H3K9me2
- easiRNA in sperm are required for the triploid block (dosage response) in the seed

Duckweeds (*Lemnaceae*): clonal reproduction without meiosis

<i>Liopsida</i>	class
<i>Aridae</i>	subclass
<i>Arales</i>	order
<i>Araceae</i>	family



Wolfia microscopica



Spirodela polyrhiza
(Landolt, 1986)

- Reduced morphology
 - Fronds (leaf-like structures)
 - Meristem-like stem cell “pocket”
- Clonal reproduction
 - 12-48 hours duplication by budding
 - Limited flowering
 - 10cm² to 1 hectare in 50 days

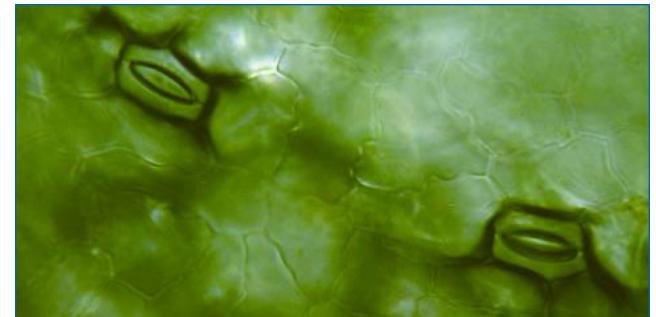
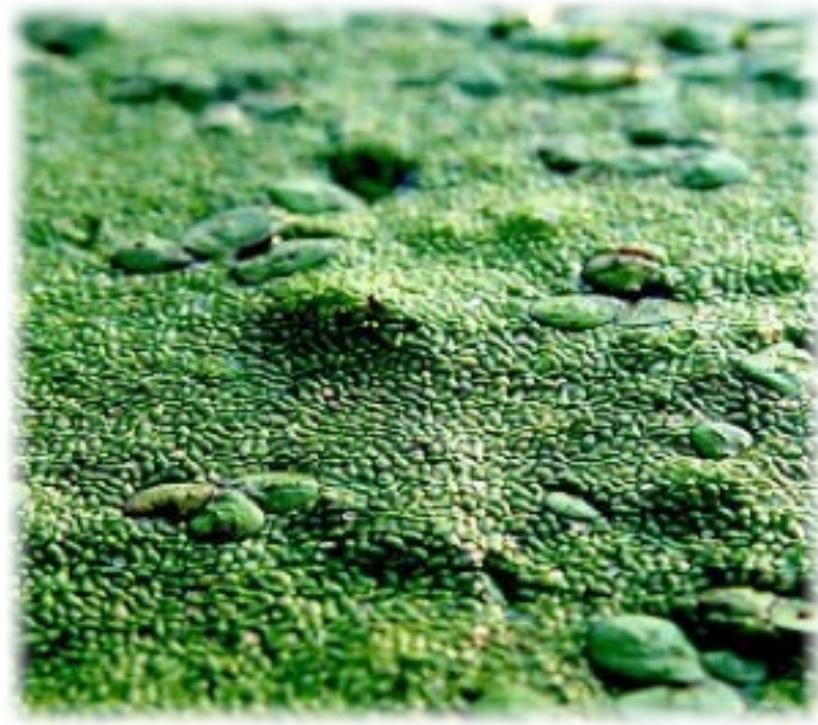
Growing duckweed (*Lemna gibba*)

in constant light conditions



Lemnaceae as an climate crop

- Duckweeds can reproduce clonally by budding without meiosis
- Excellent metabolic profile:
 - Growth rate (12-48h doubling time)
 - Starch 12-50%
 - Protein 11-40%
 - Lipid 2-9%
 - Low lignin content (5%)
 - Open stomata, unlimited access to gaseous CO₂.



The Duckweeds

Liliopsida

Aridae

Arales

Araceae

class
subclass
order
family



Lemna

aequinoctialis punctata

disperma

gibba

japonica

minor

minuta

obscura

perpusilla

tenera

trisulca

turionifera

valdiviana

yungensis

Landoltia

punctata

Spirodela

intermedia

polyrhiza

Wolffia

angusta

arrhiza

australiana

borealis

brasiliensis

columbiana

cylindracea

elongata

globosa

microscopica

neglecta

Wolffiella

caudata

denticulata

gladiata

hyalina

lingulata

neotropica

oblonga

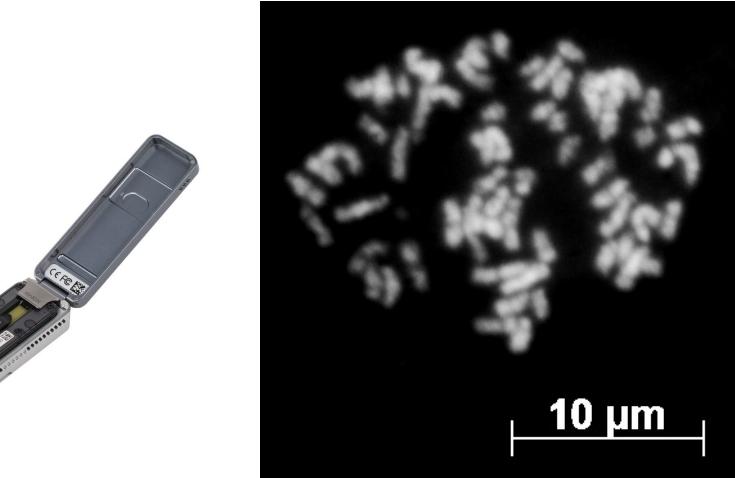
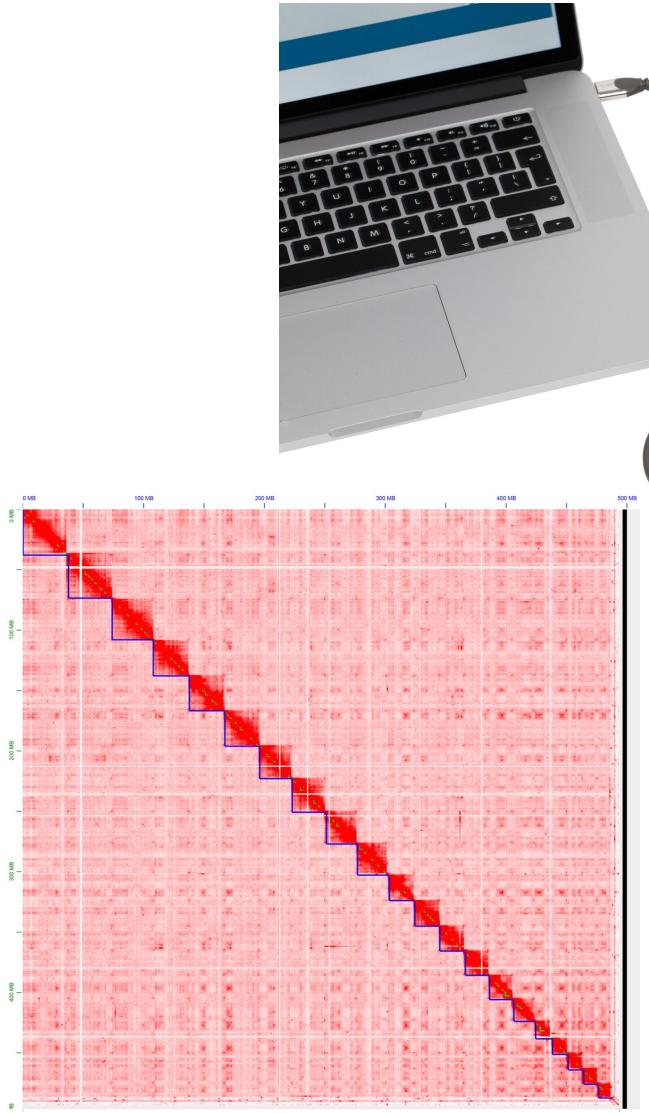
repanda

rotunda

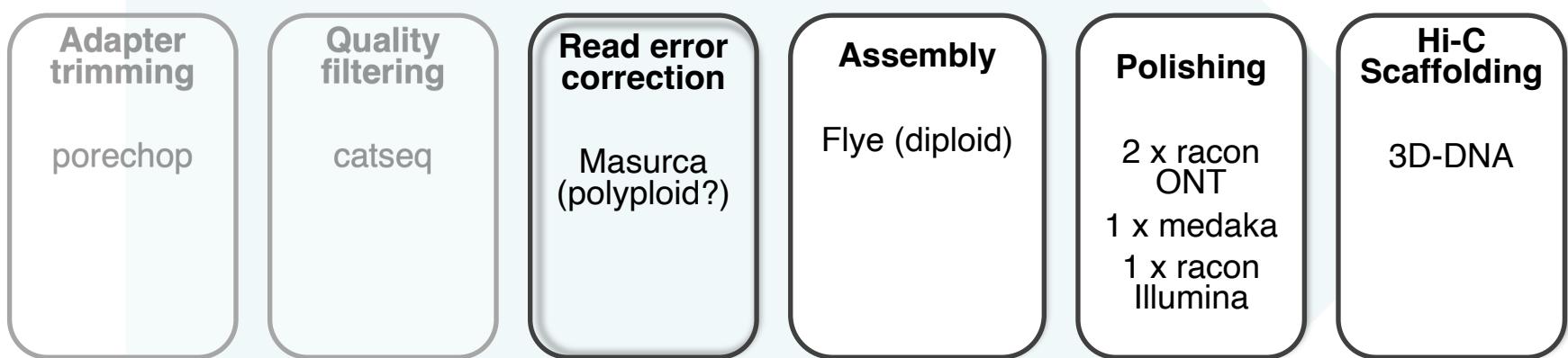
weltwischii



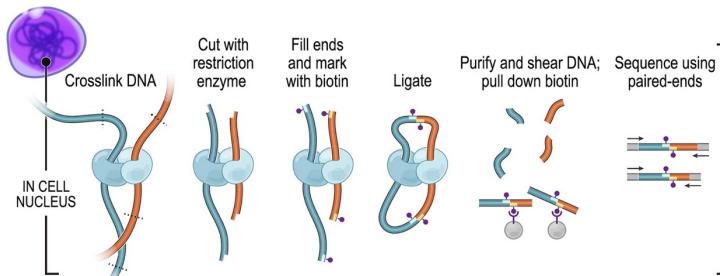
Deciphering Duckweed Genomes



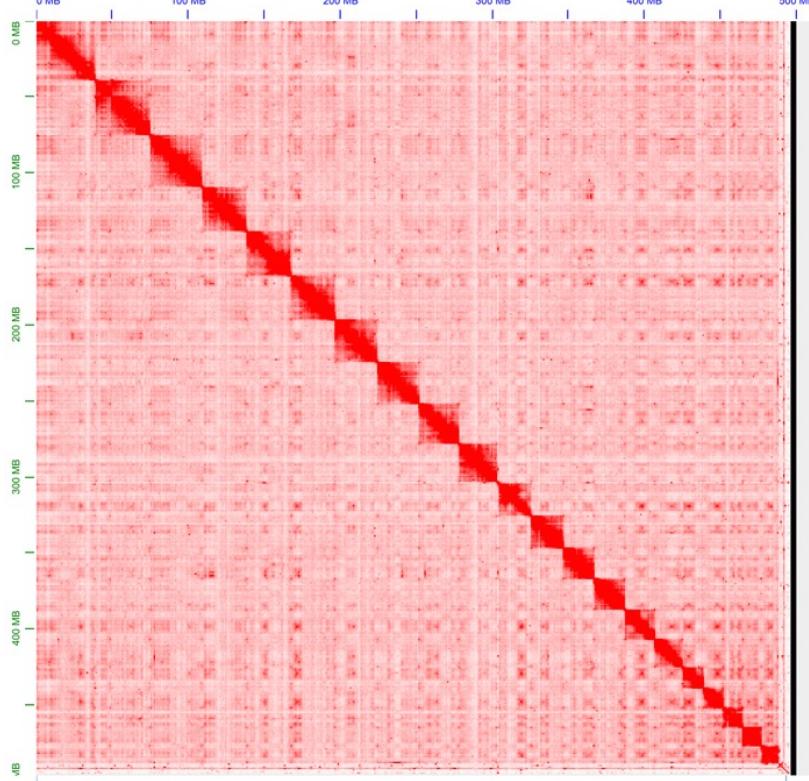
Assembly Overview (new school)



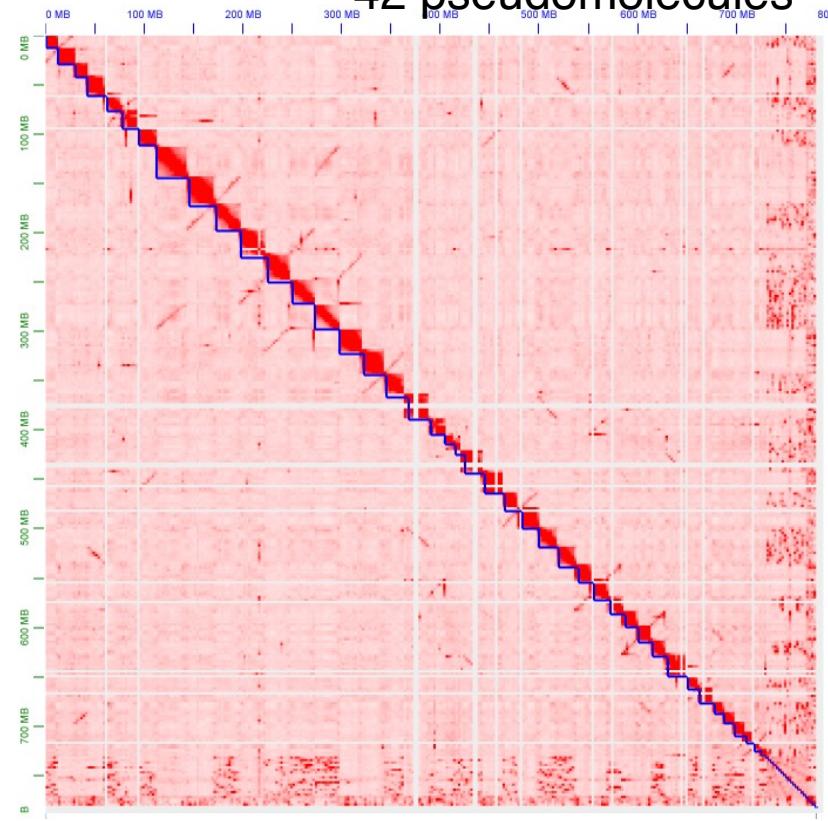
Hi-C Libraries order and orient long contigs into chromosomes



L. gibba 7742a
21 pseudomolecules



L. japonica 8627
42 pseudomolecules



Library prep: Dovetail Genomics Mapping and scaffolding: Juicebox

Lemnaceae genome project

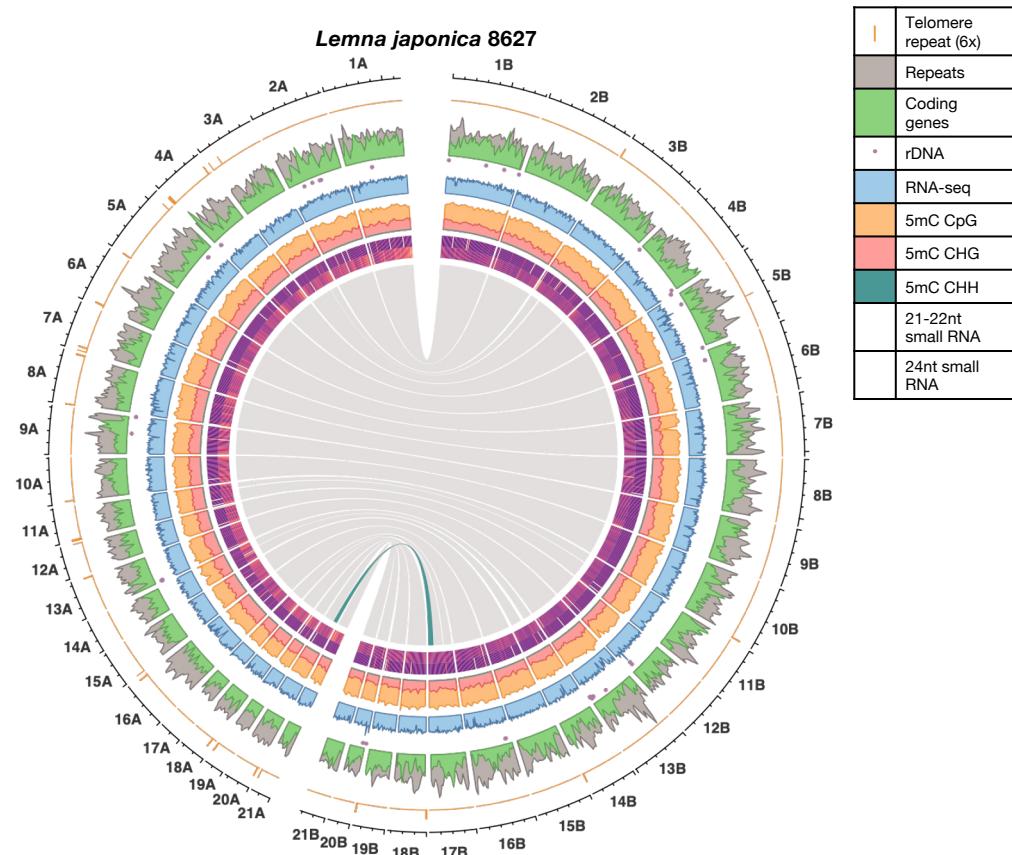


	<i>S. polyrhiza</i> greater duckweed Sp9509 v3 JCVI	<i>W. australiana</i> Australian watermeal Wa8730 v3 CSHL	<i>L. gibba</i> swollen duckweed Lg7742a v3 CSHL	<i>L. minor</i> lesser duckweed Lm9252 v1 CSHL	<i>L. minor</i> lesser duckweed Lm7210 v1 CSHL	<i>L. turionifera</i> turion duckweed Lt9434 v1 CSHL	<i>L. japonica</i> L. min. x L. tur. (3n = MMT) Lj8627 v3 CSHL	<i>L. japonica</i> L. min. x L. tur. (3n = TTM) Lj7182 v1 CSHL	<i>L. japonica</i> L. min. x L. tur. (3n = TTM) Lj9241 v1 CSHL
Genomes									
span (Mbp)	139	355	491	354	363	410	772	764	771
Repeats									
all interspersed (%) ^A	12	48	71	57	59	61	62	62	63
LTR-RT (%) ^A	7	23	57	43	42	43	43	44	42
intact LTR-RT/Mpb (#) ^A	0.1	0.6	2.0	1.2	1.7	1.3	1.3	1.3	1.6
LTR solo:intact ratio ^A	12	15	47	64	58	61	64	62	63
Protein coding genes									
Mikado genes ^B	16,827	13,095	18,427	17,138	16,701	16,613	31,990	31,761	31,746
GeMoMa genes ^C	20,389	22,416	24,717	22,754	22,362	22,862	40,651	40,523	40,203
Augustus CGP genes ^D	17,841	13,175	17,734	20,697	20,838	18,325	37,544	38,104	38,451
Combined & filtered ^{E,F,G}	20,382	20,902	24,257	22,247	21,683	22,019	40,747	40,504	40,388
PASA Updated ^H	20,342	20,882	24,222	22,205	21,649	21,988	40,692	40,444	40,336
Non-coding gene loci									
transcribed ncRNA ^B	647	865	1,852	1,252	888	976	2,100	2,095	2,107
snoRNA/snRNA ^I	541	864	1,141	751	785	1,210	2,017	2,016	2,286
tRNA ^I	237	623	470	253	260	278	486	482	479
miRNA loci ^I	59	63	74	-	-	-	108	-	-
(A) EDTA/LTR_retriever	(B) Mikado	(C) GeMoMa	(D) Augustus	(E) AGAT	(F) MAKER_P	(G) TEsorter	(H) PASA	(I) Infernal	(J) ShortStack

Evan Ernst (collaboration between Martienssen, Michael and Lam labs)

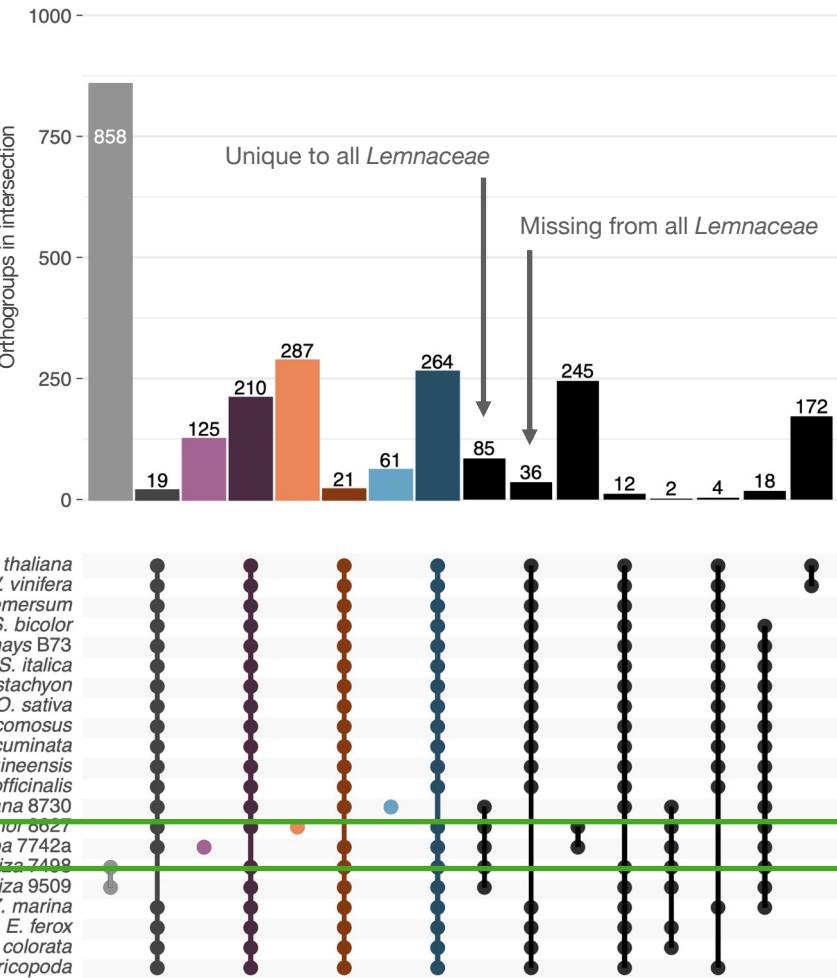
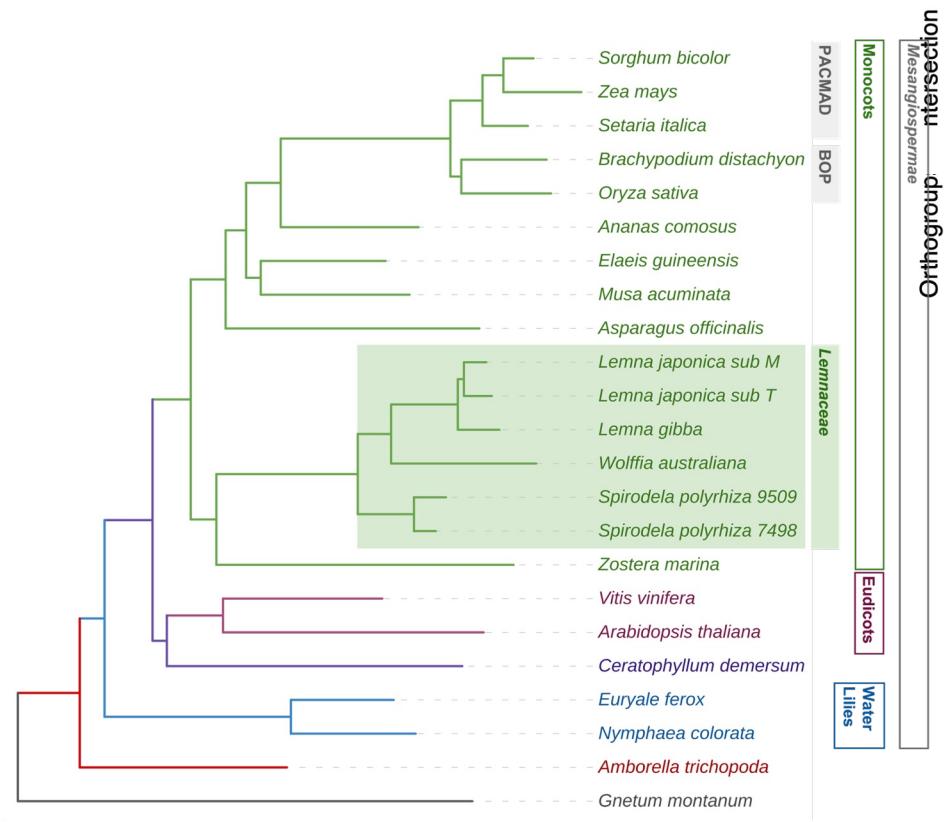
L. japonica are triploid hybrids (*L. minor* x *L. turionifera*)

- Nanopore long-read sequencing with full-length chromosome scaffolding by Hi-C gives two full chromosome sets for the clone labeled *L. minor* 8627
- gDNA read coverage indicates a 2x/1x ratio – a triploid signature
- AFLP confirm the existence of widespread interspecific *L. minor* x *turionifera* hybrids and the *L. japonica* identity of 8627 (Braglia et al. (*Frontiers in Plant Science*, 2021))
- Whole-genome sequencing of additional *L. japonica*, *L. minor* and *L. turionifera* accessions revealed that triploid formation occurs in both directions.



Homeologous chromosomes of *L. japonica* 8627. Central ribbons show synteny. A single large-scale translocation between *L. minor* and *L. turionifera* subgenomes is highlighted in teal.

Phylogenomic analysis



Evan Ernst

GO Term Enrichment: Floating aquatic plants

Unique orthogroups



Missing conserved orthogroups



area $\sim -\log_{10}(p\text{-value})$ of enrichment

Gene silencing in *Lemnaceae*: missing orthogroups and genes

Missing orthogroups (36)



Gene silencing in *Lemnaceae*: missing orthogroups and genes



Gene silencing in *Lemnaceae*: missing orthogroups and genes

Missing HOGs

GO:0031047 gene silencing by RNA

CLSY1/2
DCL2
IDM1(ROS4)
RTL1/2/3

GO:0009615 response to virus

AGO2/3

Lemnaceae have one *de novo* DNA methyltransferase (*DRM2*) capable of asymmetric CHH methylation (RNA directed DNA methylation)

But they have lost *CMT2* (as has maize, but not rice) which is the main CHH methyltransferase

All Lemnaceae have retained *DCL3* (required for 24nt siRNA)

AGO6
CMT2
DRM3

MORC1/2/3 [*Lemna* & *Wolffia*]
MORC6 [*Lemna* & *Wolffia*]
SHH1

But they have lost *AGO 2/3/6* and *DCL2* (required for some 21/22nt siRNA)

Gene silencing in *Lemnaceae*: small RNA from fronds

Missing HOGs

GO:0031047 gene silencing by RNA

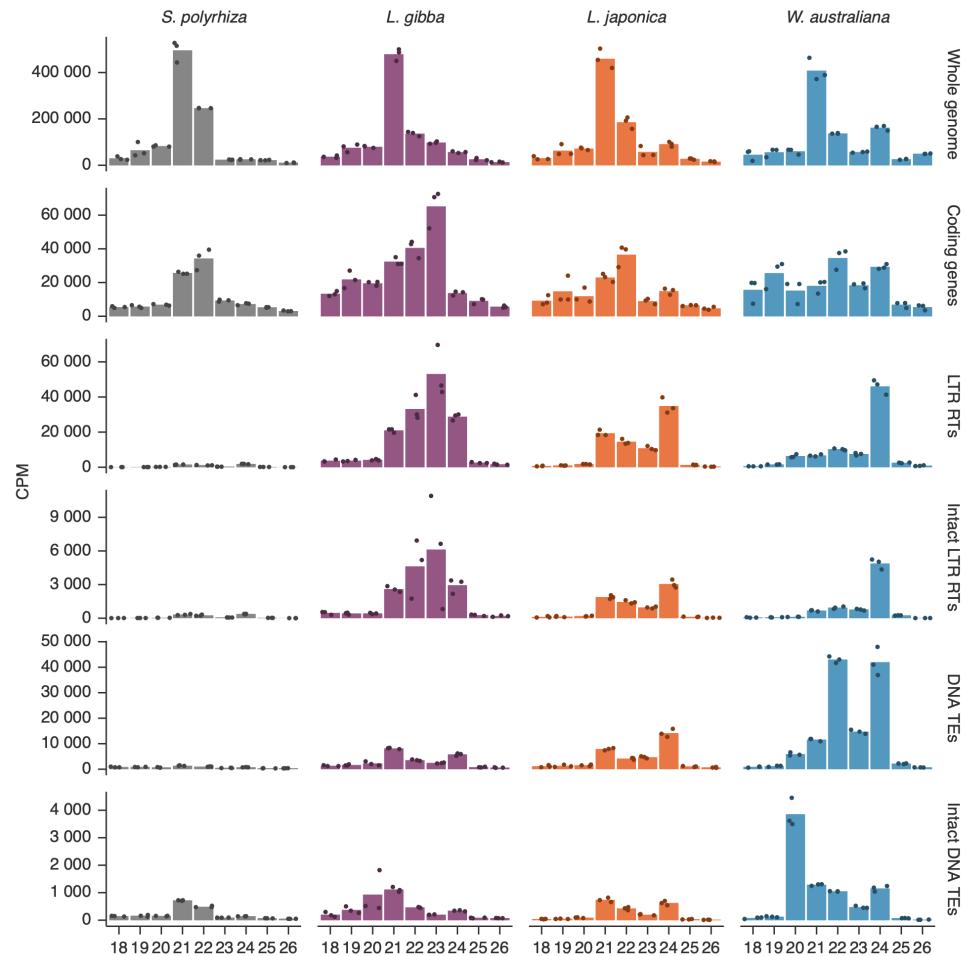
CLSY1/2
DCL2
IDM1(ROS4)
RTL1/2/3

GO:0009615 response to virus

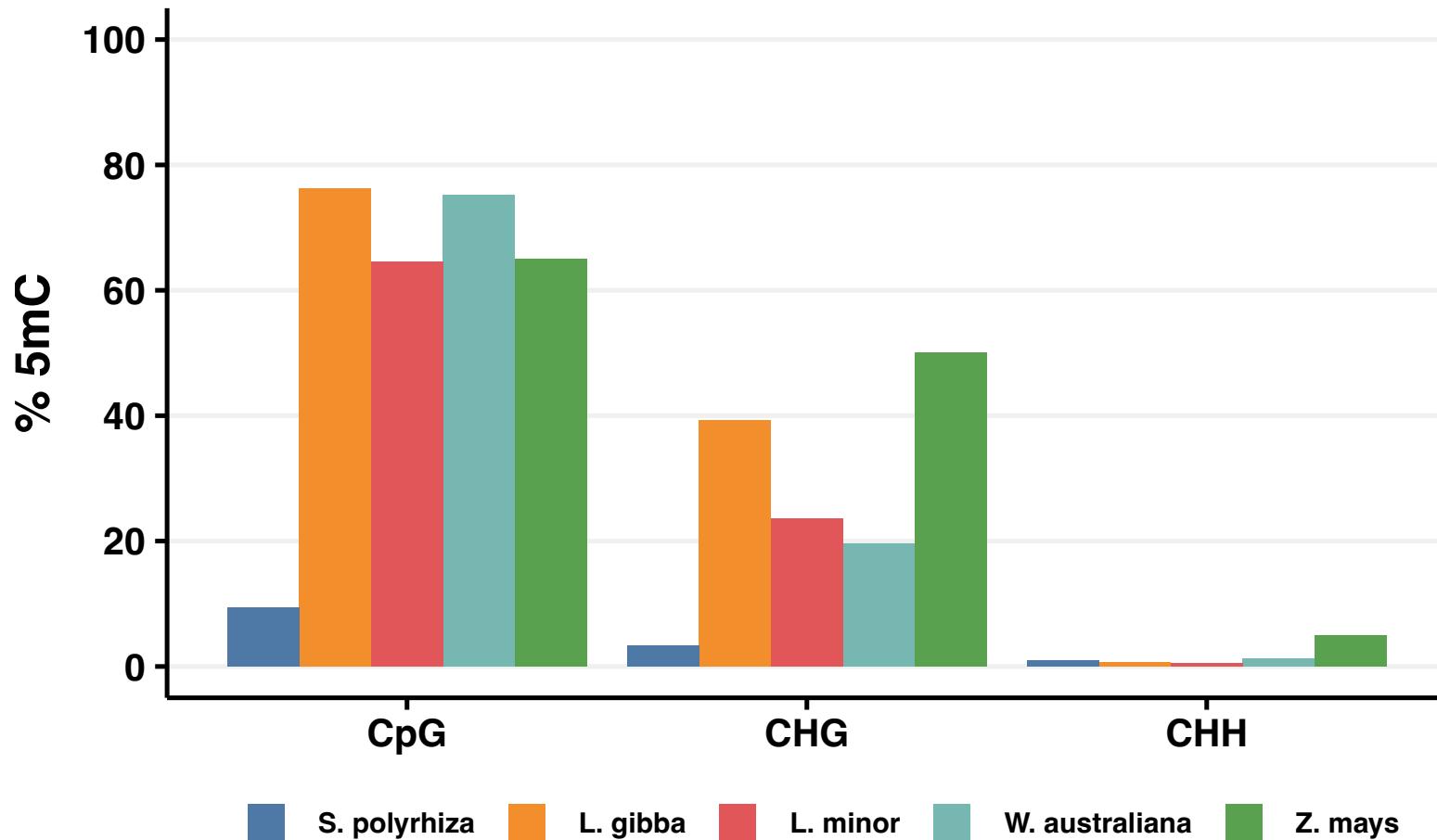
AGO2/3

Missing genes

AGO6
CMT2
DRM3
MORC1/2/3 [*Lemna* & *Wolffia*]
MORC6 [*Lemna* & *Wolffia*]
SHH1



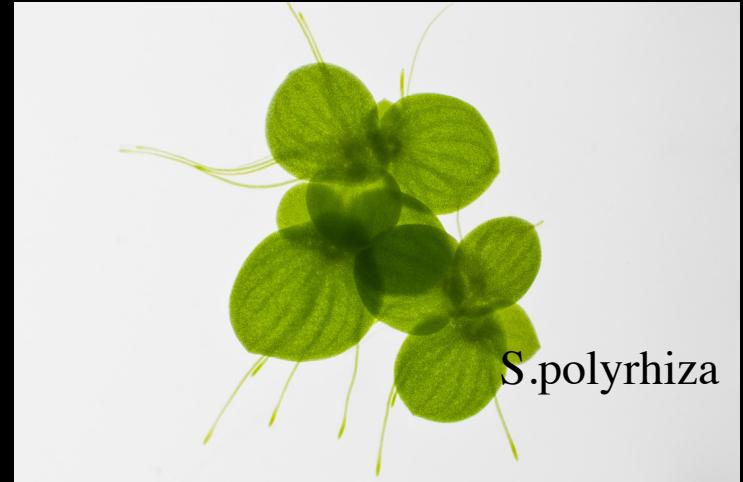
Lemnaceae have very low asymmetric (CHH) DNA methylation



Loss of fertility is accompanied by loss of transposons (and methylation)



L. minor



S. polyrhiza



L. gibba



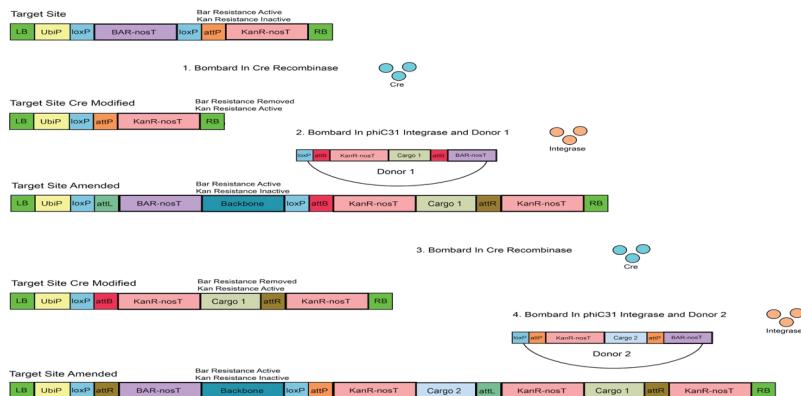
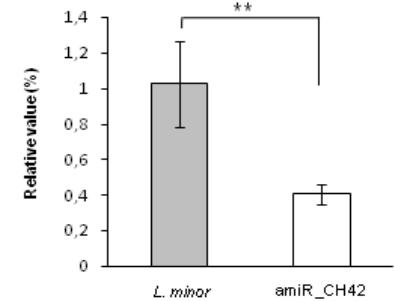
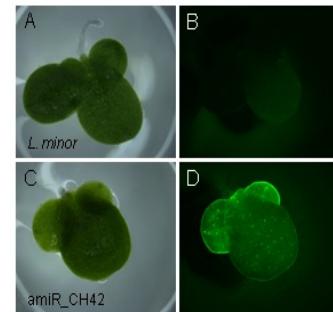
W. australiana

Reproductive strategy and transposon evolution

- Genetica. 1992;86(1-3):269-74.
- **Evolutionary dynamics of transposable elements in prokaryotes and eukaryotes.**
- Hickey DA¹.
- Author information
- 1Department of Biology, University of Ottawa, Ontario, Canada.
- **Abstract**
- This paper summarizes some recent theories about the evolution of transposable genetic elements in outbreeding, sexual eukaryotic organisms. The evolutionary possibilities available to self-replicating transposable elements are shown to vary depending on the reproductive biology of the host genome. This effect can be used to explain, in part, the differences in abundance of transposable elements between prokaryotes and eukaryotes. It is argued that the pattern of sexual outbreeding seen in mammals and plants is especially favorable to the spread of transposons. Moreover, because transposon spread is facilitated by zygote formation, the evolutionary origin of sexual conjugation may have been due to selection on transposon-encoded genes. Finally, evidence is also presented that introns could have originated as transposable genetic elements.

Synthetic Biology tools for engineering *Lemna*

- Rapid and efficient agrobacterium transformation of clonal fronds (Cantó-Pastor et al., *Plant Biol* 2014)
- Artificial miRNA for dominant gene knock down (Cantó-Pastor et al., *Plant Biol* 2014)
- CRISPR/Cas9 homozygous loss of function (Liu et al., *Plant Biotech.*, 2019)
- Transgene stacking systems
- Under development:
 - CRISPRi and CRISPRa libraries
 - artificial chromosomes



Evan Ernst, Alex Canto Pastor, Almudena Molla Morales (collaboration between Martienssen, Shanklin and Birchler labs)

Summary (Part II)

Lemnaceae genomes have been sequenced to very high accuracy and contiguity (ONT, Illumina and hi-C technology)

Gene content is 50-60% of other angiosperms, reflecting extreme reduction in morphological and physiological complexity

Sterile polyploid interspecific hybrids are common, in part due to the loss of gene silencing and the triploid block

Key developmental traits (eg. reduced roots and flowers) and physiological traits (eg. Stomatal opening in high CO₂) are associated with evolutionary loss and divergence of regulatory genes

Genome-guided development of a synthetic biology toolkit enables pathway engineering



COLD SPRING HARBOR LABORATORY



Former lab members:

Filipe Borges
Rowan Herridge
Jean-Sebastien Parent
Jakub Dolata
Atsushi Shimada

Current Lab members

Cris deSantis Alves
Seung-Cho Lee
Florencia Campatella
Jon Cahn
Ben Berube
Evan Ernst
Cristian Mateo Elizalde
Joe Simorowski
Uma Ramu
Mike Regulski
Liam Spielmann

Collaborators

Dick McCombie
Jerry Kermicle
Daniel Grimanelli
Claudia Köhler
Blake Meyers
Leemor Joshua-Tor

Acknowledgements



Office of
Science



Martienssen Lab

Rob Martienssen

Will Dahl

Ramu Umamaheswari

Joe Simorowski

Seung Cho Lee

Michael Regulski

Former Members:

Almudena Molla-Morales

Alex Cantó Pastor

BROOKHAVEN
NATIONAL LABORATORY

John Shanklin

Jörg Schwender

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Eric Lam

J. Craig Venter®
INSTITUTE

Todd Michael

Mizzou
University of Missouri

Jim Birchler



Lemnaceae genomes for all

BLAST Results | lemna.org

Not Secure | lemna.org/blast/report/fHiKm

Apps CSHL Bio Dev Web Dev ML R Stats Energy Climate Sounds Nerds laberinto

Dashboard Content Structure Tripal Appearance People Modules Configuration Reports Help Search 0 / 1 Hello admin Log out

Query Information: /tmp/2019Sep12_071223_query.fasta

Search Target: *L. gibba* 7742a 2019v2

Submission Date: Thu, 2019-09-12 07:12

BLAST Command executed: blastn -max_target_seqs 500 -evalue 0.001 -word_size 11 -gapopen 5 -gapextend 2 -penalty -2 -reward 1

Number of Results: 1

Resulting BLAST hits

The following table summarizes the results of your BLAST. Click on a triangle on the left to see the alignment and a visualization of the hit, and click the target name to get more information about the target hit.

#	Query Name (Click for alignment & visualization)	Target Name	E-Value
▲ 1	AtDc13	chr8	1.2066E-7

Hit Visualization

"chr8"

Target

AtDc13

HSP1

Bit Scores

- >= 200
- 80 - 200
- 50 - 80
- 40 - 50
- < 40

The image above shows the relationship between query and target for this particular BLAST hit.

Alignment

HSP 1

Identity= 57/69 (82.61%) , Positive= 57/69 (82.61%) Query Matches 1432 to 1500 Hit Matches = 25591980 to 25592048

Query: 1432 GATGTGGTGAAGAGGGATTCAGGTTCCAGATTGCTCATGCATGGTTGTTTGACCTG 1491

Sbjct: 25591980 GATGTGGGAAAGAGGGACTTCATTGCAAATTGCTCATCGTGATAAGCTTTGACCTG 25592039

Query: 1492 CCCAAACAA 1500

Sbjct: 25592040 CCCAAACAA 25592048