



Searching the repeat element space of plant genomes

Finding the hidden gems that are the active transposable elements contributing to genome evolution

Chris Winefield^{1,4}, Ben Crawford¹, Ting-Hsuan Chen², Darrell Lizamore³
and Peter Waterhouse^{4,5}

¹ Lincoln University

² Plant and Food Research

³ Bragato Research institute

⁴ Australian Research Council Centre of Excellence for Plant Success in Nature and Agriculture

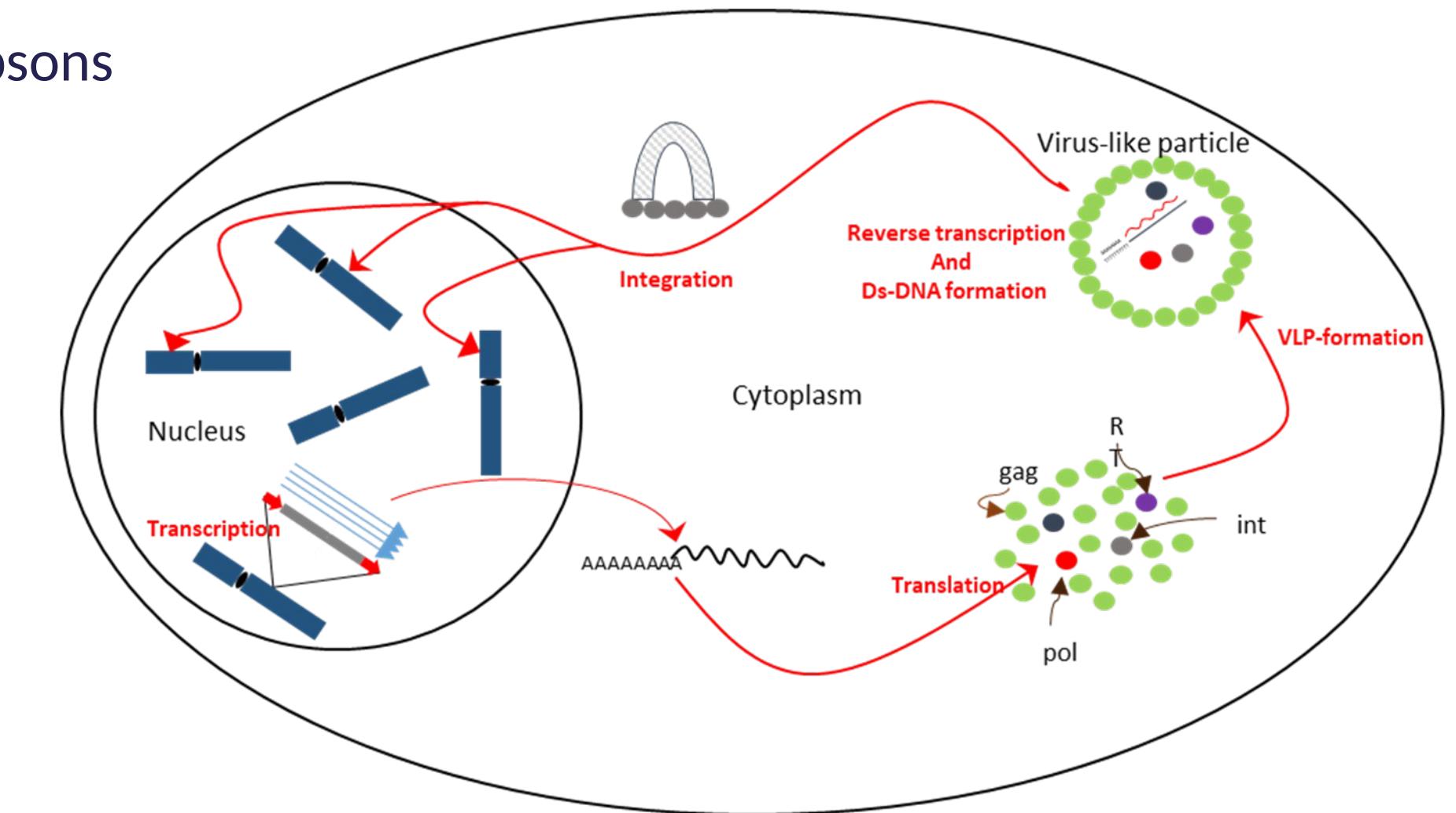
⁵ Queensland University of Technology



Plant transposons

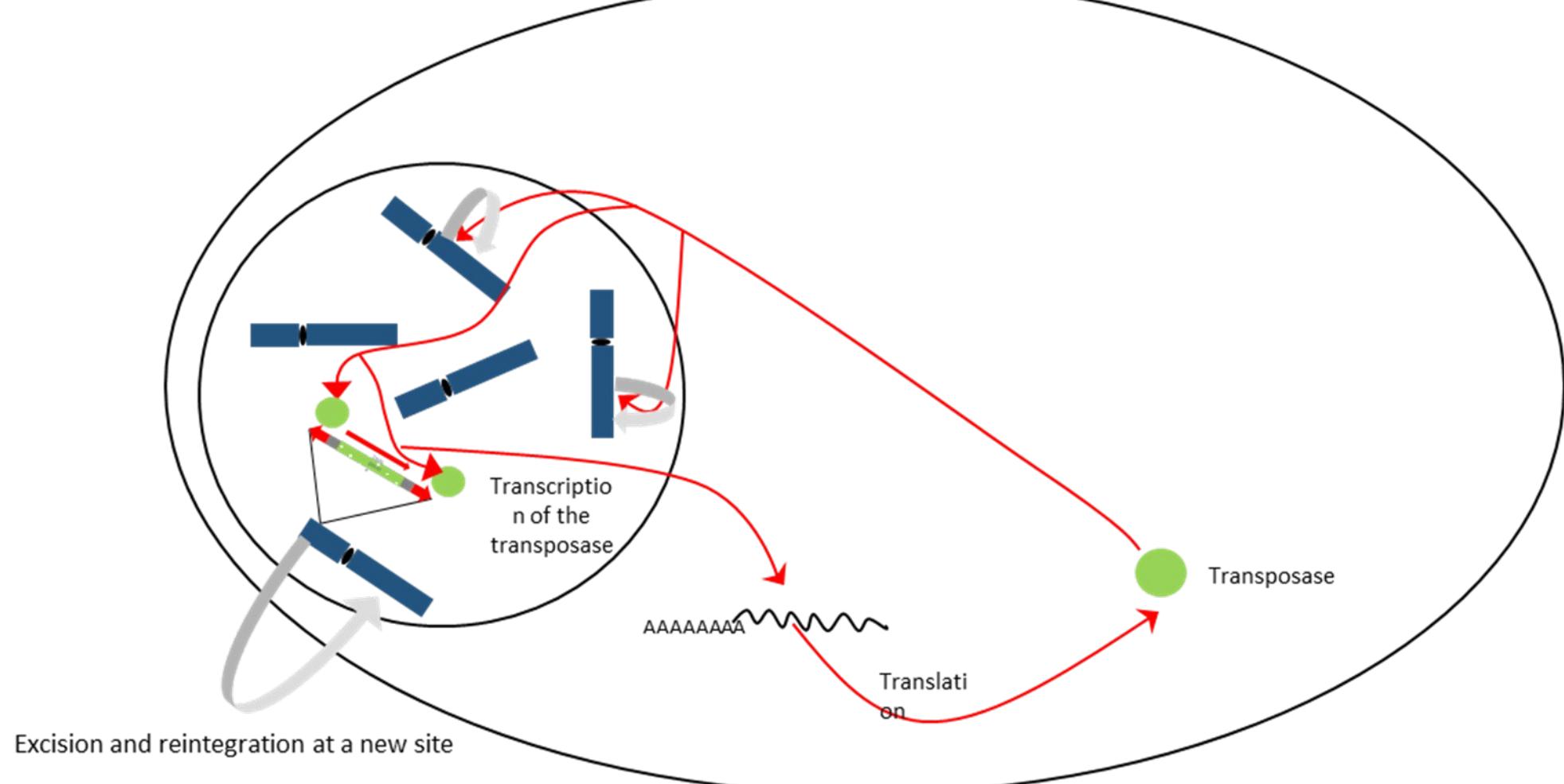
- Transposons are ubiquitous in all genomes
- Between 15 and >80% of plant genomes are comprised of transposable elements
- Two super-families of elements
 - Type 1 retro elements make up the bulk in most genomes
 - Type 2 more “elusive” but likely more important in terms of contribution to phenotypic diversity

Type 1 retrotranspsons



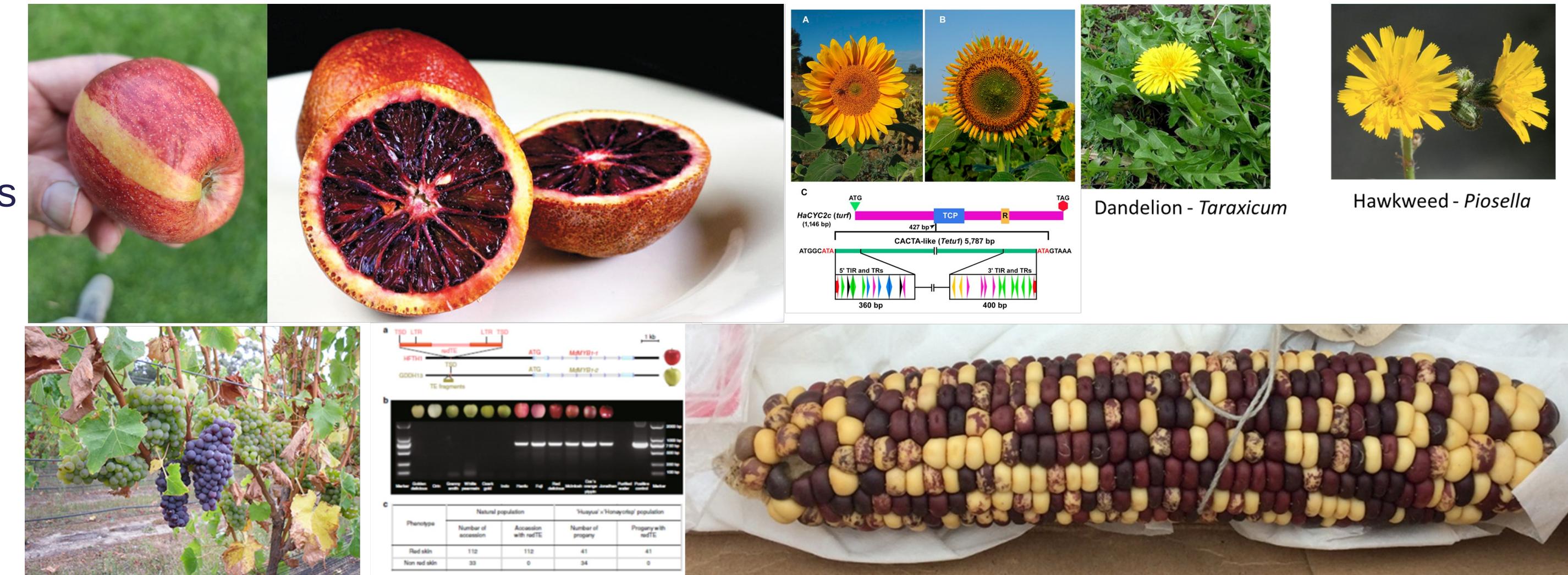
- All elements are actively repressed by a multi-layer epigenetic silencing mechanism
 - Small RNA driven transcript surveillance and destruction
 - Small RNA driven DNA methylation – RNA dependant DNA methylation
 - Alteration of chromatin structure
 - Use of tRF fragments to inhibit reverse transcription

Type 2 DNA-transposons (hAT like)



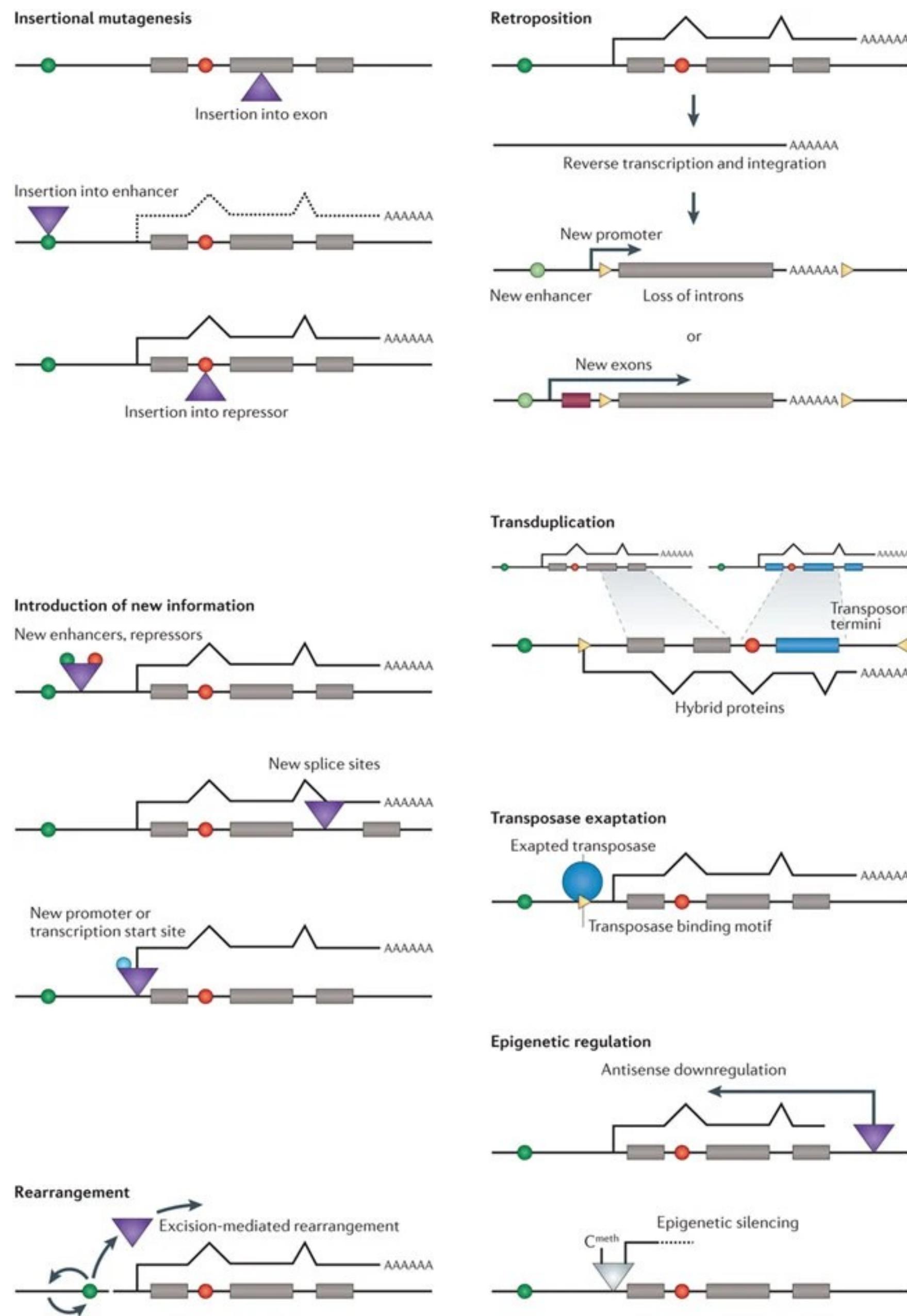
Transposons cause damage.....and new diversity

- Transposons if unchecked are highly damaging to genomes
 - *ddm* mutations in *Arabidopsis* show heightened TE mobilisation
 - After 4-5 generations the progeny are very 'sick'
- However, numbers of phenotypes and physiologies in plants are due, solely or in part, to transposon insertions
 - Vernalization
 - Pigment accumulation in blood orange
 - Dioecious flowering in Poplar (and other species?)
 - Apomixis in *Taraxicum* and *Hieracium*
 - Red grapes vs white grapes
 - RPM-1 in grapes (bunch architecture)
 - Red fleshed apples
 - Cyclodia mutations in sunflower (radial flower formation)
 - Tomato fruit shape
 - Floral variegation in numbers of tissues (flowers and seed coats) in numerous flowering plant species



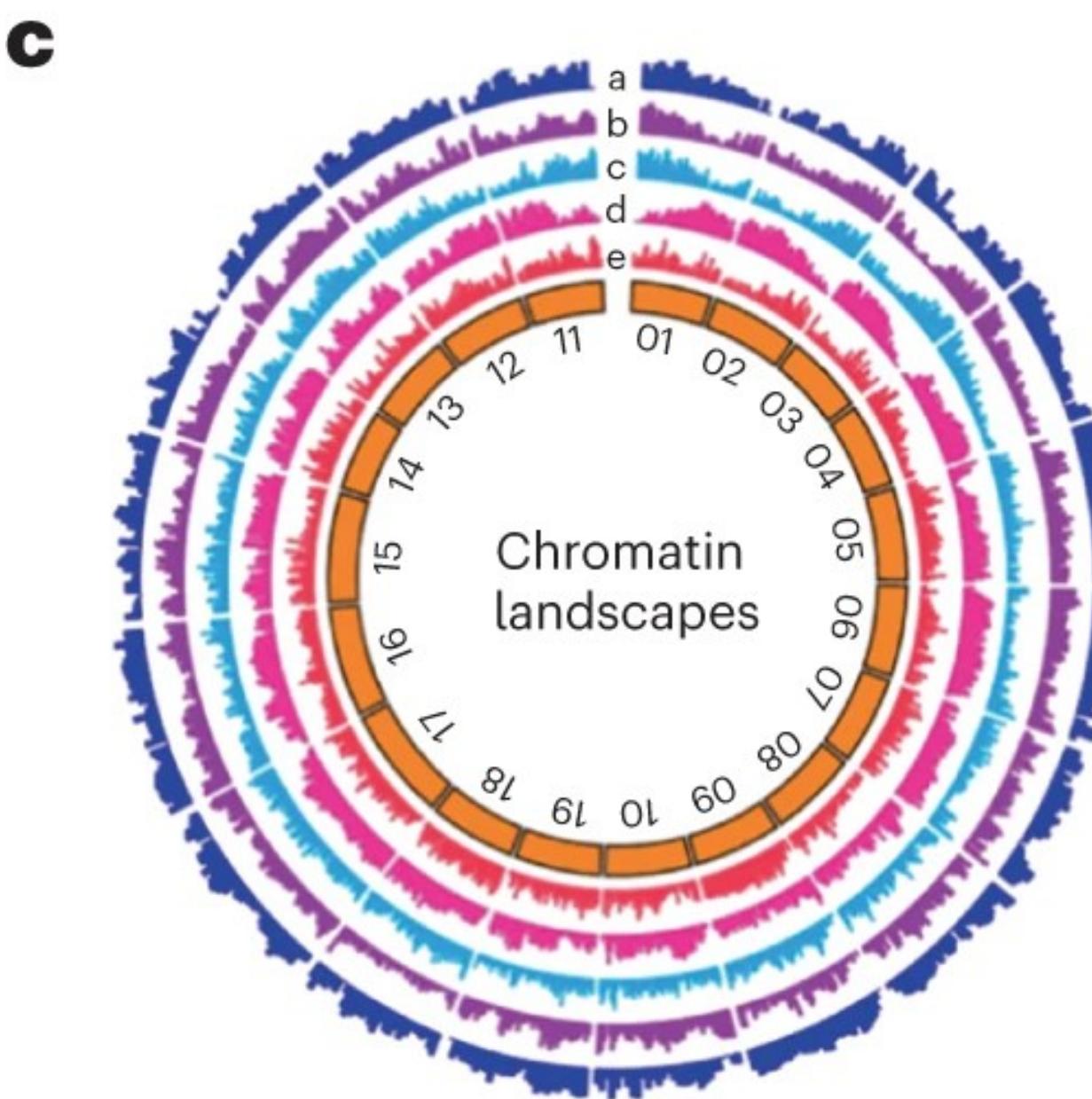
Transposons can produce novel (and useful) diversity

- Insertions
 - Insertion into an enhancer
 - Insertion into a repressor
 - Insertion into an exon
 - Insertion into an intron
- Introduction of new information
 - New enhancer
 - New splice site
 - New cis-regulon
- Introduction of new epigenetic regulation

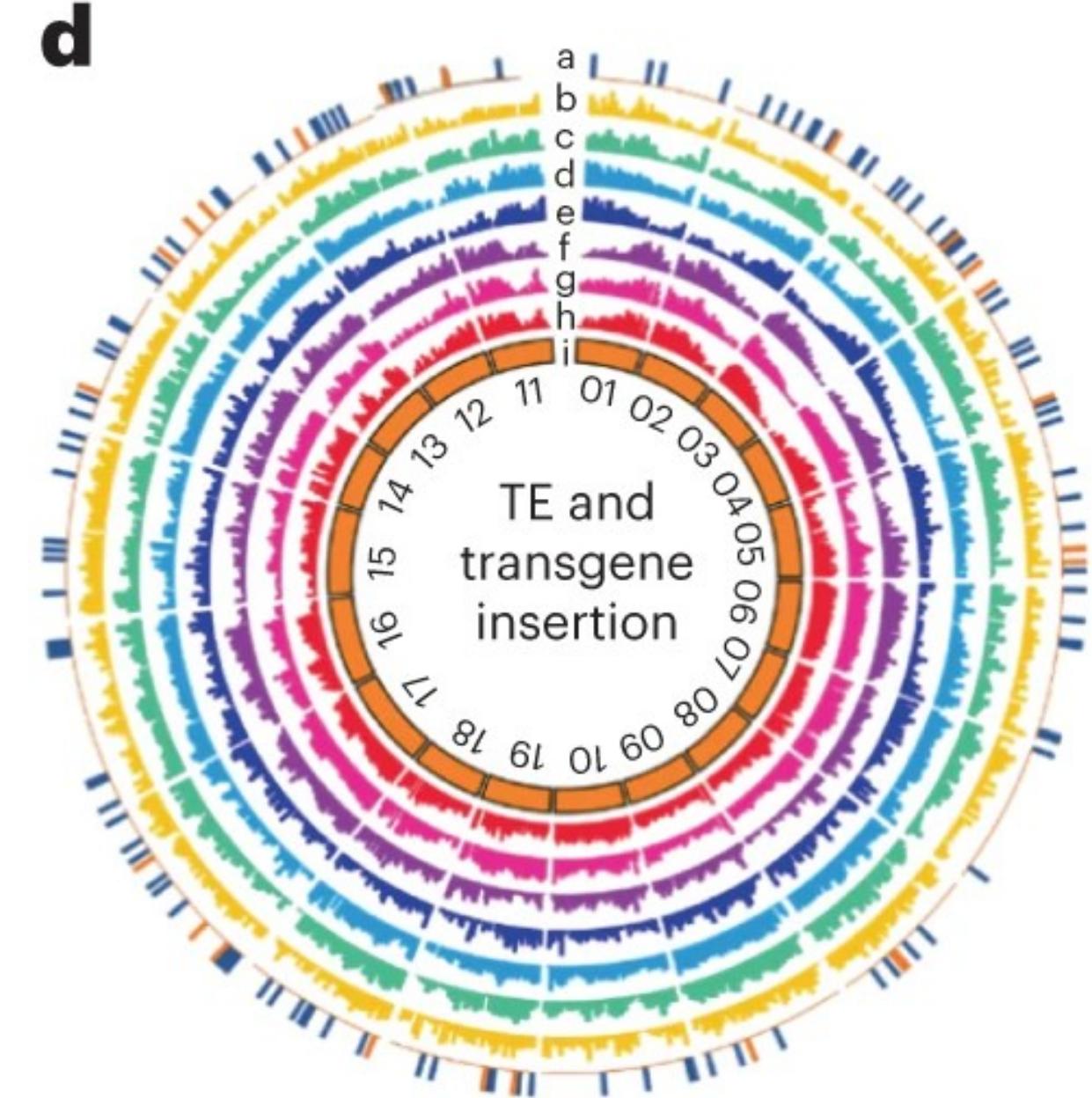


Transposable elements across a genome – resolution improves with improved genome assemblies.....

- Poor assemblies reduce/omit the contribution of repeat elements to the genomic landscape
 - Collapse of repeat element space due in part to inability to accurately assemble/map repeat containing fragments
 - Fragmentation of TE fragments (naturally) and a focus on geneic space has largely underplayed the presence of TEs and TE fragments in and around genes
- Long Read sequencing technologies have been revolutionary in acquiring structurally resolved de novo genomes
 - Still issues in assemblies due to the ‘same-by-same’ nature of many element containing DNA fragments where differences may only be 1-2 SNVs – so sequencing technology accuracy very important



Tracks a and b represent respectively the location of **permissive** histone marks H3K27ac and H3K4me3
Track c depicts the gene density across the LAB genome,
Tracks d and e represent the location of **repressive** histone marks H3K9me2 and H3K27me3, respectively.



Track a, transgene insertion sites; red ‘ticks’ represent insertions derived from stable transformation, blue ‘ticks’ represent insertions derived from transient agro-infiltration.
Track b, insertions of intact Copia TEs
Track c, insertion of all annotated Copia TEs, including fragmented elements.
Track d, distribution of CHH methylation marks
Track e, gene density across the LAB genome.
Track f, insertions of all annotated Gypsy TEs, including fragmented elements.
Track g, distribution of CG methylation marks.
Track h, distribution of CHG methylation marks.
The innermost circle represents the numbered chromosomes.



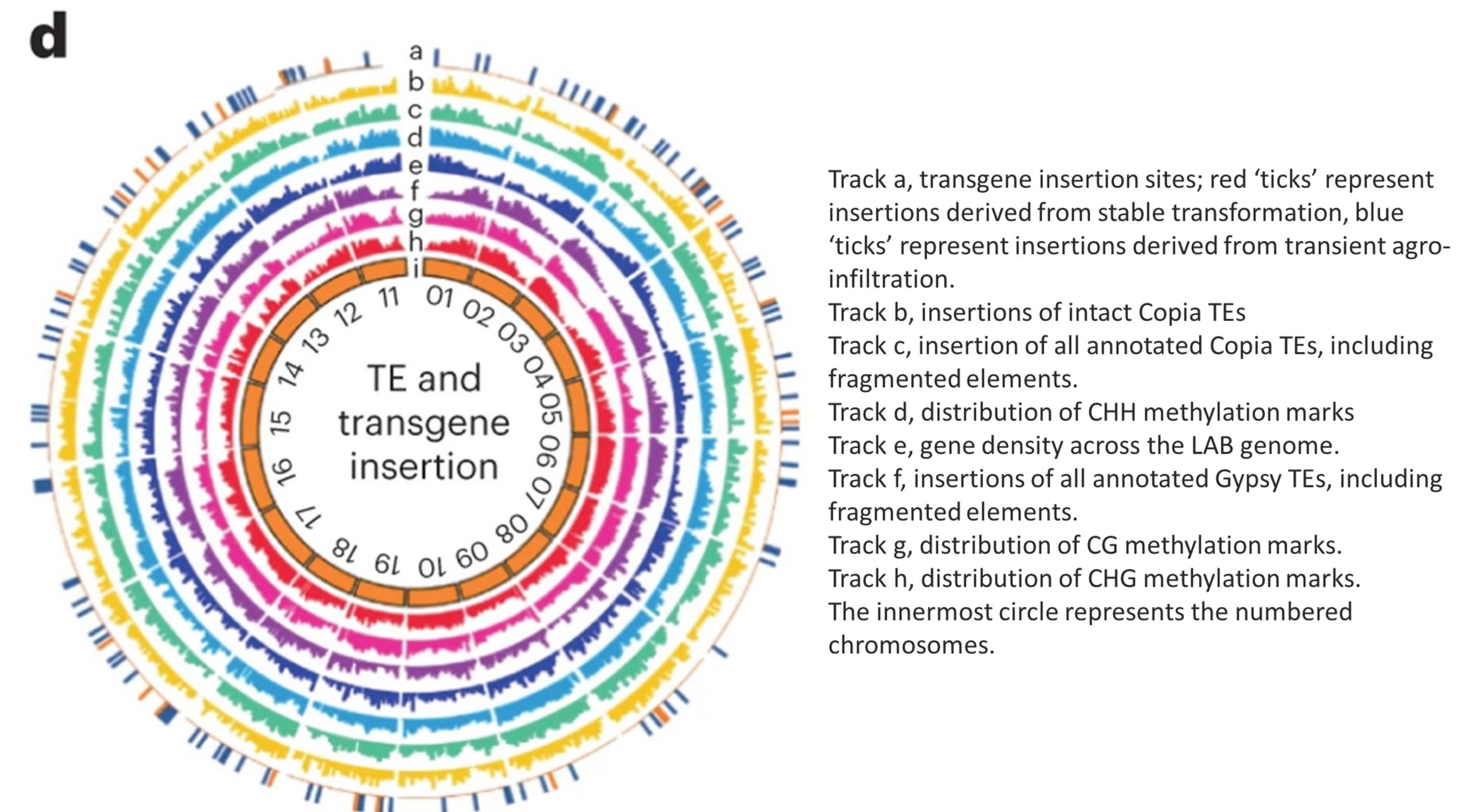
TEs aren't located only in repeat islands

- Transposons are (in general) very common in and around genes



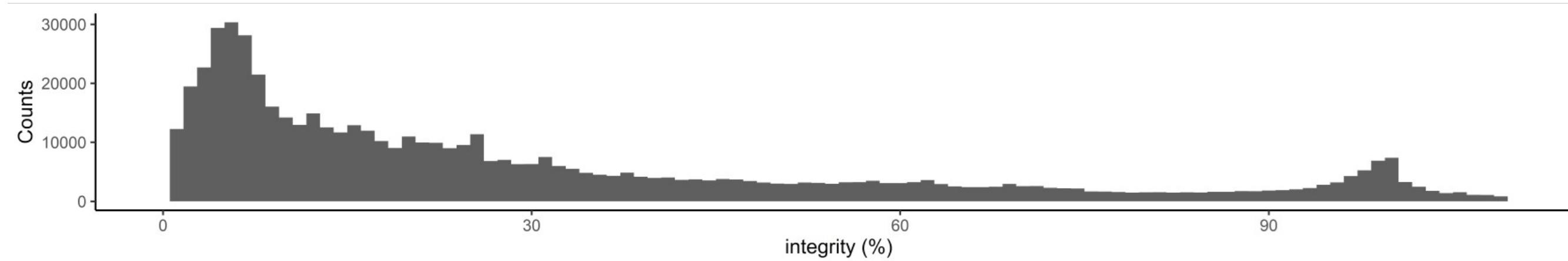
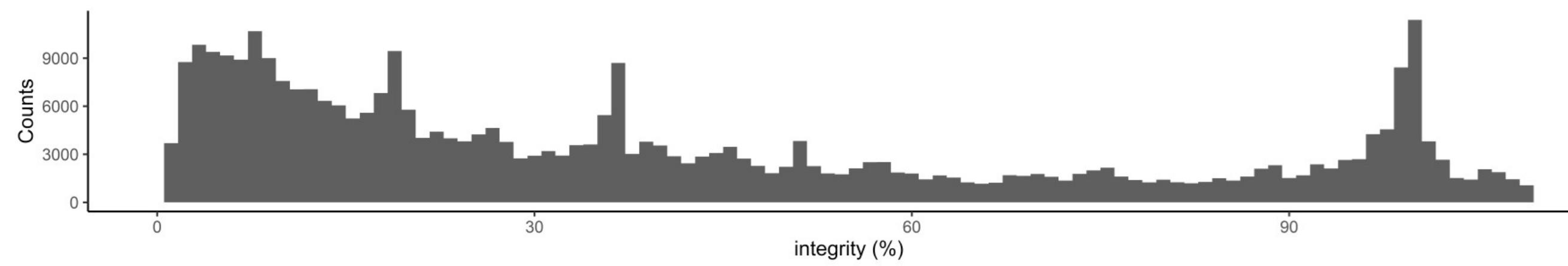
- Newly mobilised elements appear to accumulate in or near gene rich regions of the genome

- In *Nicotiana benthamiana* we see larger numbers of fully intact Copia elements in and around gene rich regions compared to the older and more fragmented Gypsy elements



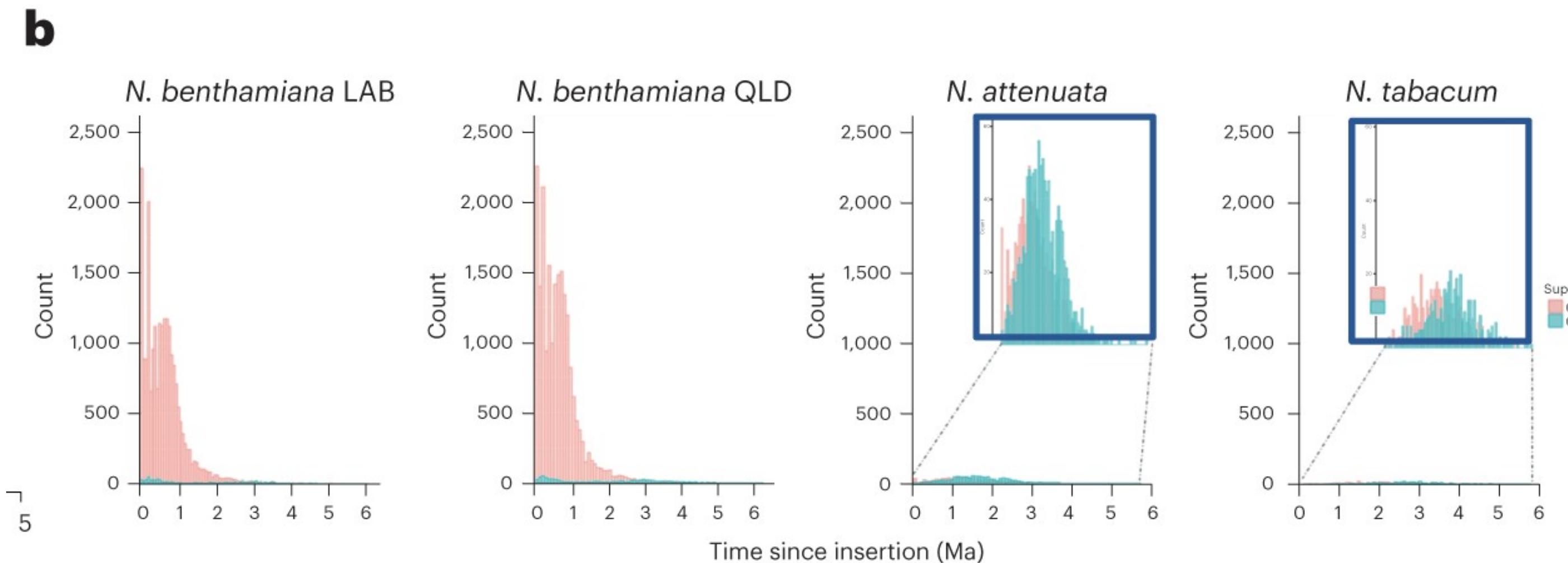
Most TEs are fragments of previous activity

- Genomes are largely comprised of dead/dying elements
 - Epigenetic ‘tagging’ of elements is very rapid
 - Once tagged and rendered inoperable rapid mutation (deletion, fragmentation, accumulation of SNVs) follows
- Bursts of activity are evident but likely accumulate slowly from a very limited number of parental loci



Natural transposon mutants

- Natural transposition is thought to be rare per generation
- Yet across time massive accumulation of new elements can be observed
 - In some cases, we see very large and very ‘recent’ bursts
 - Remembering that what we see in genomes are those elements transmissible via germ cell lineages – very little is known around rates of somatic TE based mutation
- Natural TEs have been exploited to develop tagged-mutant populations in a number of species
 - Ac/Ds Maize
 - Mutator Maize
 - Spm1 Maize
 - TNT-1 Tobacco/Medicago
 - Tam3 Antirrhinum
 - dTPH1 Petunia



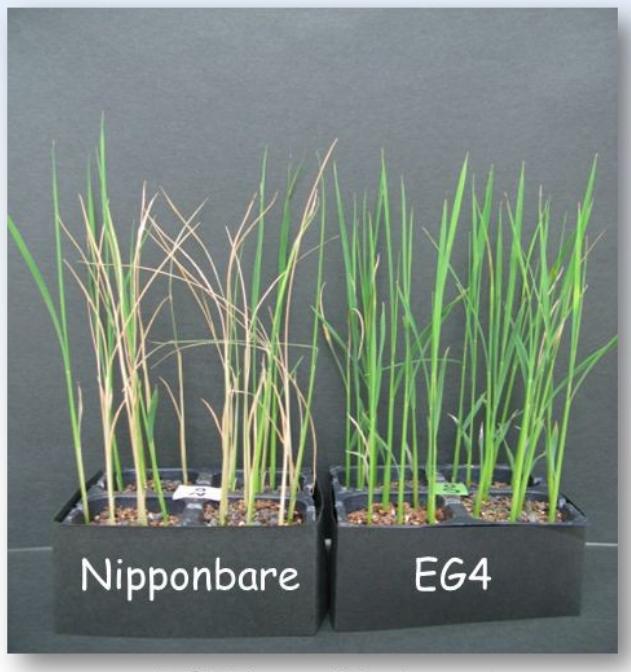
Ranawaka, B., An, J., Lorenc, M.T. et al. A multi-omic *Nicotiana benthamiana* resource for fundamental research and biotechnology. *Nat. Plants* (2023). <https://doi.org/10.1038/s41477-023-01489-8>

A guide to enhancing transposon activity

- Utilise plant lines with active transposition
 - Maize
 - Rice
 - Petunia
 - Tobacco
 - Antirrhinum
- Utilise “active” transposons from one species in another
 - i.e. horizontal TE transfer
 - 1-2 generations of activity prior to silencing/stabilization
 - Transgenic
- Convince native TEs to become active
 - Suppress epigenetic silencing
 - Use mutants of silencing machinery like DDM
 - Use chemical inhibitors of silencing components



 **BIO5**
Institute



EG4 is salt tolerant

 iPlant
Collaborative



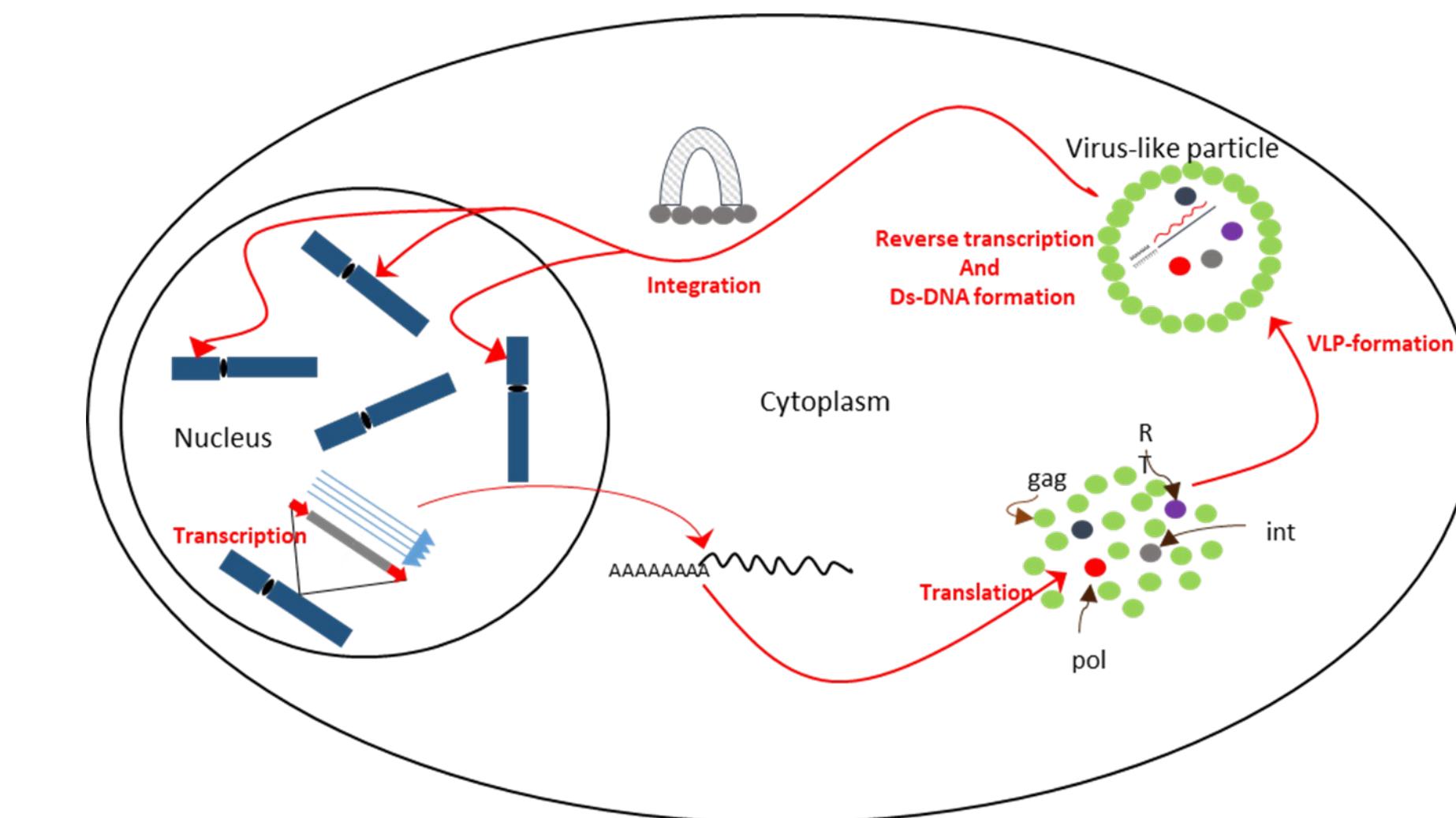
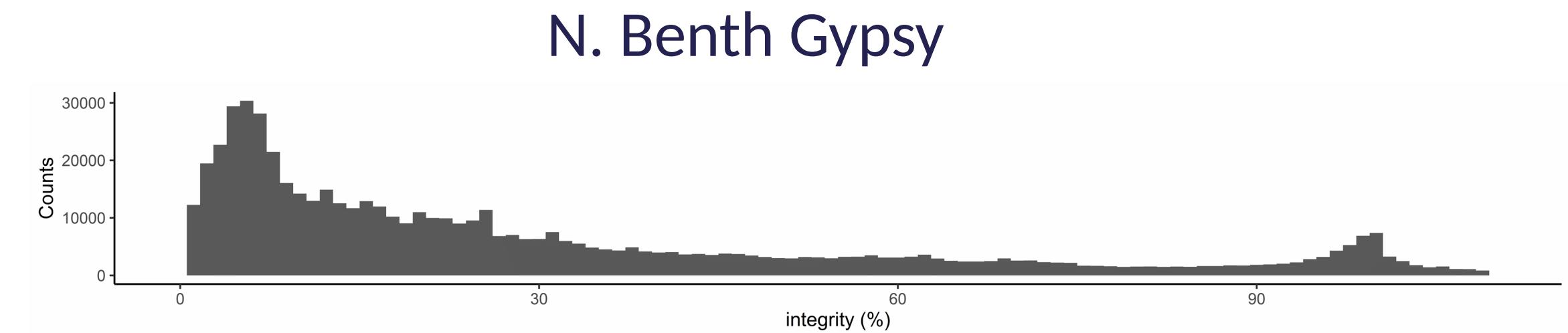
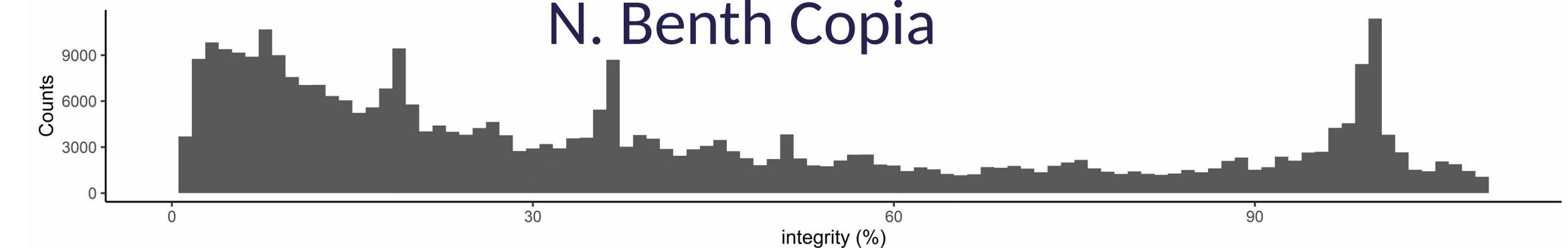
Chemical epigenetic inhibitors

- Inhibition of DNA methylation
 - 5-Azacytidine and 5-aza-2'-deoxycytidine
 - Zebularine
- Inhibitors of RNA polymerase I, II and to a extent III
 - α-Amanitin
- Inhibitors of histone de-acetylation (HDAC inhibitors)
 - 4-phenyl butyric acid (4PBA)
 - Trichostatin A
 - Others including dietary compounds such as butyrate, biotin, lipoic acid, garlic organosulfur compounds, and metabolites of vitamin E
- Stress
 - Stress treatments will de-repress regions where stress regulated genes are resident
 - TEs residing locally to this will become transcriptionally active



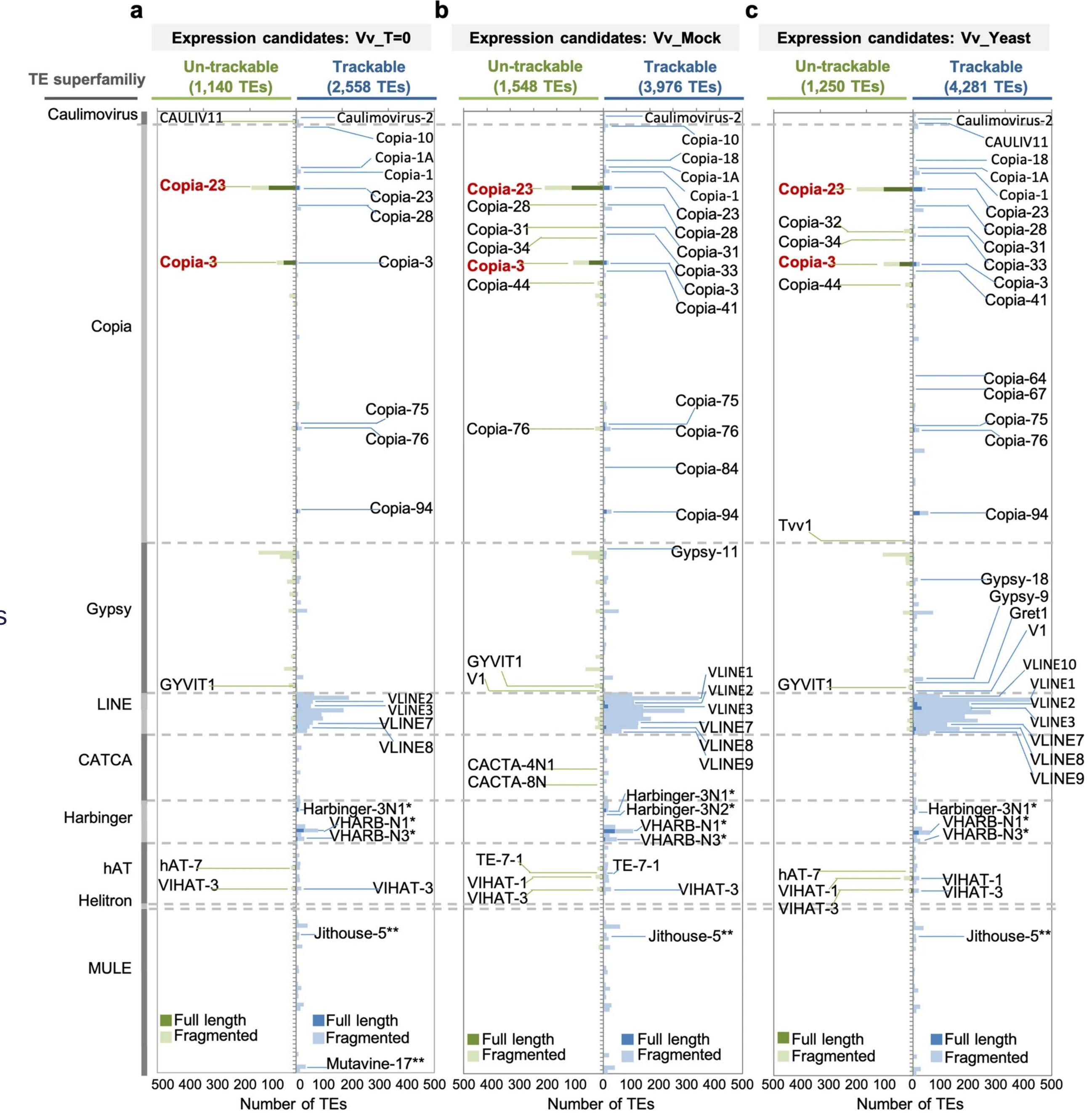
Which transposons become active....

- Surprisingly few element families are competent for mobilisation
- Even fewer members of these families are intact
- Likely only 1 or 2 loci 'seed' new insertions under normal conditions
- Mostly activation is recorded as transcription
 - This is misleading as many elements and *element fragments* can become transcriptionally active
 - Short read approaches for global analysis worst
 - Long read approaches much better
- Best approach is to recover plants and re-sequence to detect new insertions
 - Not easy
 - Design specific capture panels to reduce sequencing and computational cost
 - Best to identify pool of family candidates
 - Use a technique called VLP-Seq (virus-like protein sequencing)



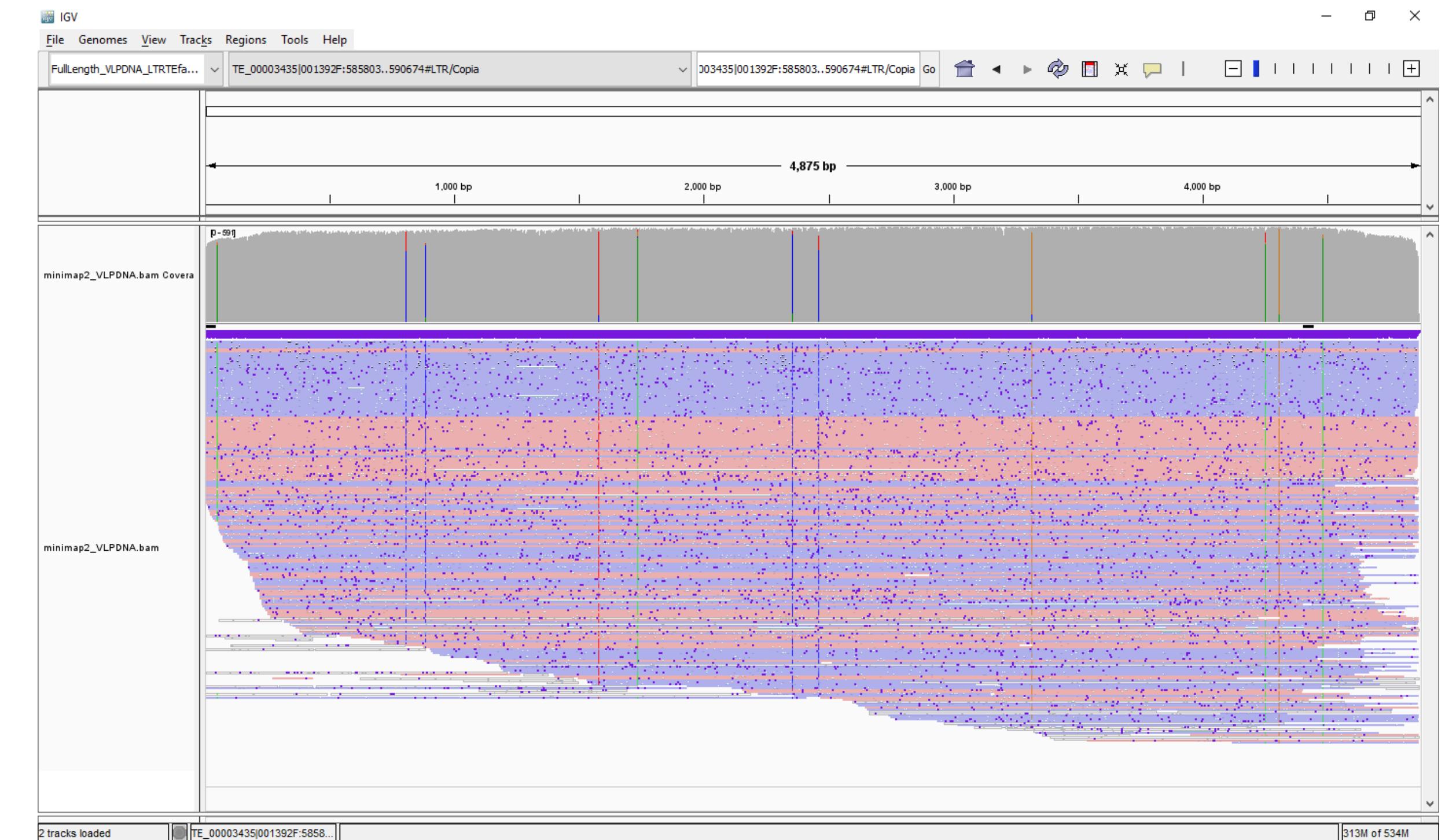
Transcriptional activatable elements in grapevine

- Focusing on transcribable elements allows sub-setting of loci most likely to contribute new transposable type -1 elements
- 2 families (both copia) meet criteria
 - Untrackable elements are those that have no discriminating SNVs to map unambiguously to a locus – thus must have retrotransposed <7000 years ago
 - Trackable elements are those with discriminating SNVs that allow mapping to a unique locus
 - Criteria for recent mobility therefore is transcribed ‘full-length’ elements that are untrackable as multiple identical insertions
 - Short-read RNA-seq very problematic!



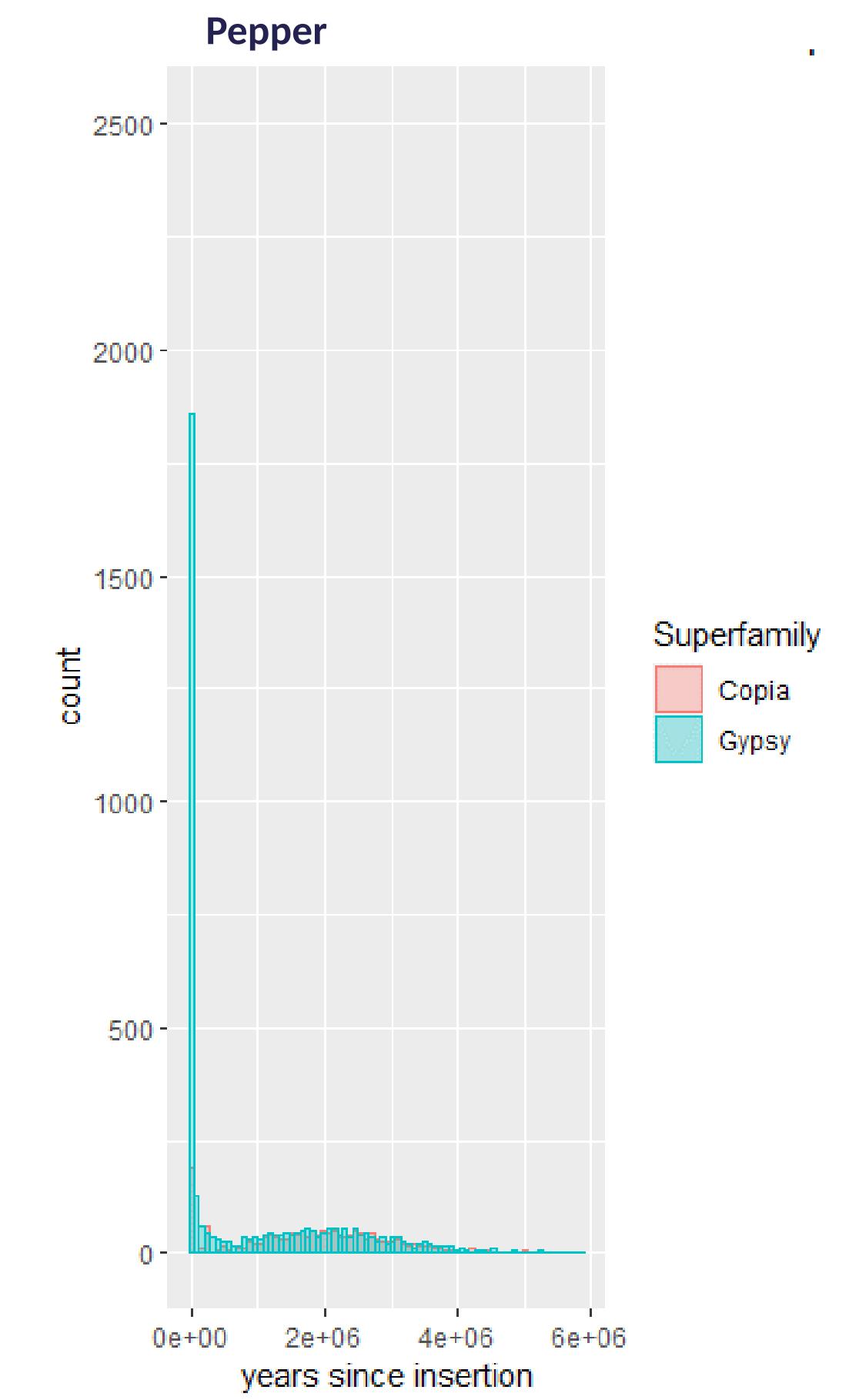
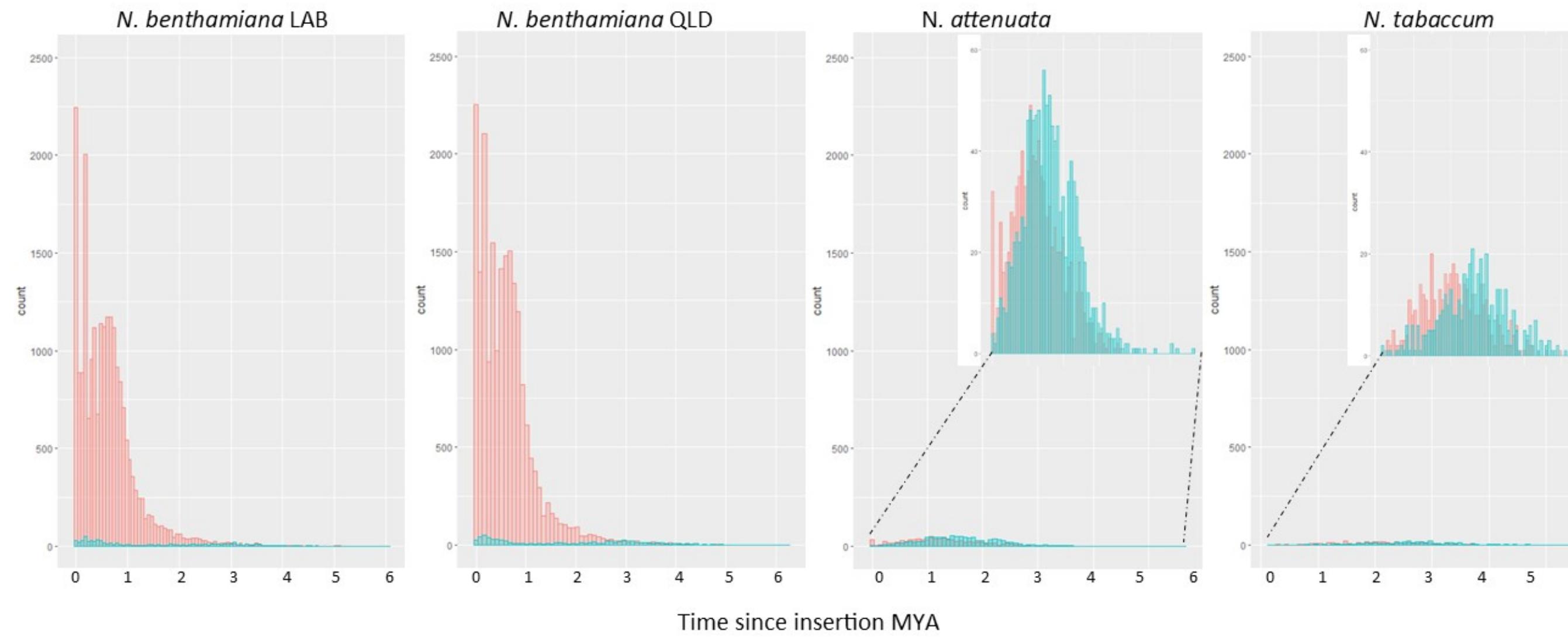
VLP-Seq – a way to discover mobilising (retro)elements in plants

- Long read RNA-seq is a potentially very useful tool to explore transcript architecture
- VLP-seq relies on isolation of VLPs from plant tissues, capture of the DNA within these vesicle like structures and sequencing the products using ONT-seq
- Map data back against TE library to identify intact element
 - If have numbers of intact RNAs mapping to loci that are also intact then have resolved a limited number of loci capable of ‘seeding’ new elements.



Nicotiana benthamiana & *Capsicum annuum* – recent and ongoing TE burst an opportunity to discover the natural causes of periodic TE expansion in plant genomes

- TE annotation and analysis of the *N. bent* genomes of LAB and QLD
 - Very recent and ongoing burst of Copia retro-transposable elements
 - 4 families contributing to the burst
- *Capsicum annuum*
 - Also a very recent burst of activity - but Gypsy rather than Copia



Summary

- TEs have an important role in generating useful new variants in nature
- We have means to artificially stimulate transposon bursts
- We suspect that certain TE families will become active and mobilise under certain environmental queues
- It is likely that new insertions will target genes in open chromatin (i.e the same chromatin ‘opened’ due to stress applied)
- Exploring the impact on genomes of this mobility



Acknowledgments

- Funding
 - BRI
 - Hapi research
 - Kiwinet
 - ARC centre of excellence in Plant success in nature and agriculture
- Rob Martienssen
- The Waterhouse group and the “newcotiana” consortia
- BRI team at Lincoln University
- Freestyle Farms



Apomixis – a process enabled by TE activity

- Our recent work to characterise the genes regulating parthenogenesis revealed a transposon fragment in the promoter of the PAR gene
 - Underwood, C.J., Vijverberg, K., Rigola, D. et al. A PARTHENOGENESIS allele from apomictic dandelion can induce egg cell division without fertilization in lettuce. *Nat Genet* **54**, 84–93 (2022).
<https://doi.org/10.1038/s41588-021-00984-y>
 - This MITE is a fragment of a hAT type 2 DNA transposon
 - Its parental element appears to possess a novel transposase that has signatures of DMM and crypton transposases, suggesting some specificity in insertion preference
 - In Dandelion and Piosella there is a different MITE but of nearly identical size and insertion pattern
 - Recombination analysis of the locus shows a zone of no recombination indicating 2 epistatically linked loci that are required to be moved together.
 - We are currently functionally characterising the functional element

