Outline

- Resources/methods/tools for studying gene function
  - Induced mutations
    - Site-nonspecific mutations induced by chemical, physical, and insertional mutagen
    - Forward and reverse genetic approaches
  - Site-specific mutation
  - Gene silencing
  - QTL/association mapping
  - Gene expression profiling
  - Multiomics

Mutation

- Mutations are changes in DNA sequence
  - Large-scale chromosomal structure changes
    - ✓ Deletion, insertion, duplication, and inversion of a chromosome fragment
  - Small-scale changes
    - ✓ One or a few of nucleotides change, insertion, or deletion

DNA sequence polymorphism

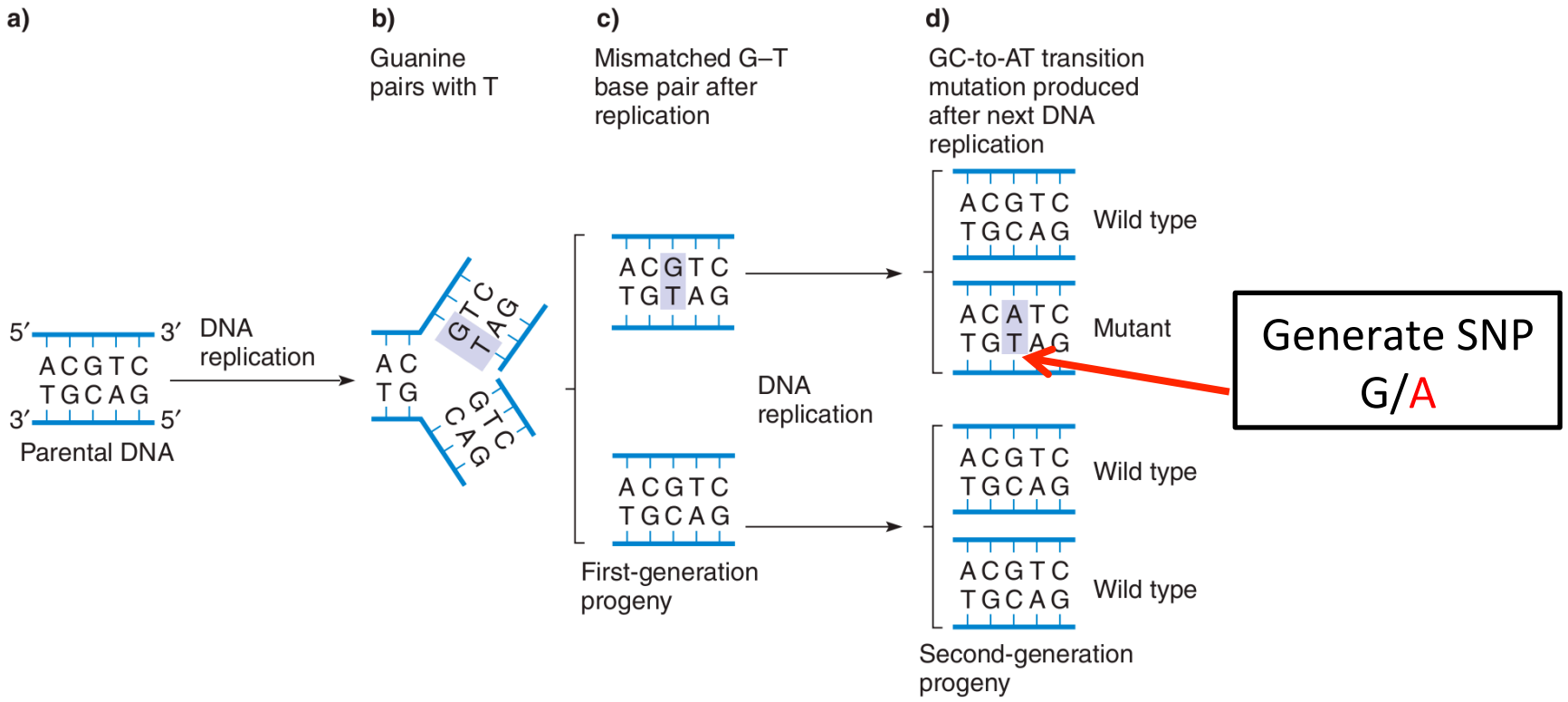
- Structure variation
  - Present/absent variation (PAV)
  - Copy number variation (CNV)
- Insertion/deletion (InDel)
- Single nucleotide polymorphism (SNP)

Mutation

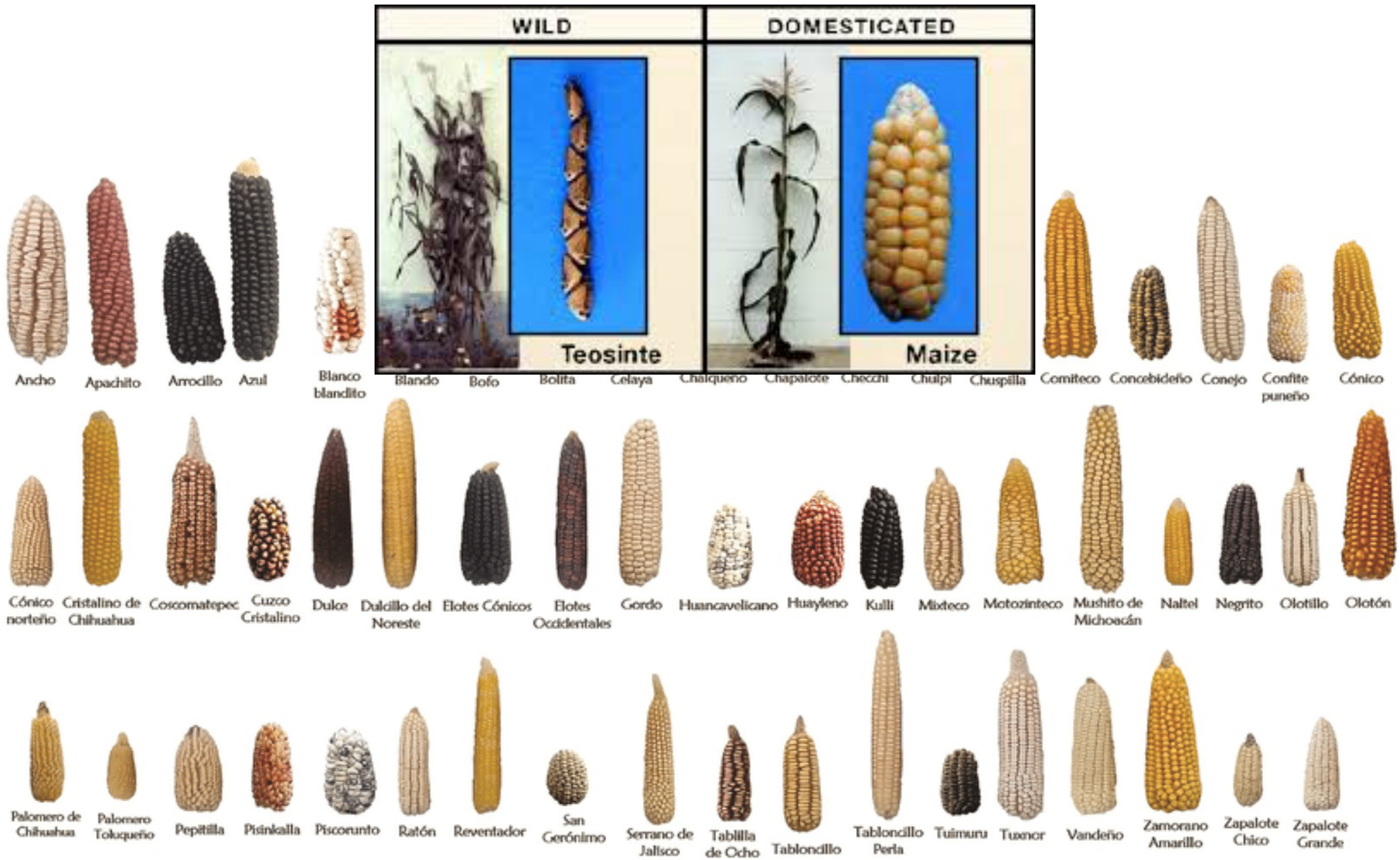
- Mutations can result from DNA copying mistakes during cell division, exposure to ionizing radiation, chemicals, or infection by viruses, etc.
- Germ line mutations occur in the egg and sperm can be passed on to offspring, while somatic mutations occur in body cells and are not passed on

Spontaneous point mutation due to DNA replication error

- Spontaneous mutation rate is very low, about 10^{-7}-10^{-11} per gene per generation

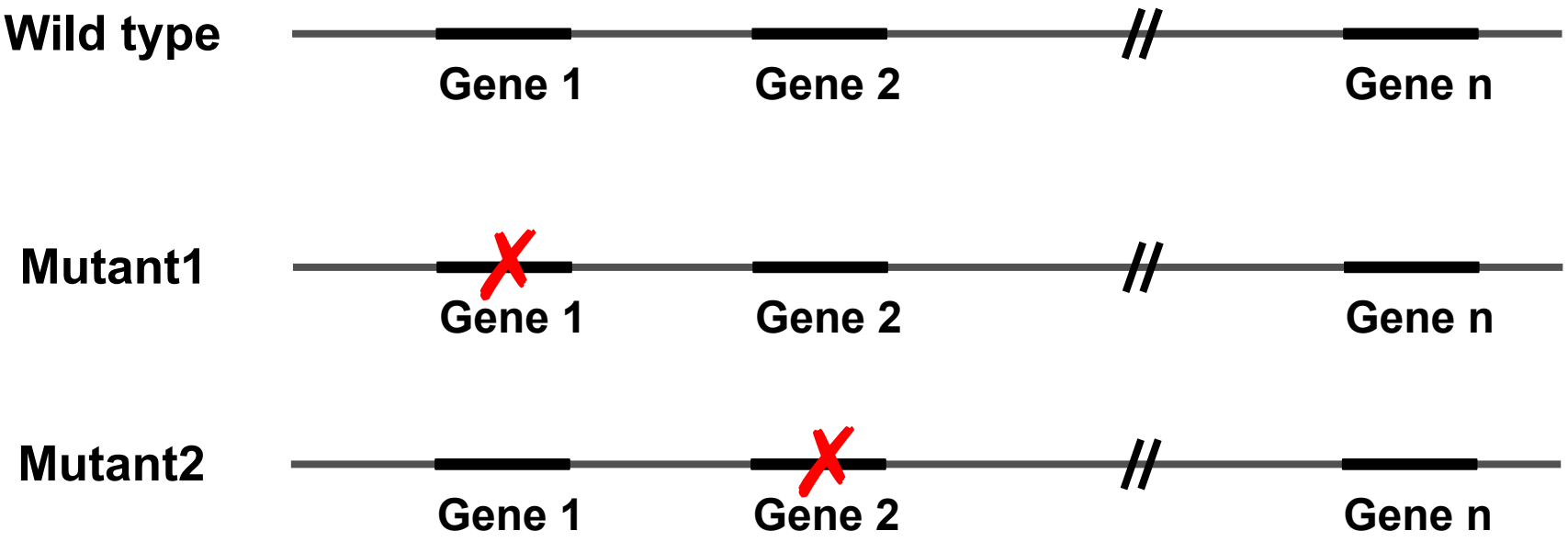


Mutation is a main cause of diversity



Induced mutations are valuable resources to study gene function

- **Induced mutations** provide possibility to understand gene's function, where other genes are same between wild type and mutant. Any phenotype change can be associated with the mutated gene



Induced mutation

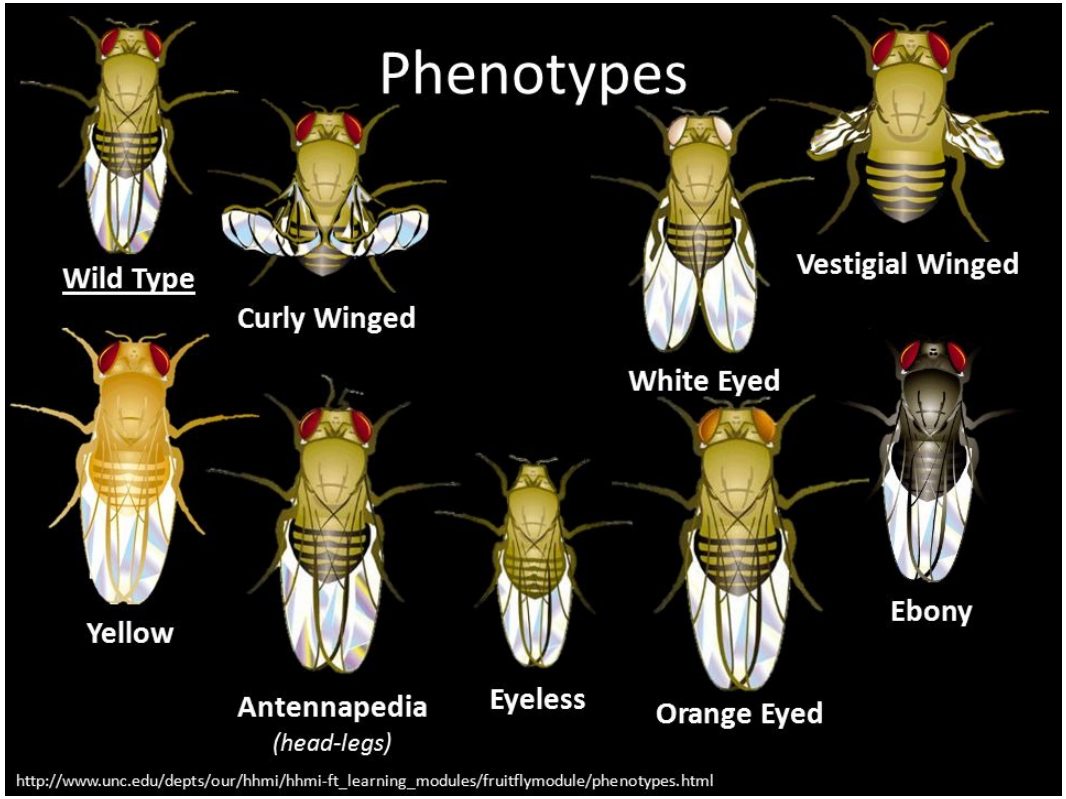
- Site-nonspecific mutations induced by
  - Chemical mutagen
  - Physical mutagen
  - Insertional mutagen
- Site-specific mutation
  - Homologous recombination
  - Engineered enzymes, like CRISPR-Cas

Site-nonspecific mutation

- Chemical, physical, and insertional mutagen
  - ✓ How do those mutagens create mutations?
  - ✓ What are characters of the induced mutations
  - ✓ How to use the induced mutations to study gene's function
    - Reverse genetics
    - Forward genetics
  - ✓ How to create gene-index catalogue for induced mutants (locate mutation sites for all mutants)

Chemical mutagens

- 1940s, Charlotte Auerbach and J.M. Robson found that mustard gas can cause mutations in fruit flies
- Large number of chemical mutagens have been identified, especially after the development of the Ames test in 1970s by Bruce Ames

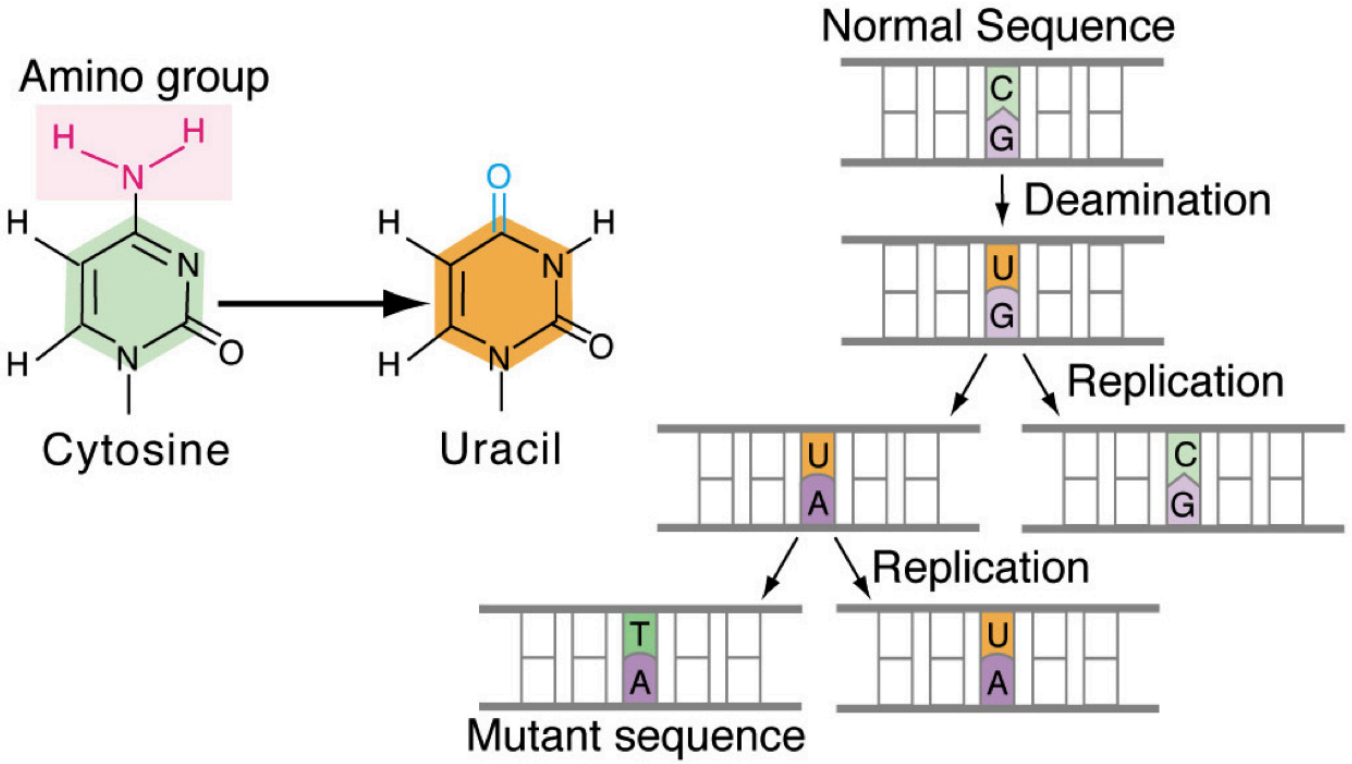


Chemical mutagen induces point mutation

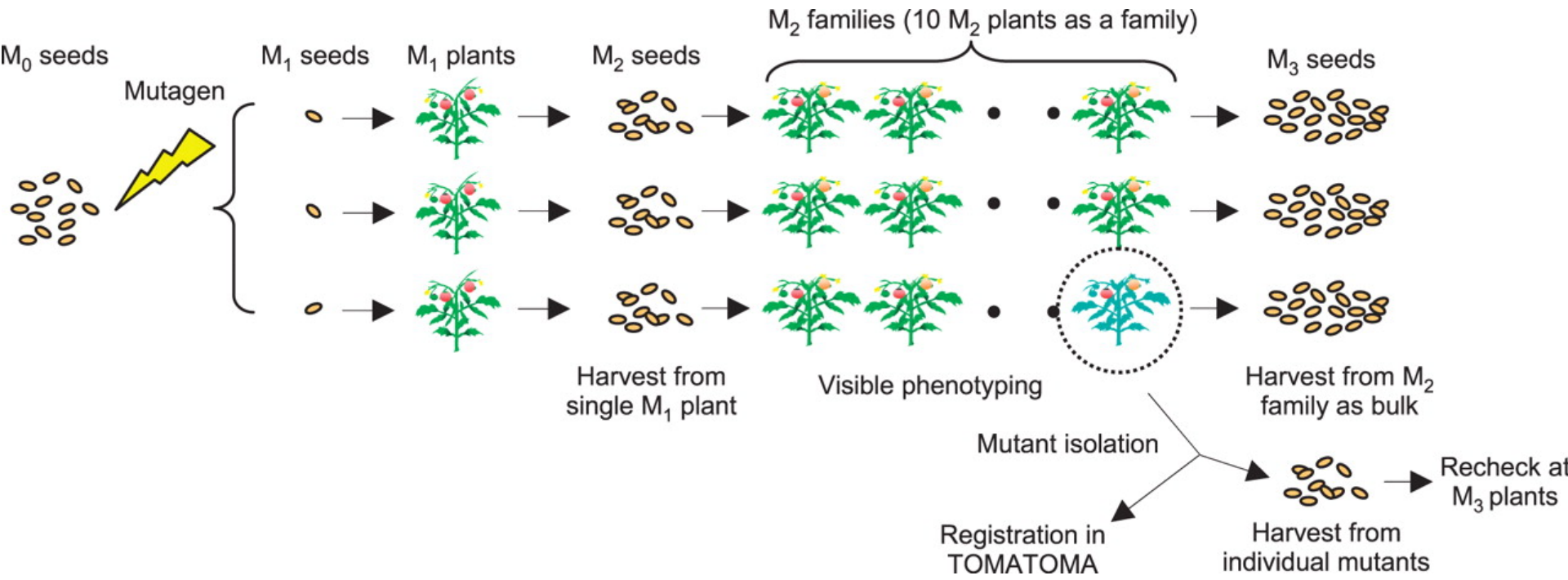
- Chemical mutagen causes DNA damages (base modification, change, and loss); the DNA damages lead to mismatches; some mismatches are not repaired by mismatch repair system and result in point mutations
- **Point mutation** is a single nucleotide change of DNA

Chemical mutagen - Deamination

- Nitrous acid causes deamination of cytosine (C) and produces uracil (U), which is a normal base in RNA
- If the uracil (U) is not replaced, an adenine (A) will be incorporated into the new DNA strand, resulting in a CG-to-TA transition mutation



Flow chart for constructing mutagenized population using chemical mutagen

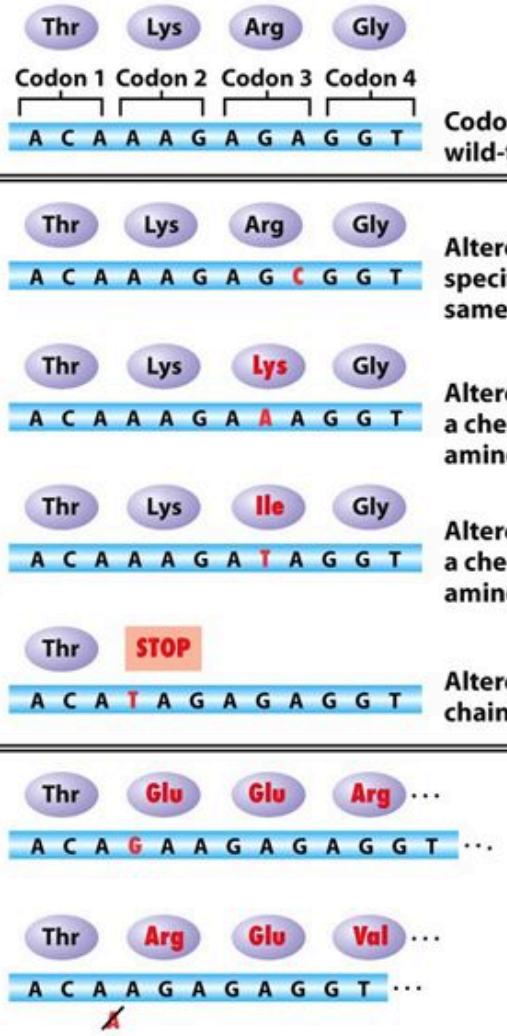


Types of point mutations

- **Synonymous mutation:** no change on the encoded amino acid
- **Missense mutation:** a change in one DNA base pair that results in the substitution of one amino acid for another
  - Conservative missense mutation changes a same type of another amino acid, which may not change structure of the protein and do not change its function
  - Non-conservative missense mutation changes to another type of amino acid
- **Nonsense mutation**: the altered DNA sequence prematurely signals the cell to stop building a protein and results in a shortened protein that may function improperly or not at all

Types of point mutations

Types of mutations at the DNA level	Results at the molecular level	
No mutation	Wild type	Codons specify wild-type protein.
Transition or transversion	Synonymous mutation	Altered codon specifies the same amino acid.
	Missense mutation (conservative)	Altered codon specifies a chemically similar amino acid.
	Missense mutation (nonconservative)	Altered codon specifies a chemically dissimilar amino acid.
	Nonsense mutation	Altered codon signals chain termination.
Indel Base insertion	Frameshift mutation	
Base deletion	Frameshift mutation	



DNA codon table

Amino acids biochemical properties | nonpolar | polar | basic | acidic

Termination: stop codon

Standard genetic code

1st base	2nd base T		2nd base C		2nd base A		2nd base G		3rd base
T	TTT	(Phe/F) Phenylalanine	TCT	(Ser/S) Serine	TAT	(Tyr/Y) Tyrosine	TGT	(Cys/C) Cysteine	T
	TTC		TCC		TAC		TGC		C
	TTA	(Leu/L) Leucine	TCA		TAA[B]	Stop (*Ochre*)	TGA[B]	Stop (*Opal*)	A
	TTG		TCG		TAG[B]	Stop (*Amber*)	TGG	(Trp/W) Tryptophan	G
C	CTT		CCT	(Pro/P) Proline	CAT	(His/H) Histidine	CGT	(Arg/R) Arginine	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	(Ile/I) Isoleucine	ACT	(Thr/T) Threonine	AAT	(Asn/N) Asparagine	AGT	(Ser/S) Serine	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	ATG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GTT	(Val/V) Valine	GCT	(Ala/A) Alanine	GAT	(Asp/D) Aspartic acid	GGT	(Gly/G) Glycine	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GTG		GCG		GAG		GGG		G

Whole genome sequencing of sorghum EMS-induced mutants

- Sorghum is a plant model species (~730 Mbp)
- 6400 mutants from EMS-mutagenized BTx623 seeds
- Total of 256 mutants were sequenced with an average coverage of 16x
- Revealed >1.8 million EMS-induced mutations, affecting 95% of genes in the sorghum genome
- On average, ~7,600 SNPs/mutant and ~10 SNPs/Mbp
- **The vast majority (97.5%) of the induced mutations were distinct from natural variations**

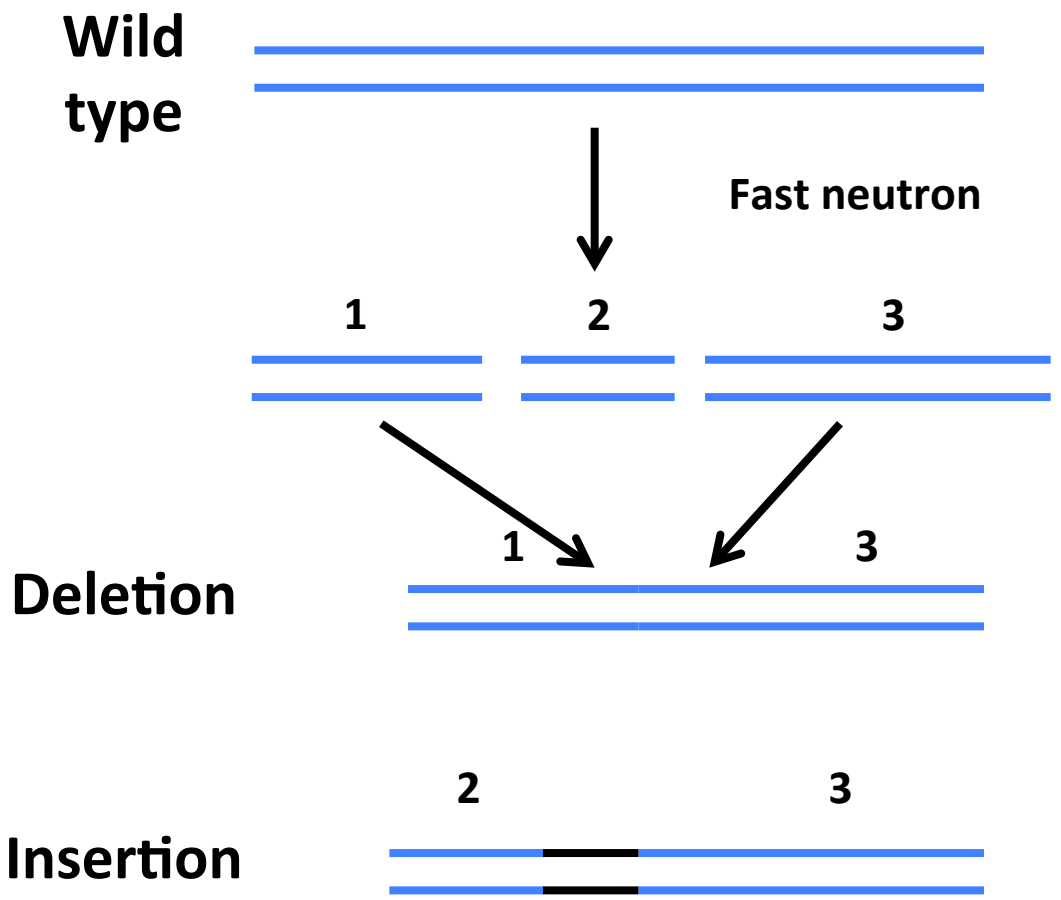
Jiao et al., The Plant Cell, 2016

Physical mutagens

- In 1927, Hermann Muller discovered that X-ray can cause genetic mutations in fruit flies
- Lewis Stadler showed the mutational effect of X-rays on barley in 1928 and ultraviolet (UV) radiation on maize in 1936
- Physical mutagens cause DNA breakage or other damages and induce mutations

Ionizing radiations and double strand break repair

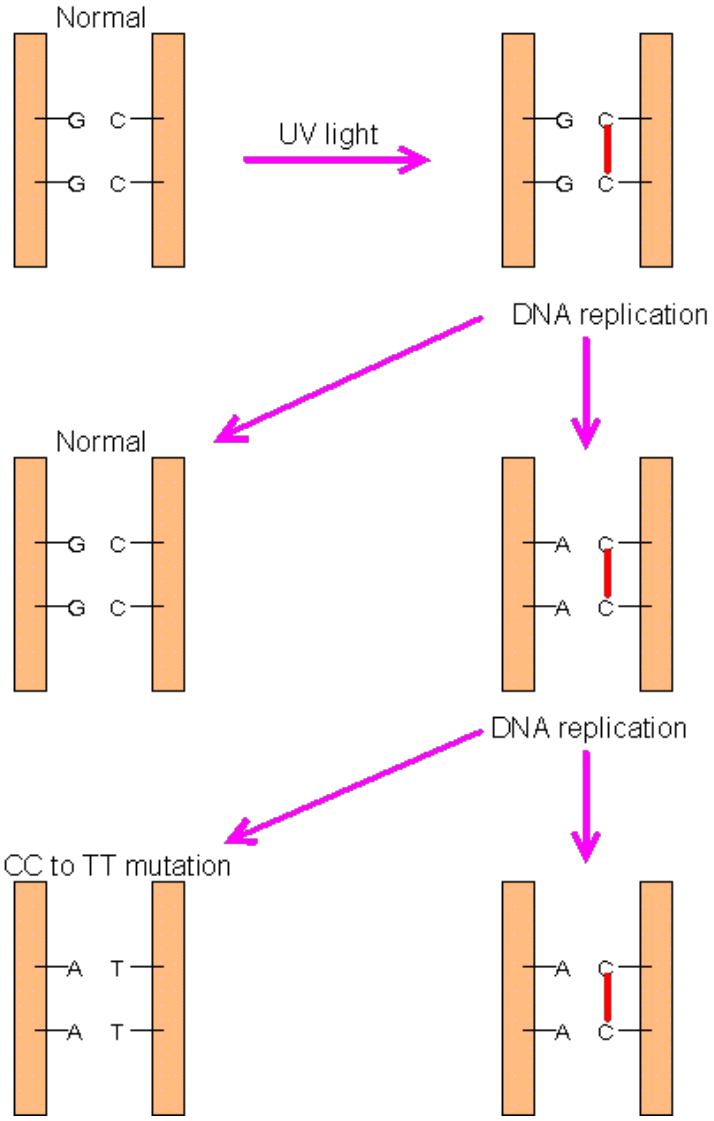
- Ionizing radiations
  - Fast moving particles such as fast neutrons have sufficient energy to physically 'punch holes' in DNA directly
  - Fast neutron may induce deletion and insertion mutations



UV (a non-ionizing radiation) induces pyrimidine dimers mutation

- Causes two consecutive pyrimidine bases on one strand to bind together
- Leads to a CC to TT mutation

Wild type



Mutant

Sequencing 1504 rice mutants facilitate functional genomic studies

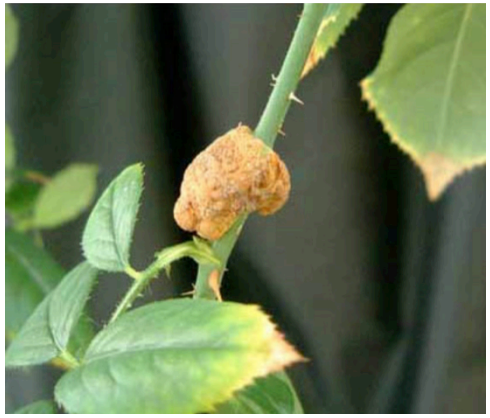
- Sequenced a fast-neutron-induced mutant population of 1504 lines in rice (~400Mbp) and identified 91,513 mutations
- On average, 61 mutations/mutant and 0.15 mutations/ Mbp
- Including 43,483 single base substitutions, **31,909 deletions,** 7,929 insertions, 3,691 inversions, and 4,436 translocations
- **Deletions** were found for 27,614 genes; the average deletion size is 8.8 kb, deletions smaller than 100 bp account for nearly 90% of all deletions

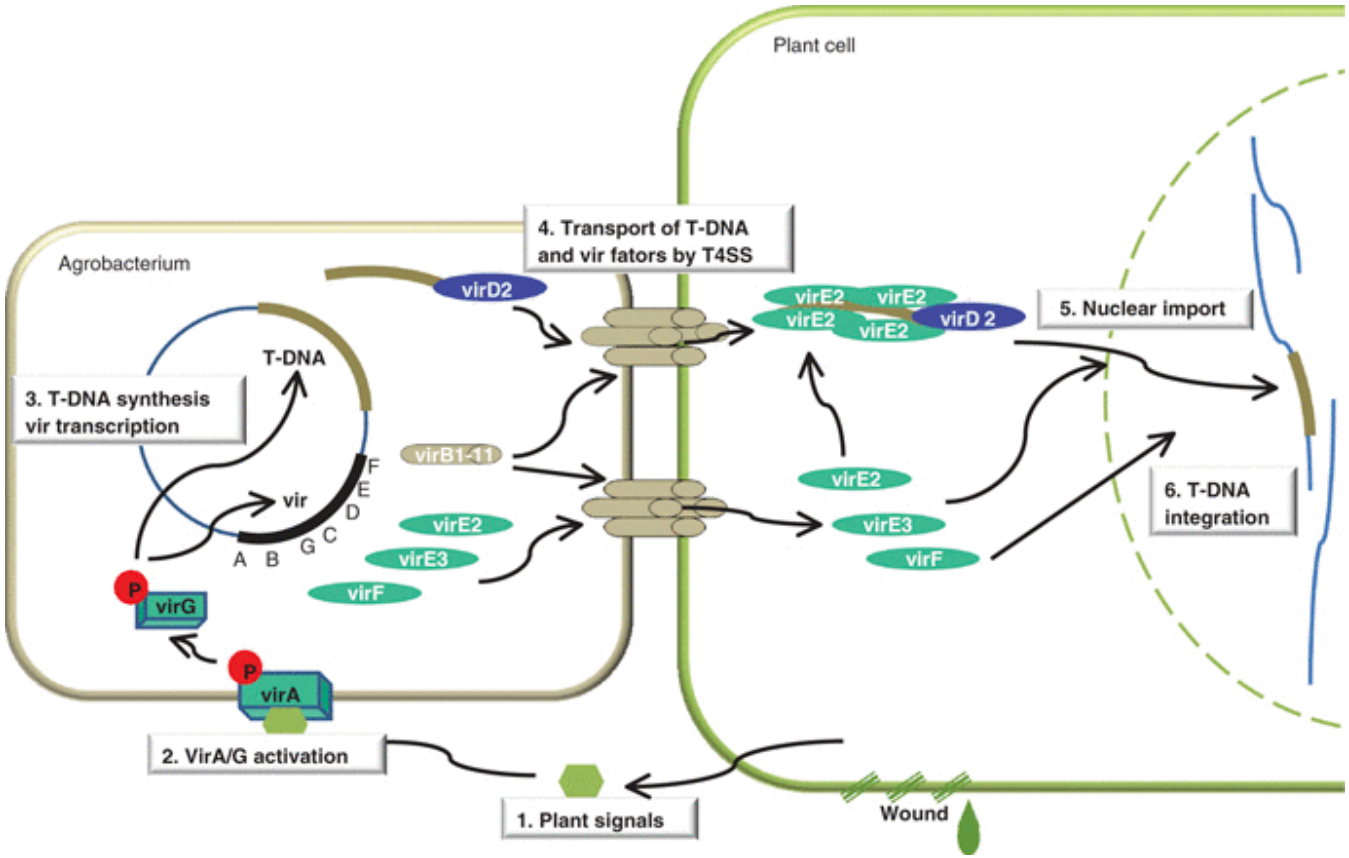
Li et al., Plant Cell 2017

Insertional mutagen

- **Insertional mutagen** induces mutations of DNA via incorporation of additional bases
- Insertional mutations can mediated by bacteria, transposon, or virus

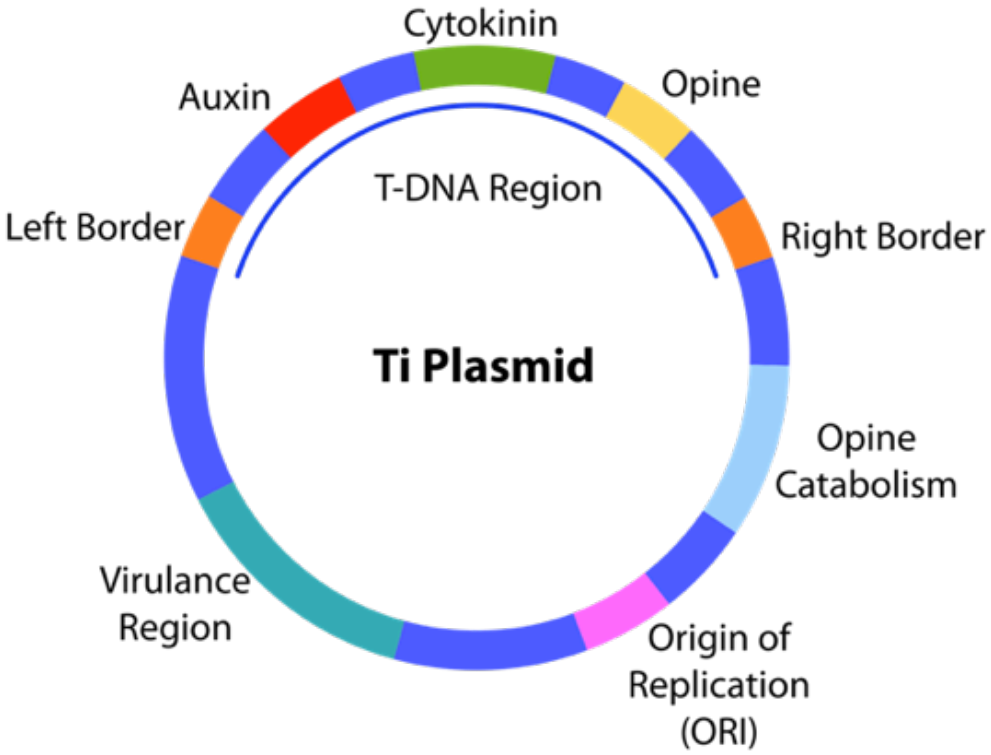
T-DNA inserts into plant genome from agrobacterium

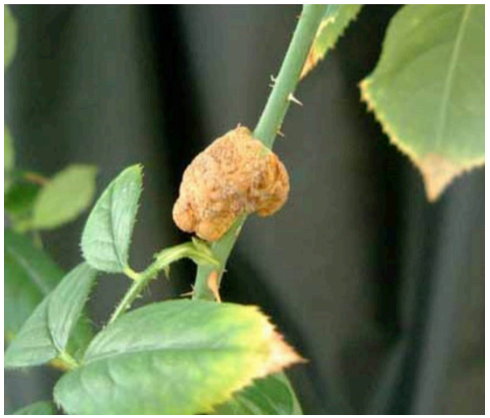




Tzfira and Citovsky, Trends in cell biology 2002

Plasmid DNA and T-DNA structure

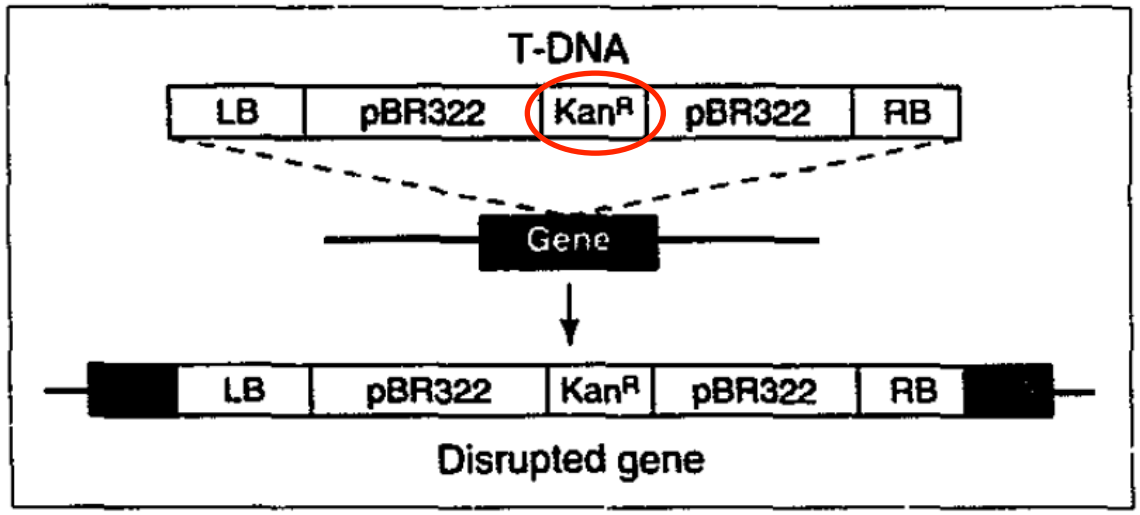




- Tumor-promoting gene: Auxin and Cytokinin are plant hormones that enable the plant cell grow uncontrollably, thus forming the tumors
- Biosynthetic genes: opine is amino acid derivatives used by the bacterium as a source of carbon and energy

Agrobacterium as insertional mutagenesis

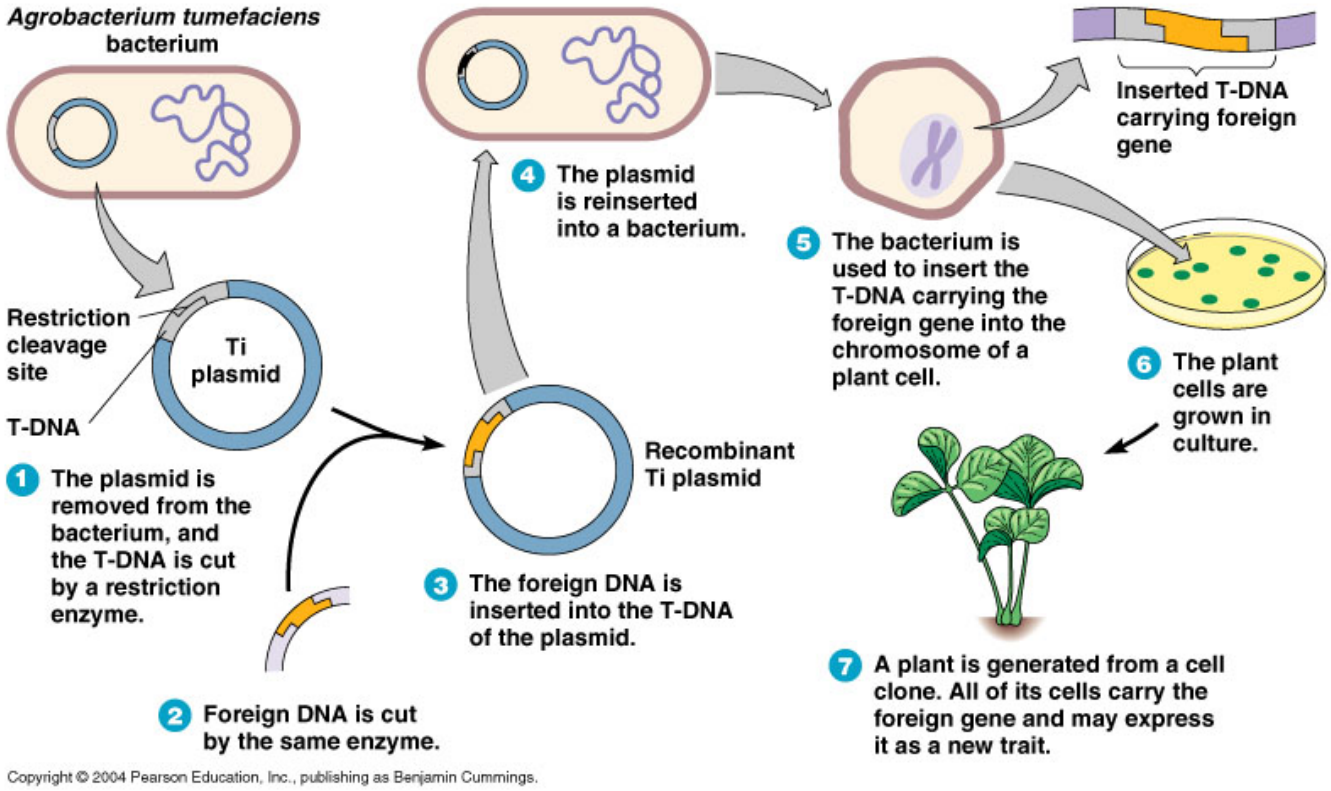
- Insertional mutations can be artificially created in lab
- Kan^R gene confers kanamycin resistance, allowing selection of the transformed plants (or mutants)
- The length of the insert is 17 kb in this case, **causing loss of gene function**



Azpiroz-Leehan and Feldmann, Trend in genetics 1997

Agrobacterium-mediated transformation --Tissue culture approach

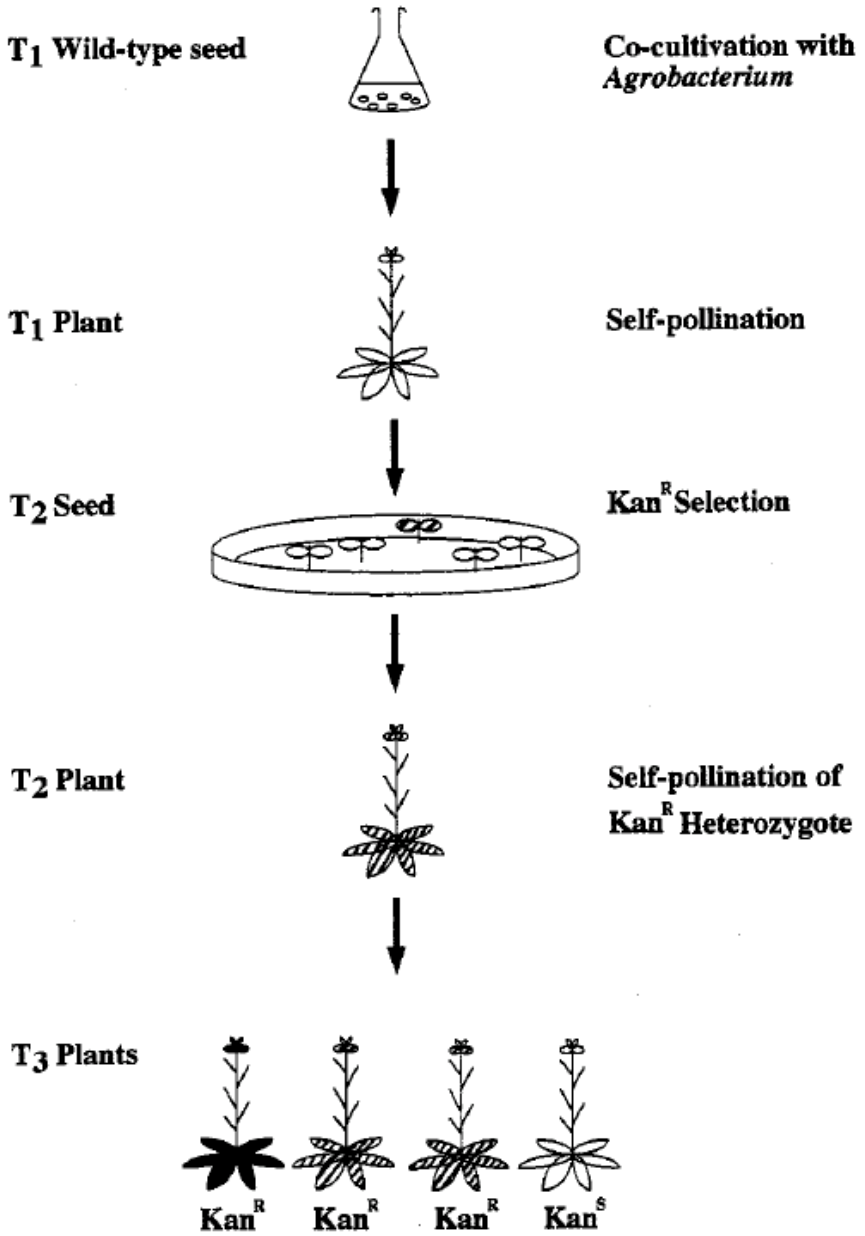
- Regeneration of whole plants generally requires weeks to months
- Time/labor intensive to construct large scale of mutational lines using tissue-culture method



kanamycin

Agrobacterium-mediated transformation
--seed infection method

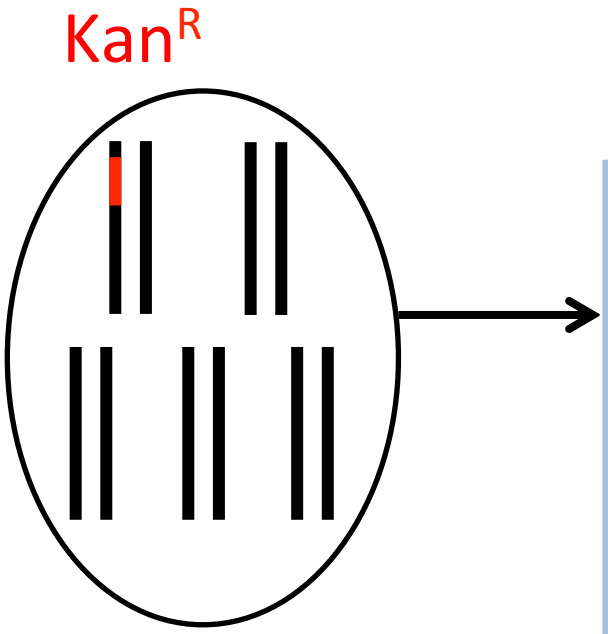
- If a cell with inserts in a germinated seed forms reproductive tissues of the T1 plant, then some T2 seeds from the T1 plant have T-DNA inserts
- **<u>How many T-DNA inserts in a T2 mutant?</u>**



Forsthoefel et al., 1992

Number of inserts per mutant can be inferred from segregation ratio of its progenies

- If there is one T-DNA insert in a mutant, the segregation ratio of transformed plant vs non-transformed plants in its progenies is 3:1



		Pollen	
		Kan^R (50%)	**- (50%)**
Egg cell	**Kan^R (50%)**	Kan^R/Kan^R (25%)	Kan^R/- (25%)
	- (50%)	Kan^R/- (25%)	-/- (25%)

Number of inserts per mutant can be inferred from segregation ratio of its progenies

- One insert in one T2 mutant, the segregation ratio of transformed plants (kan^R) vs non-transformed plants (-/-) in its progenies is 3:1
  - Kan^R : -/- = 3:1
- Two inserts in one T2 mutant
  - Kan^R : -/- = 15:1
- Three inserts in one T2 mutant
  - Kan^R : -/- ??

❖ **Average 1.5 inserts per Arabidopsis mutant**

How many insertional mutants are needed to study most genes in Arabiposis

- Arabidopsis, a plant model species
  - Genome size: ~120,000 Kbp
  - Gene number: ~30,000
  - Average gene size: ~2 Kbp

➢ If insertional mutations are randomly distributed on chromosomes, how many mutants do we need to create so that we can get insertional mutation for every single gene?

Total number of T-DNA inserts is function of size of gene and size of genome

- The chance that a T-DNA insert is in a gene of x Kbp: x/120,000
- The chance that a insert is not in a gene of x Kbp: 1-x/120,000
- Given n inserts, the chance of none of the n inserts is in a gene of x Kbp: $(1-x/120000)^n$
- Given n inserts, the chance of at least one insert is in a gene of x Kbp: $\boldsymbol{p = 1-(1-x/120000)^n}$

How many mutants are needed to saturate the genome?

- Total number of T-DNA inserts is function of size of gene (x kb) and size of genome

$$p = 1 - (1 - x/120{,}000)^n$$

- A 2.1-kb gene requires **280,000 T-DNA inserts** to achieve 99% probability of being mutated
- Average **1.5 inserts per mutant**, **~186,000 mutants** are needed

Krysan et al., Plant Cell 1999

T-DNA insertion collections in Arabidopsis

Institution	Population size	Reference
Salk Institute	150,000	Alonso et al., 2003
Bielefeld University, Germany	71,000	Kleinboelting et al., 2012
Syngenta	100,000	Sessions et al., 2002

Characters of insertional mutant induced by T-DNA/agrobacterium

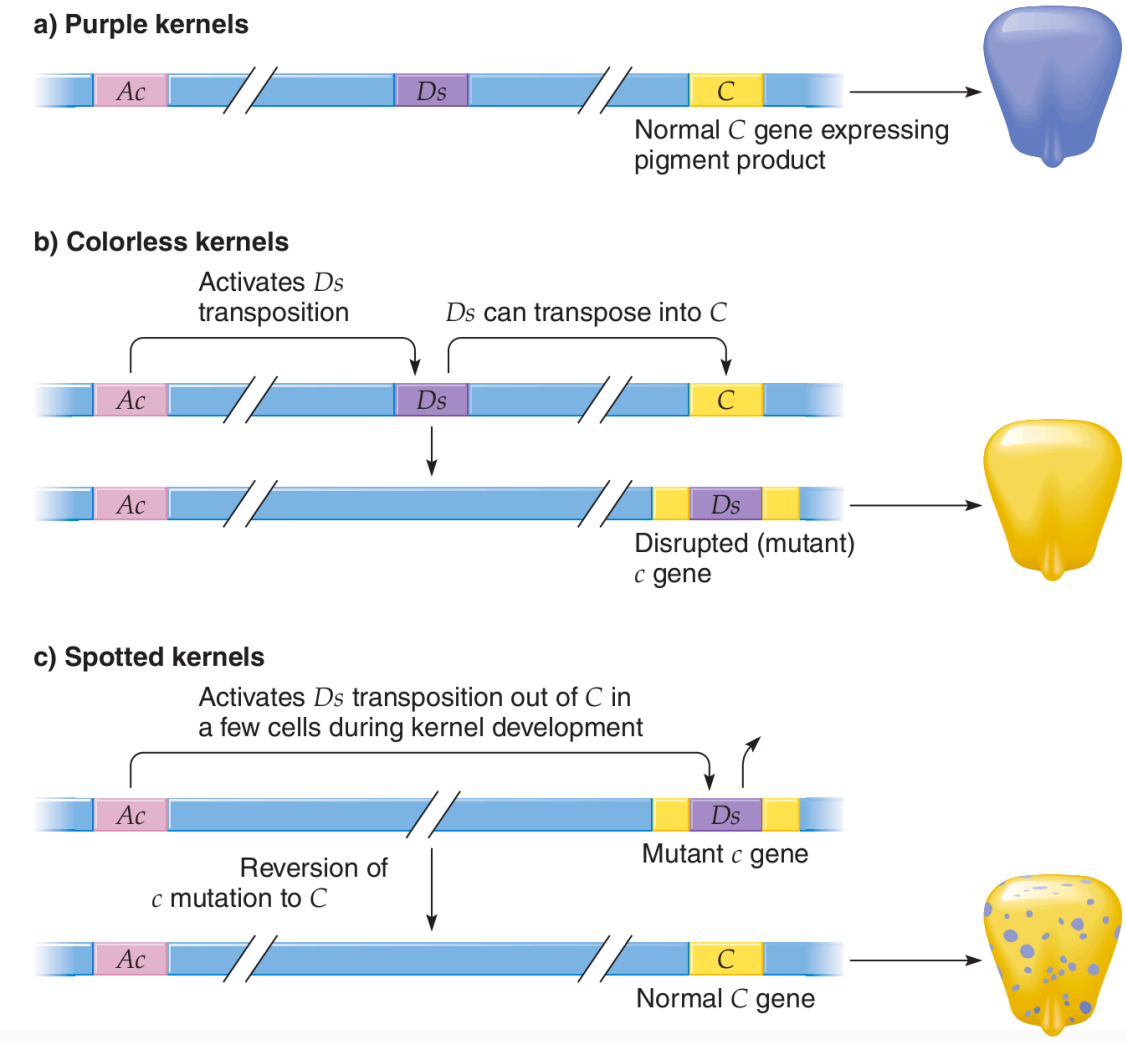
- T-DNA/agrobacterium induces mutants with large fragment insertion, which causes loss of the gene function
- T-DNA insertional mutations can not be generated for all genes. For genes required for life, insertion leads to lethal
- Low T-DNA insertional rate, ~1.5 inserts per mutant; gene with small size has low chance to get insertional mutation
- T-DNA/agrobacterium does not work for all species

Transposon

- Transposons or transposable elements are DNA fragments that can move from one location to another location within the genomes

Transposon discovery

- Transposon was first identified in maize by Barbara McClintock more than 60 years ago

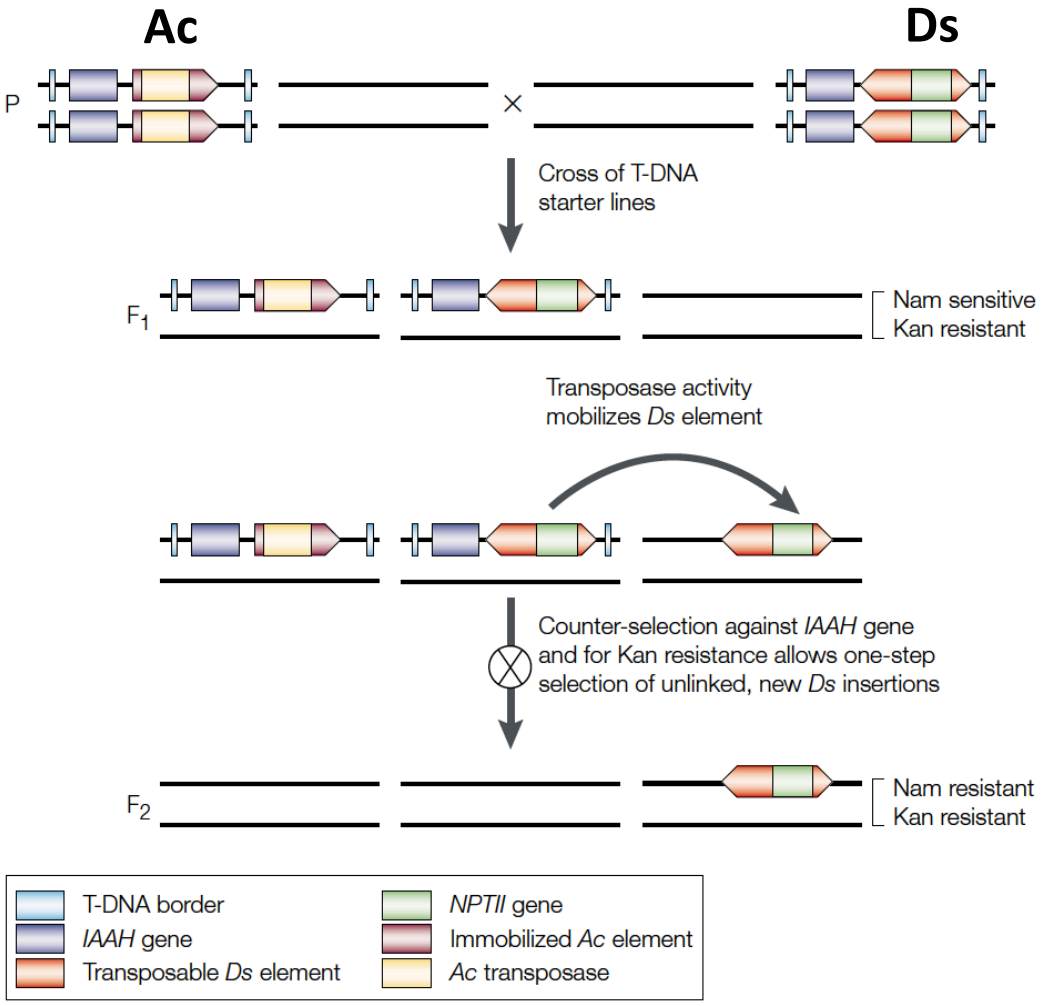


Silencing of transposons

- Transposons are found in almost all organisms
- For example, transposons make up approximately 50% of the human genome and up to 90% of the maize genome
- Most transposons are silenced by epigenetic modification or other ways. Most transposons are not actively move around the genome and change phenotypes

Example 1: Ac-Ds transposon to create insertional mutations in Arabidopsis

- Two transformed plants as parents, one has Ac insert and the other has Ds insert
- Cross the two parents and then self-pollinate F_1 hybrid to get many F_2 progenies
- What genotypes for the two loci Ac and Ds are in F_2 progenies?



Page and Grossniklaus, 2002

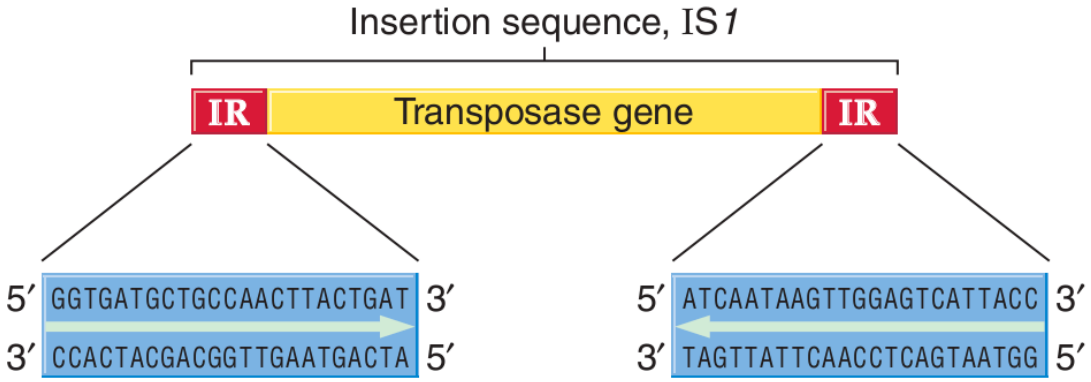
Genotypes of two loci Ac and Ds in F_2 population

- Parent1: Ac/Ac -/-; Parent2: -/- Ds/Ds
- **F_1: Ac/- Ds/-**
- Genotypes of Ac and Ds loci in the F_2 progenies

		Pollen			
		Ac Ds	Ac -	- Ds	- -
Egg	Ac Ds	Ac/Ac Ds/Ds	Ac/Ac Ds/-	Ac/- Ds/Ds	Ac/- Ds/-
	Ac -	Ac/Ac Ds/-	**Ac/Ac -/-**	Ac/- Ds/-	**Ac/- -/-**
	- Ds	Ac/- Ds/Ds	Ac/- Ds/-	**-/- Ds/Ds**	**-/- Ds/-**
	- -	Ac/- Ds/-	**Ac/- -/-**	**-/- Ds/-**	**-/- -/-**

A mini-transposon for insertional mutagenesis in the bacteria *S. pneumoniae*

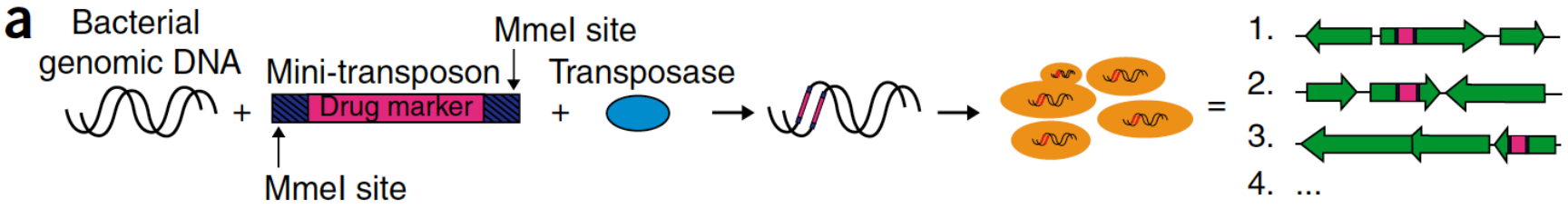
- The mini-transposon has ends with terminal inverted repeats (IRs) of 9 to 41 bp



- The artificial mini transposon contains Kan^R gene and *MmeI* restriction site within each inverted repeat

Example 2: Create a insertional mutation library of *S. pneumoniae* using mini-transposon

- A gene disruption library is constructed by first transposing the mini-transposon into bacterial genomic DNA **in vitro** and then transforming a bacterial population with the transposed DNA
- The transpose inserts randomly in the genome, requiring only a TA dinucleotide at the insertion site



van Opijnen et al., Nature methods 2009

Mutagen	Main characteristics
Chemical agents (e.g., EMS)	✓ Works for all species ✓ Mainly point mutations ✓ High efficiency, hundreds to hundreds of thousands of mutations per mutant ✓ Provides allelic series, and not just knockouts, which can yield refined insights into gene function
Physical agents (Fast neutrons, X-rays, etc.)	✓ Works for all species ✓ Break DNA and cause deletions ✓ Medium efficiency
Biological agents (e.g., T-DNA and transposons)	✓ Not work for all species ✓ Insertion of specific DNA sequence and cause loss of function of a gene ✓ Low efficiency, 1-3 mutations per line

What need to know for final exam

- Can describe and explain how chemical and insertional mutagens induce mutation, how to use those mutagens to create a population of mutants, characters of mutations induced by chemical, physical, and insertional mutagens
- Can describe synonymous mutation, missense mutation, and nonsense mutation and their differences

References

1. Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R. and Gadrinab, C., 2003. Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science*, *301*(5633), pp.653-657.
2. Alonso, J.M. and Ecker, J.R., 2006. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nature Reviews Genetics*, *7*(7), p.524.
3. Azpiroz-Leehan, R. and Feldmann, K.A., 1997. T-DNA insertion mutagenesis in Arabidopsis: going back and forth. *Trends in Genetics*, *13*(4), pp.152-156.
4. Clough, S.J. and Bent, A.F., 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The plant journal*, *16*(6), pp.735-743.
5. Feldmann, K.A. and Marks, M.D., 1987. Agrobacterium-mediated transformation of germinating seeds of Arabidopsis thaliana: a non-tissue culture approach. *Molecular and General Genetics MGG*, *208*(1-2), pp.1-9.
6. Forsthoefel, N.R., Wu, Y.W., Schulz, B., Bennett, M.J. and Feldmann, K.A., 1992. T-DNA insertion mutagenesis in Arabidopsis: prospects and perspectives. *Functional Plant Biology*, *19*(4), pp.353-366.

References

7. Henikoff, S. and Comai, L., 2003. Single-nucleotide mutations for plant functional genomics. *Annual Review of Plant Biology*, *54*(1), pp.375-401.
8. Henry, I.M., Nagalakshmi, U., Lieberman, M.C., Ngo, K.J., Krasileva, K.V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., Tai, T.H. and Comai, L., 2014. Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *The Plant Cell*, *26*(4), pp.1382-1397.
9. Huang, M., Bai, Y., Sjostrom, S.L., Hallström, B.M., Liu, Z., Petranovic, D., Uhlén, M., Joensson, H.N., Andersson-Svahn, H. and Nielsen, J., 2015. Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proceedings of the National Academy of Sciences*, *112*(34), pp.E4689-E4696.
10. Jiao, Y., Burke, J.J., Chopra, R., Burow, G., Chen, J., Wang, B., Hayes, C., Emendack, Y., Ware, D. and Xin, Z., 2016. A sorghum mutant resource as an efficient platform for gene discovery in grasses. *The Plant Cell*, pp.tpc-00373.
11. Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R.H., Wang, X., Borrill, P. and Fosker, C., 2017. Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences*, *114*(6), pp.E913-E921.
12. Krysan, P.J., Young, J.C. and Sussman, M.R., 1999. T-DNA as an insertional mutagen in Arabidopsis. *The Plant Cell*, *11*(12), pp.2283-2290.

References

13. Krysan, P.J., Young, J.C., Tax, F. and Sussman, M.R., 1996. Identification of transferred DNA insertions within Arabidopsis genes involved in signal transduction and ion transport. *Proceedings of the National Academy of Sciences*, *93*(15), pp.8145-8150.
14. Li, G., Jain, R., Chern, M., Pham, N.T., Martin, J.A., Wei, T., Schackwitz, W.S., Lipzen, A.M., Duong, P.Q., Jones, K.C. and Jiang, L., 2017. The sequences of 1504 mutants in the model rice variety Kitaake facilitate rapid functional genomic studies. *The Plant Cell*, *29*(6), pp.1218-1231.
15. McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S., 2000. Targeting induced locallesions in genomes (TILLING) for plant functional genomics. *Plant physiology*, *123*(2), pp.439-442.
16. Page, D.R. and Grossniklaus, U., 2002. The art and design of genetic screens: Arabidopsis thaliana. *Nature Reviews Genetics*, *3*(2), p.124.
17. Tzfira, T. and Citovsky, V., 2002. Partners-in-infection: host proteins involved in the transformation of plant cells by Agrobacterium. *Trends in cell biology*, *12*(3), pp.121-129.
18. Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y., Sundin, L., Goeminne, G., Kim, H., Cross, J., Morreel, K., Araujo, P. and Welsh, L., 2013. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science*, *341*(6150), pp.1103-1106.
19. van Opijnen, T., Bodi, K.L. and Camilli, A., 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature methods*, *6*(10), pp.767-772.