Outline

- Resources, approaches, technologies, and tools used in functional genomics studies
  - Induced mutation
  - Gene silencing
  - Site-specific mutation
  - QTL/Gene mapping and clone
  - Gene expression profiling
  - Metabolite profiling

Mutation

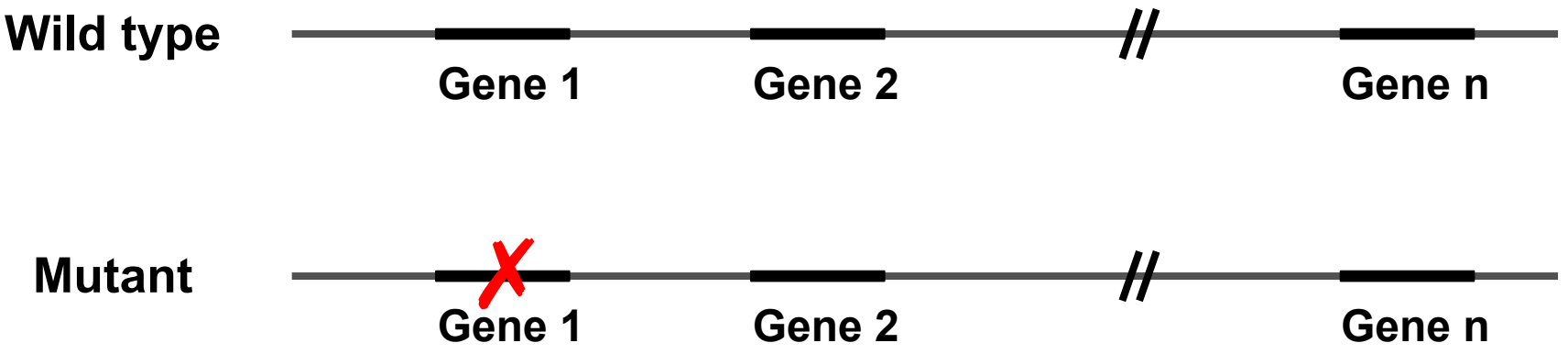
- **Mutations** are changes in DNA sequence
  - Large-scale chromosomal structure variation
    - ✓ Deletion, insertion, duplication, and inversion of a chromosome fragment
  - Small-scale mutation
    - ✓ One or a few of nucleotides change, insertion, or deletion

Mutation

- Mutations can result from DNA copying mistakes during cell division, exposure to ionizing radiation or 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
- Mutation is a main cause of diversity among organisms and individuals

Induced mutations are valuable resources to study gene function

- **Mutations** can be induced by treating a organism with a mutagen
- **Induced mutations** of a gene 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 gene.



Three major types of mutagens induce mutations randomly distributed on genome

- Mutagen types
  - Insertional mutagen
  - Physical mutagen
  - Chemical mutagen

✓ How to use mutagens to create mutations?

✓ Characters of induced mutations

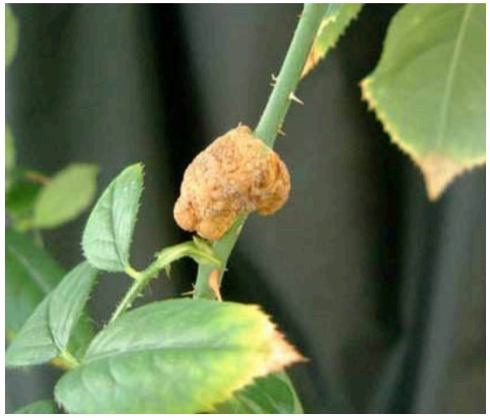
✓ How to identify mutation site in a mutant?

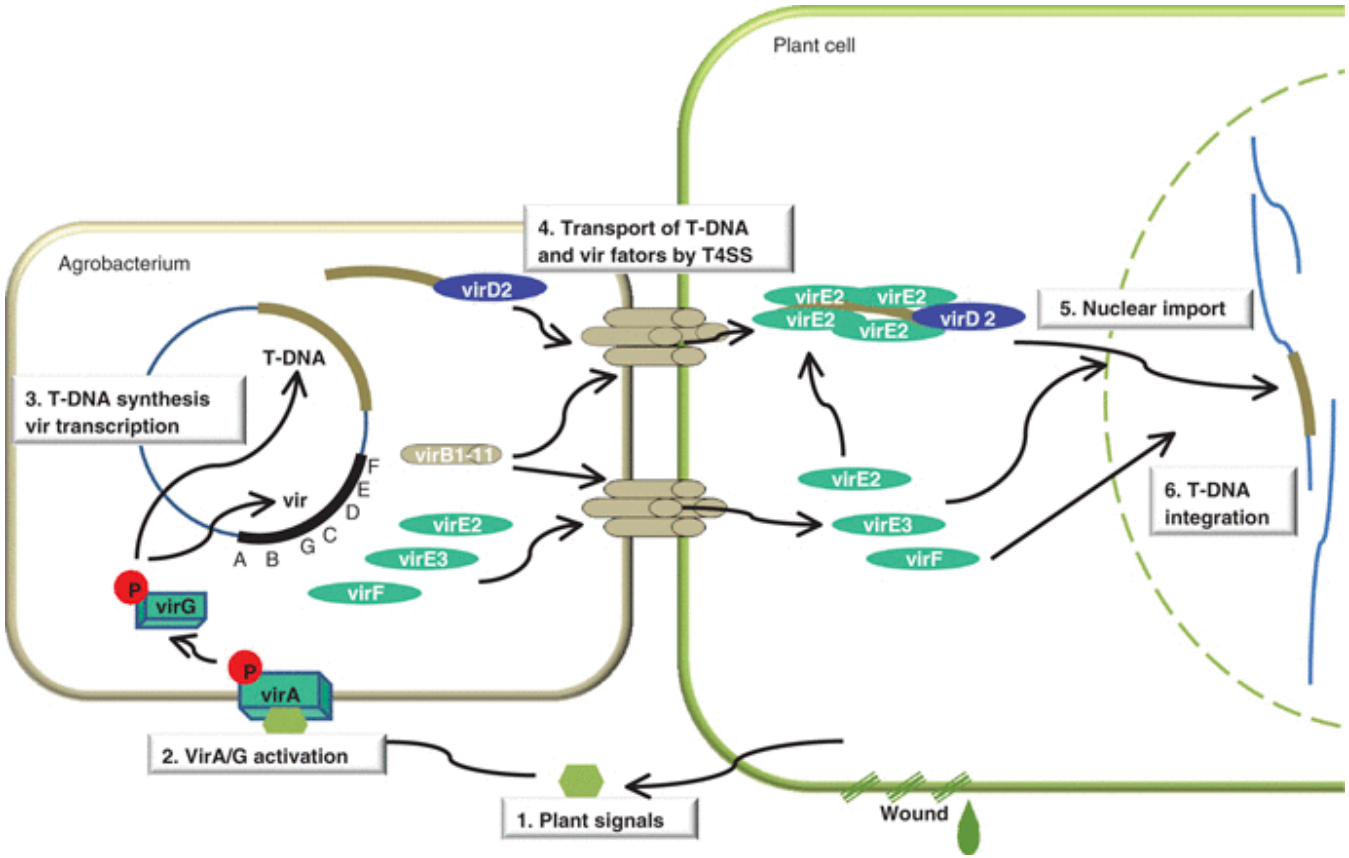
✓ How to use the induced mutations to study gene function

Insertional mutagen

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

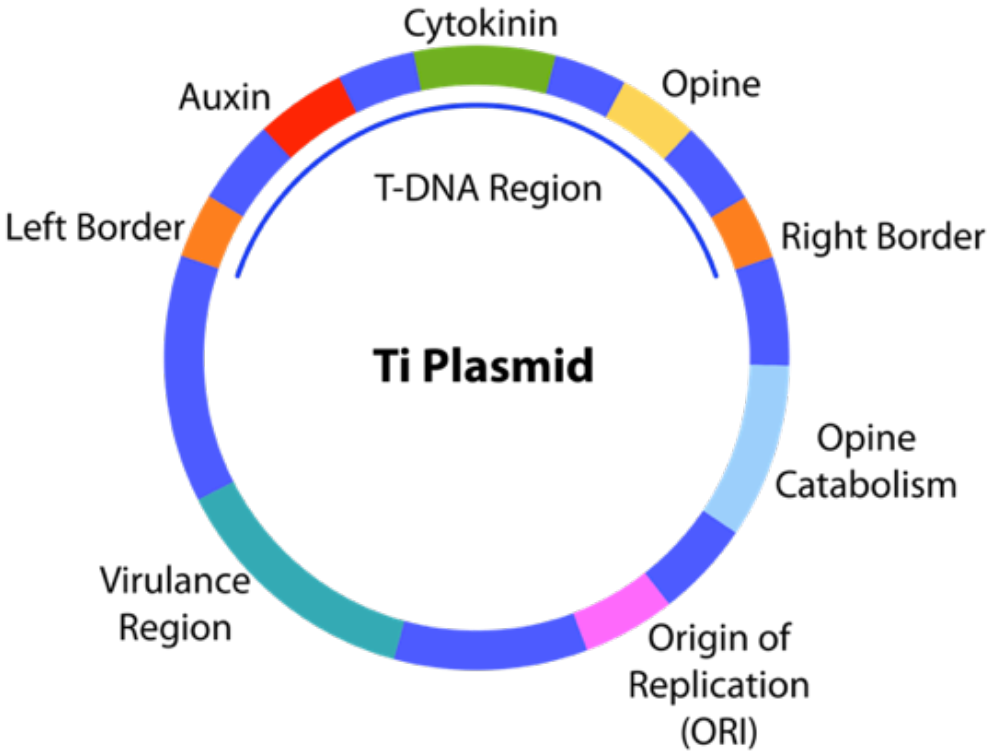
T-DNA inserts into plant genome from agrobacterium

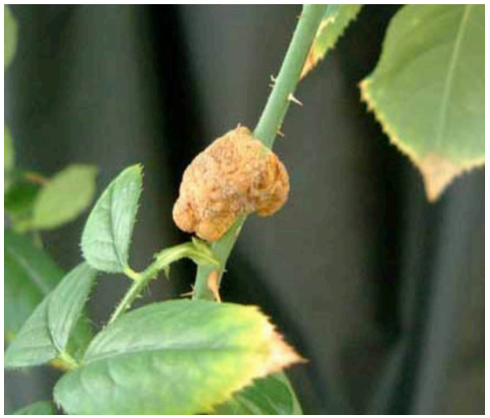




Tzfira and Citovsky, Trends in cell biology 2002

Plasmid DNA and T-DNA structure

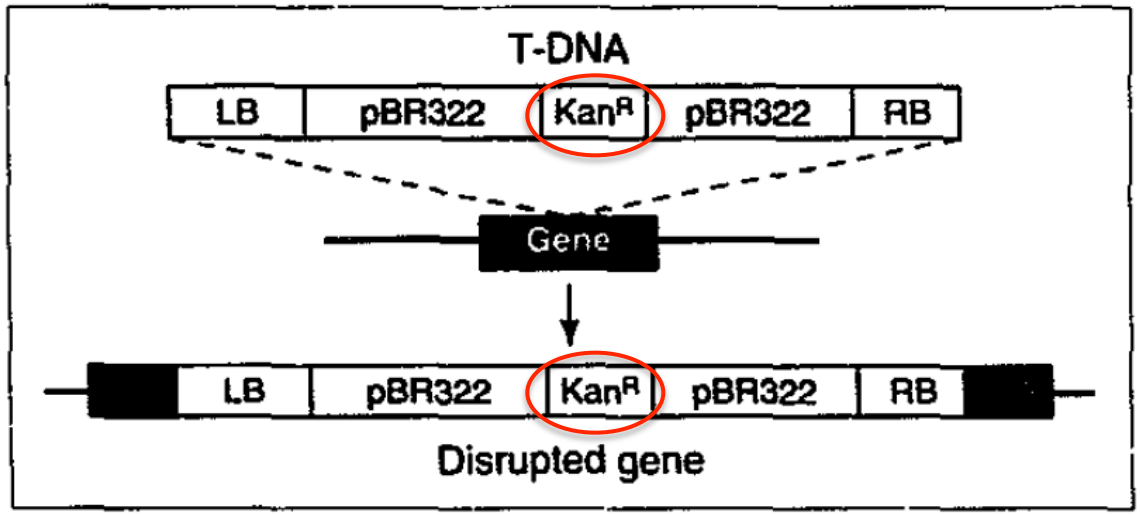




- Tumor-promoting gene: Auxin and Cytokinin are plant hormones that enable the plant cell grow uncontrollably, thus forming the tumors
- Biosynthetic genes: opine is amino acid derivatives used by the bacterium as a source of carbon and energy

Agrobacterium as insertional mutagenesis

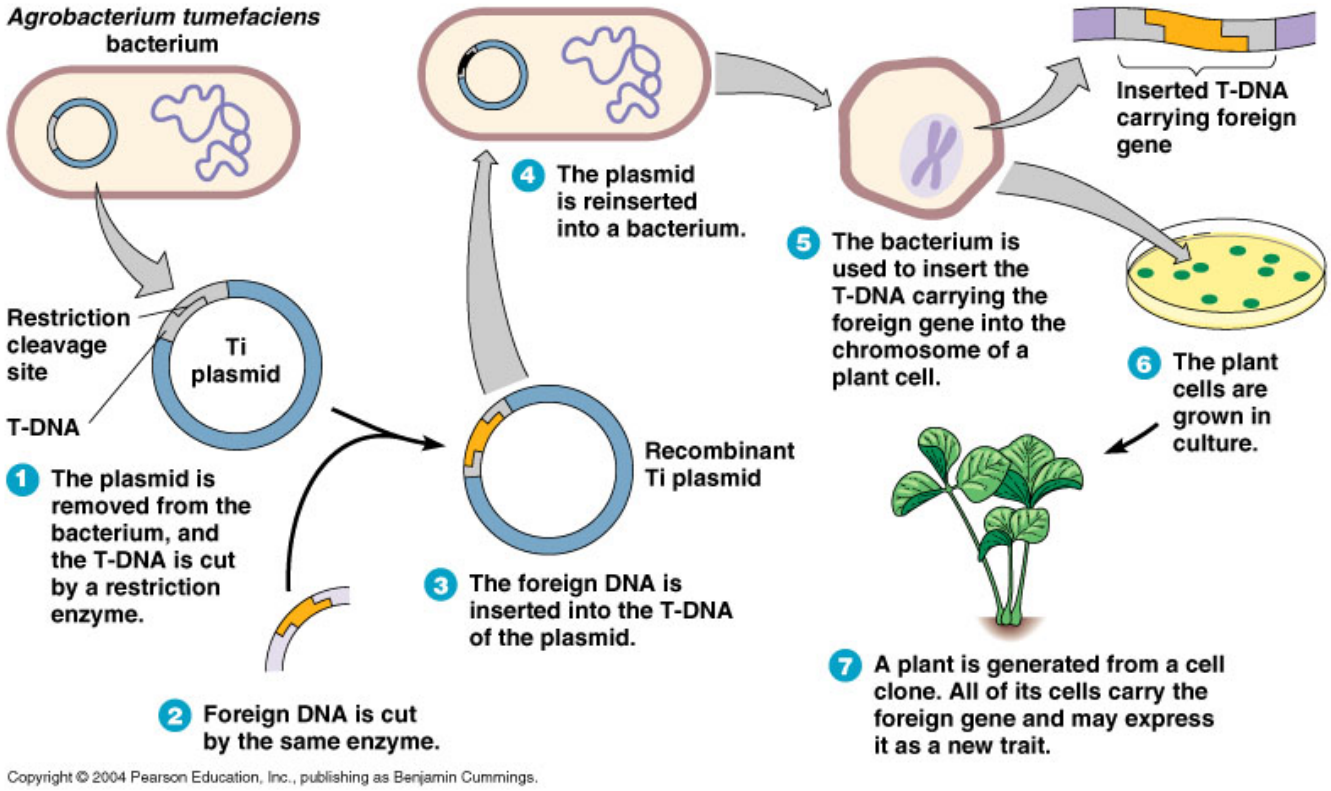
- Insertional mutations can be artificially created in the lab
- Kan^R gene confers kanamycin resistance, allowing selection of the transformed plants (or mutants)
- The length of the insert is 17 kb in this case, **causing loss of gene function**



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

How many insertional mutants do we need to study all genes in Arabiposis

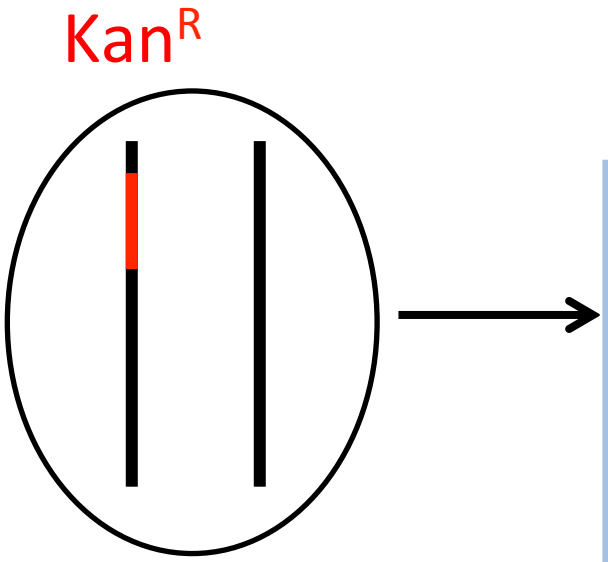
- Arabidopsis, a plant model species
  - Genome size: ~120,000 Kbp
  - Gene number: ~30,000
  - Average gene size: ~2 Kbp

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

➢ How many T-DNA inserts in one mutant?

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 a 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 mutant
  - Kan^R : -/- = 15:1
- Three inserts in one mutant
  - Kan^R : -/- ??

❖ **Average 1.5 inserts per Arabidopsis mutant**

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: $\underline{p = 1-(1-x/120000)^n}$

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

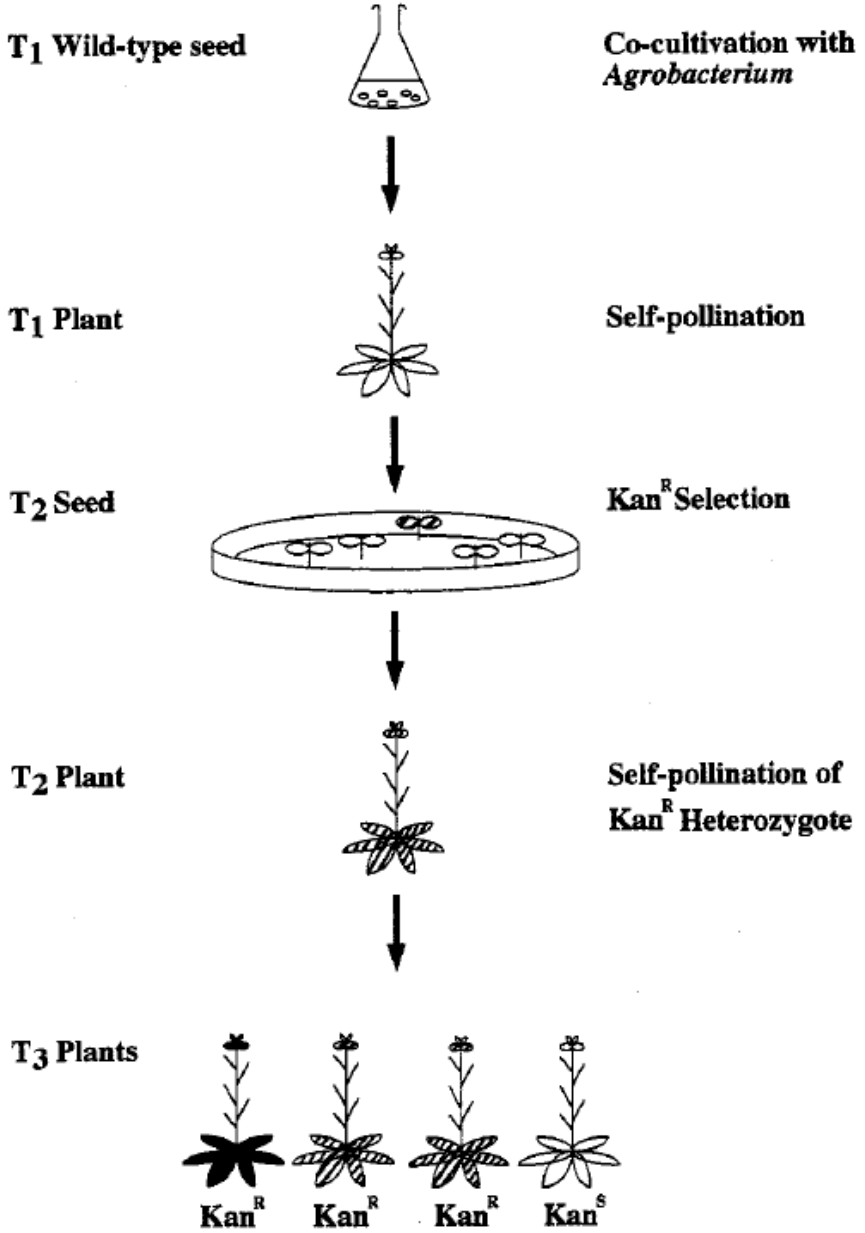
$$p = 1 - (1 - (x / 120{,}000))^n$$

- 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

Agrobacterium-mediated transformation
--Seed infection method

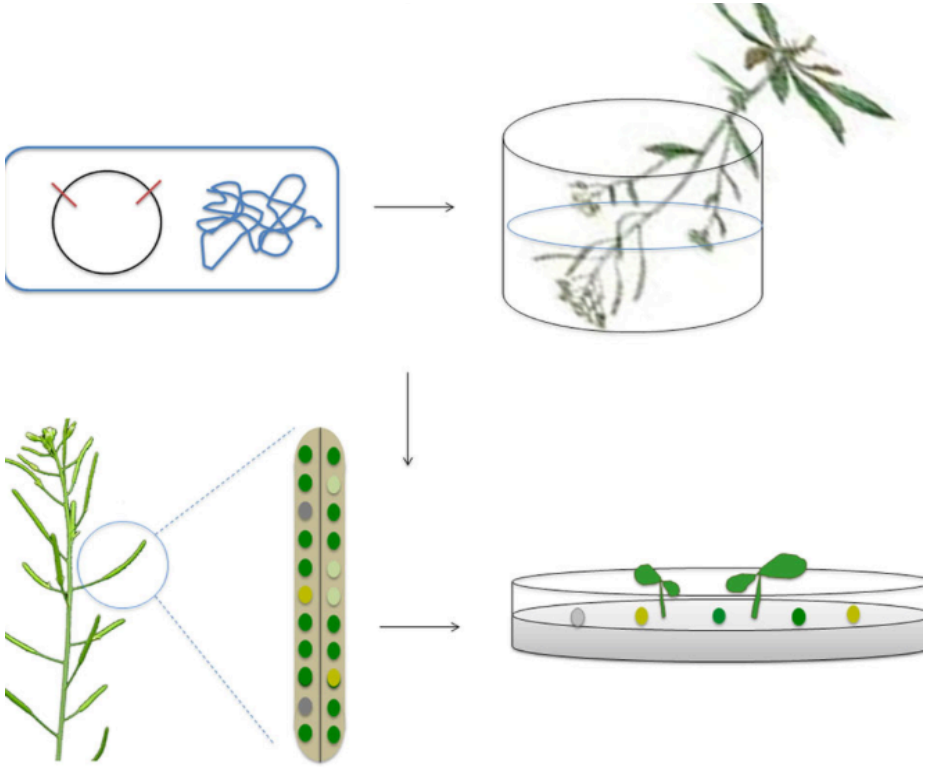
- If a cell with inserts in a germinated seed forms reproductive tissues of the T1 plant, then some T2 seeds from the T1 plant have inserts



Forsthoefel et al., 1992

Agrobacterium-mediated transformation --Flora dip method

- Dipping flowering plants in *Agrobacterium* that are suspended in a solution
- Collect T2 seeds from the treated plants
- Plant T2 seeds on medium containing kanamycin
- Transformed Kan-resistant plants with green cotyledons can be identified
- Kanamycin-resistant seedlings are transferred to soil and grow to mature to collect T3 seeds



Clough et al., The Plant Journal 1998

T-DNA insertion collections in Arabidopsis

Institution	Population size	Genotype	Reference
Salk Institute	150,000	Columbia-0	Alonso et al., 2003
Bielefeld University, Germany	71,000	Columbia-0	Kleinboelting et al., 2012
Syngenta	100,000	Columbia	Sessions et al., 2002

Reverse genetics using induced insertional mutations to study a gene's function

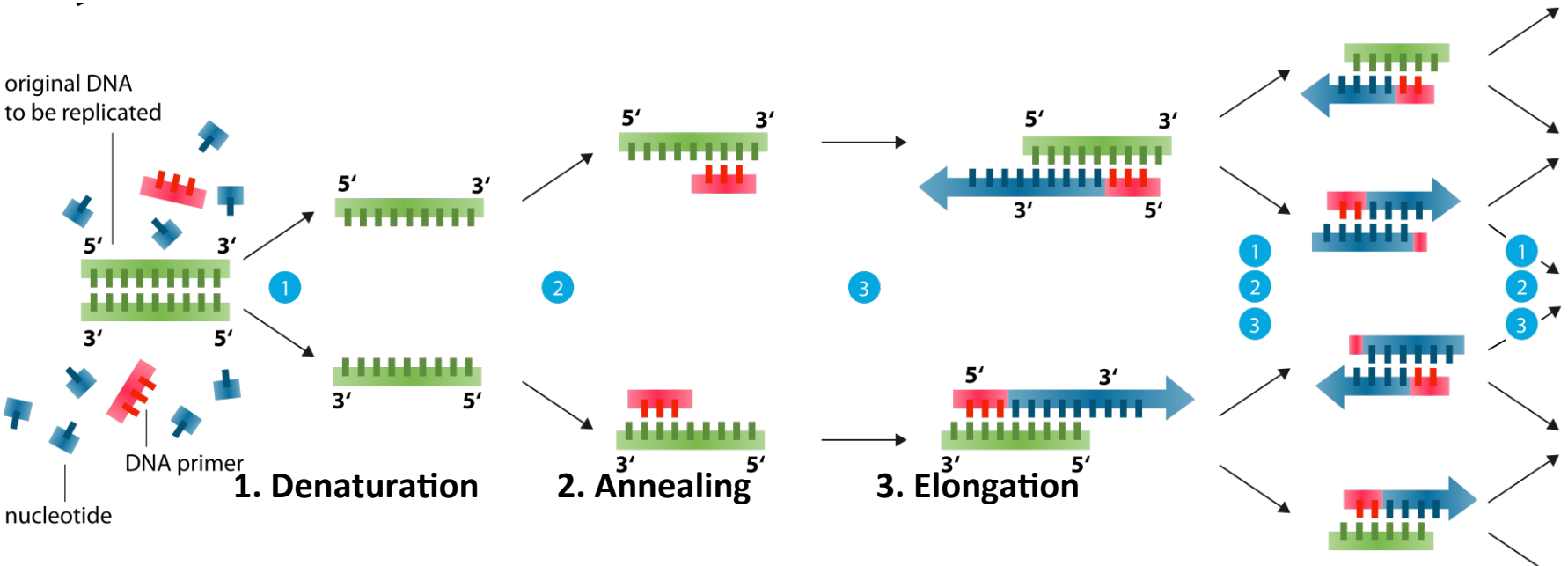
1. A candidate gene with known sequence → 2. Identify the mutant for the candidate gene → 3. Evaluate phenotype of the mutants → 4. Define gene function

- Induced insertional mutations are randomly distributed along the genome
- Identifying a mutant of the candidate gene needs to screen many mutants

➢ **Gene specific <u>PCR</u> followed by <u>gel analysis</u> is a method to identify an insertional mutant for a candidate gene**

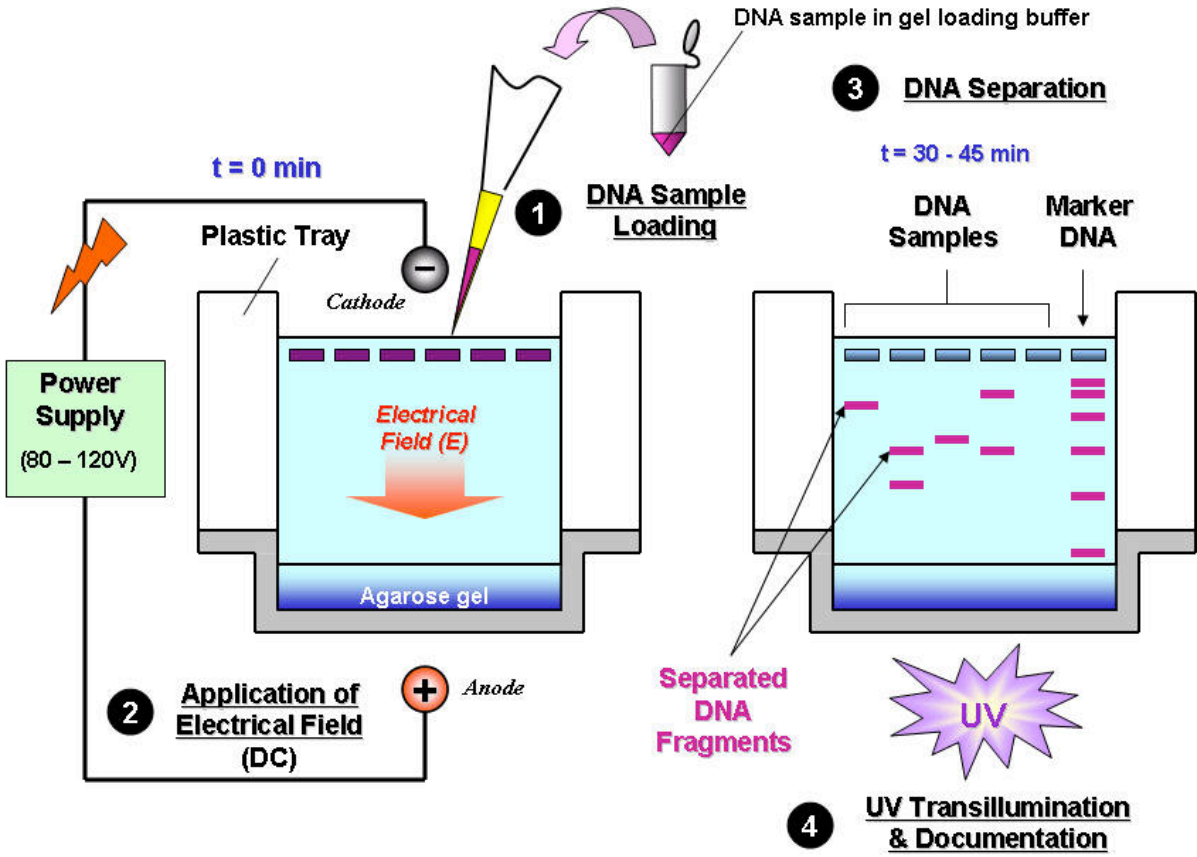
Polymerase chain reaction (PCR)

- A technique of amplifying a few copies of a piece of DNA to generate thousands to millions of copies
- A pair of primers to amplify a DNA fragment between the two primers; amplify DNA fragment of 0.1 to 10 Kbp.
- https://dnalc.cshl.edu/resources/3d/19-polymerase-chain-reaction.html



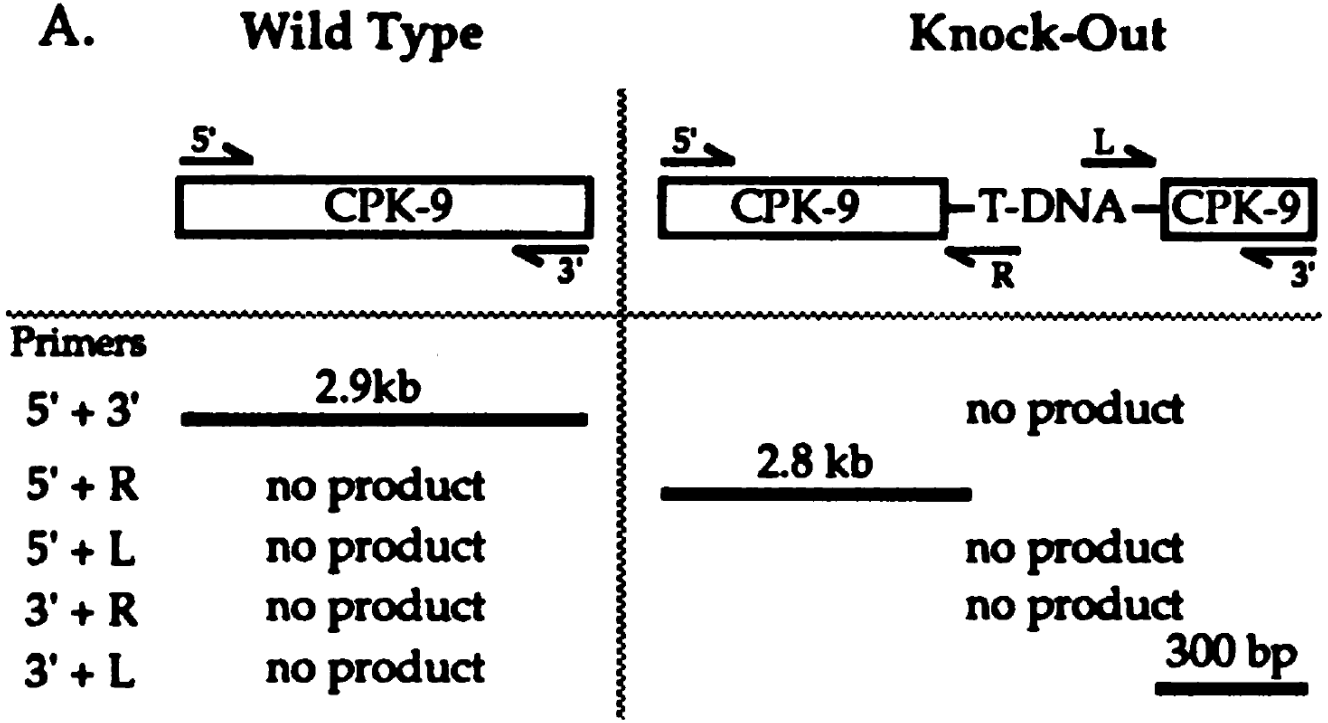
Gel analysis to detect DNA fragments

- **Gel electrophoresis** is a method for separation and analysis of DNA, protein, and other molecules based on their fragment size
  - https://dnalc.cshl.edu/resources/animations/gelelectrophoresis.html



Example: Identification of a T-DNA insertion within a gene CPK-9

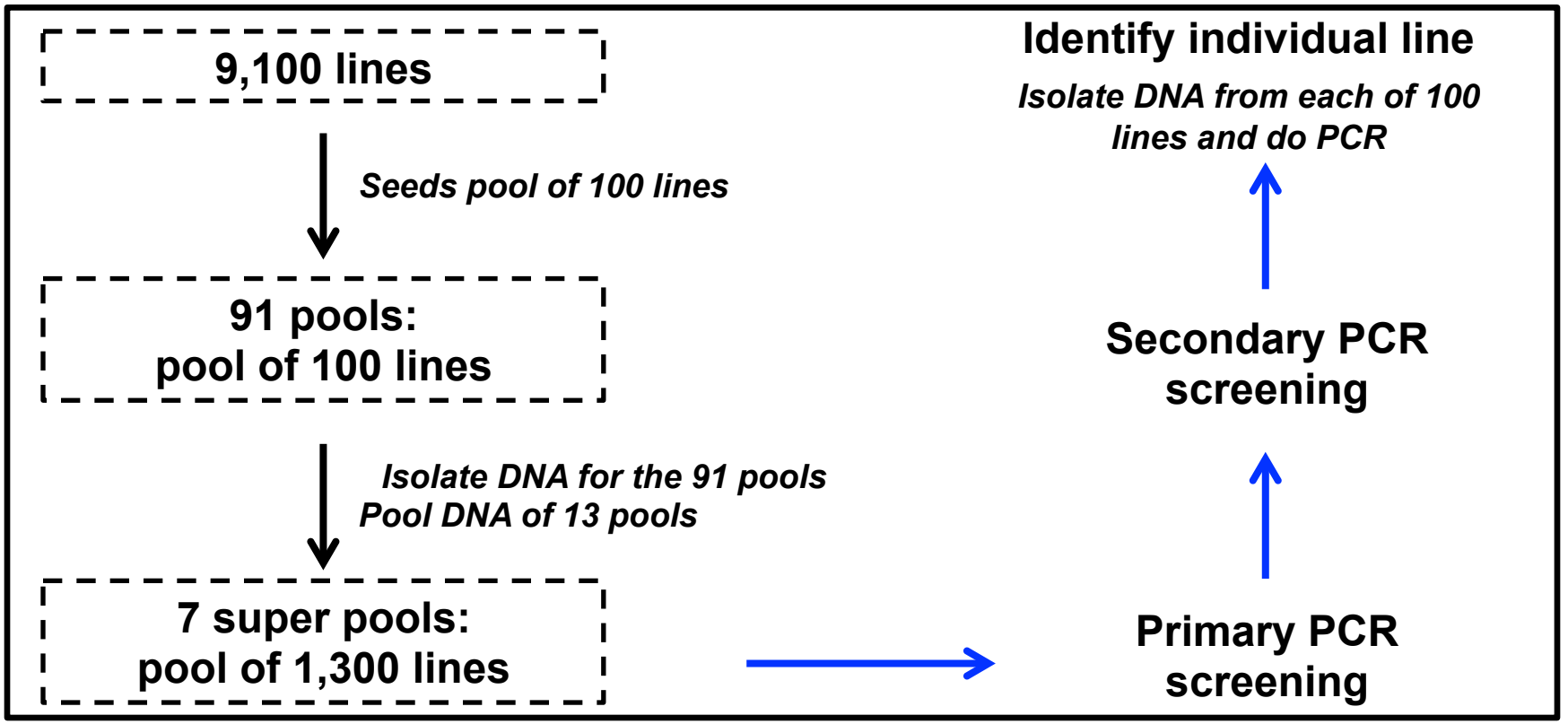
- A gene specific PCR to find mutation lines for a known target gene. Five different combinations of PCR primers can be used (5'+3', 5'+L, 5'+R, 3'+L, and 3'+R)
- The T-DNA is 17 Kbp long, PCR could generate fragments from 0.1 to 10 Kbp long



Krysan et al., PNAS 1996

Pool strategy to reduce work of DNA preparation and PCR screening

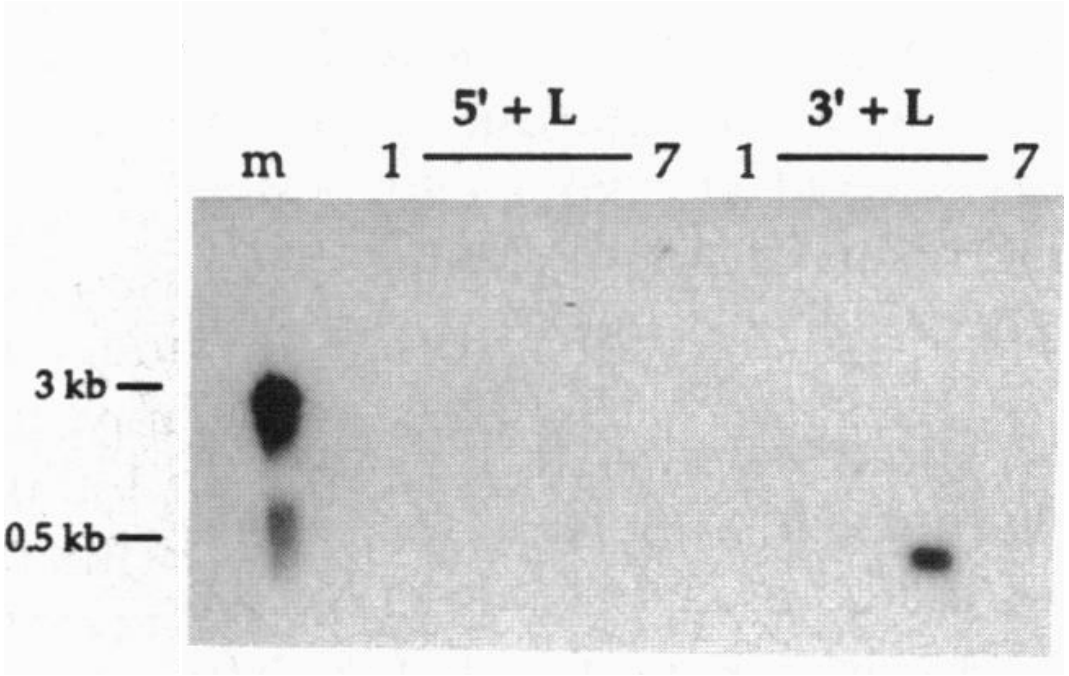
- Single mutant screening, 9,100 DNA isolation and PCR
- Using pool strategy, 191 DNA isolations and 121 PCR screens



Krysan et al., PNAS 1996

Example: Identification of a T-DNA insertion within a gene CPK-9

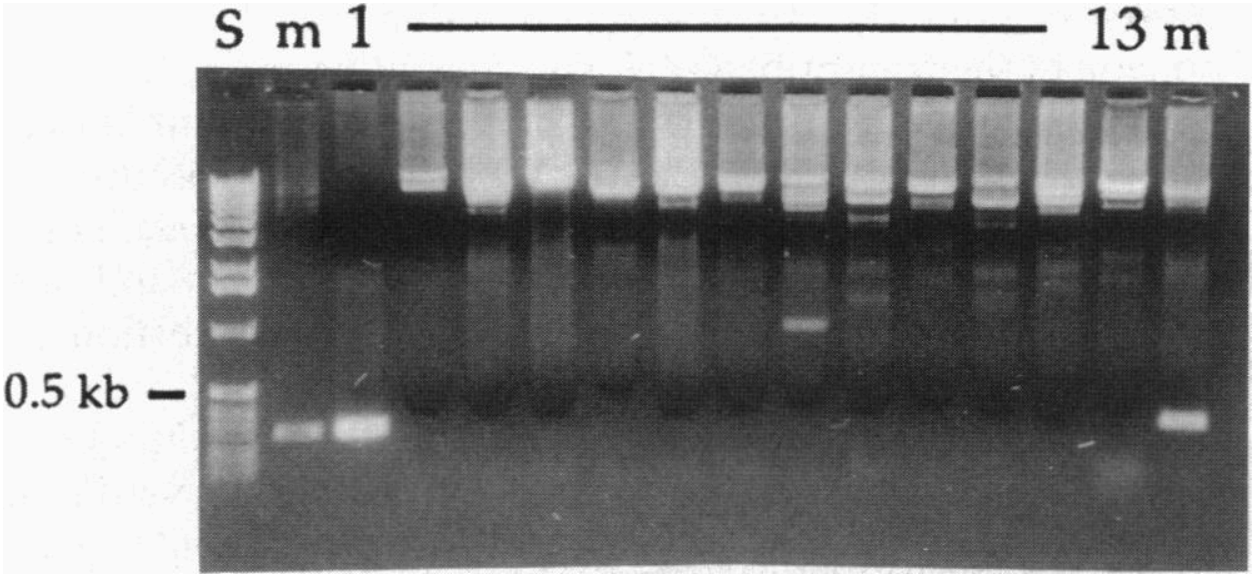
- Screen the seven super pools
  - Four different combinations of PCR primers can be used (**5'+L**, 5'+R, **3'+L**, and 3'+R)
  - Super pool 5 contained a plant carrying a T-DNA within the CPK-9 gene



Krysan et al., PNAS 1996

Example: Identification of a T-DNA insertion within a gene CPK-9

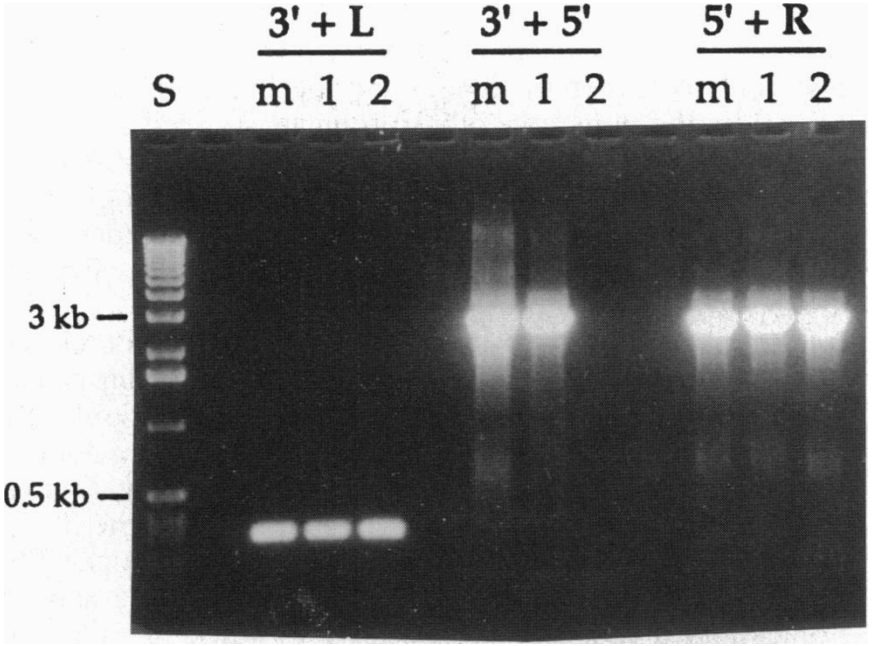
- Screen the 13 pools of 100 within the super pool 5 using primers 3'+L
  - Lanes 1-13 correspond to these 13 pools
  - S, DNA ladder size standard
  - m, amplification of super pool 5 using primers 3' + L

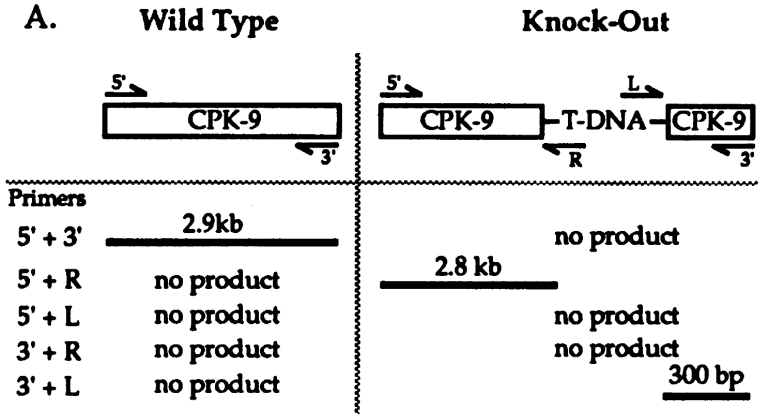


Krysan et al., PNAS 1996

Example: Identification of a T-DNA insertion within a gene CPK-9

- Screen the 100 individual plants within the pool 1
  - Lanes 1 and 2 correspond to these two individual plants found to carry T-DNA inserts within CPK-9
  - m, amplified using DNA from pool of 100 plants in pool 1
  - S, DNA ladder size standard

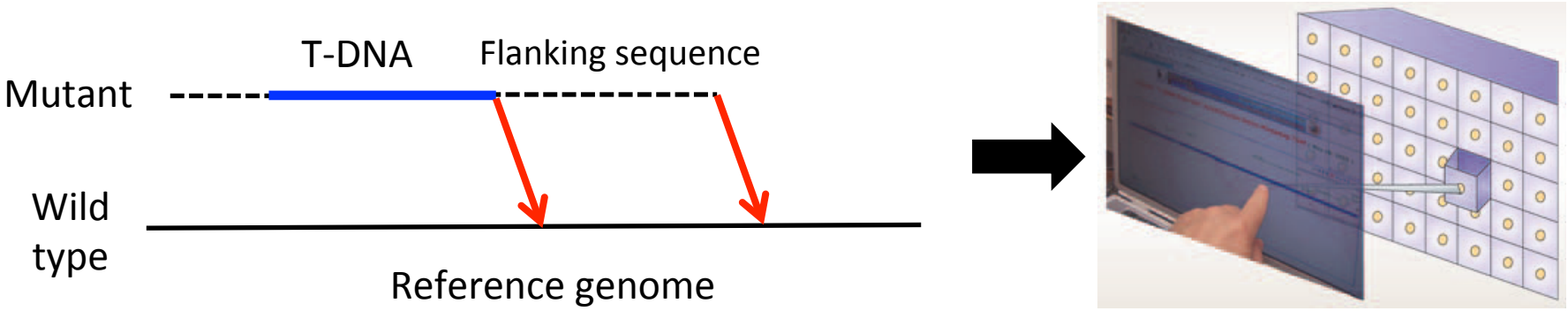




Krysan et al., PNAS 1996

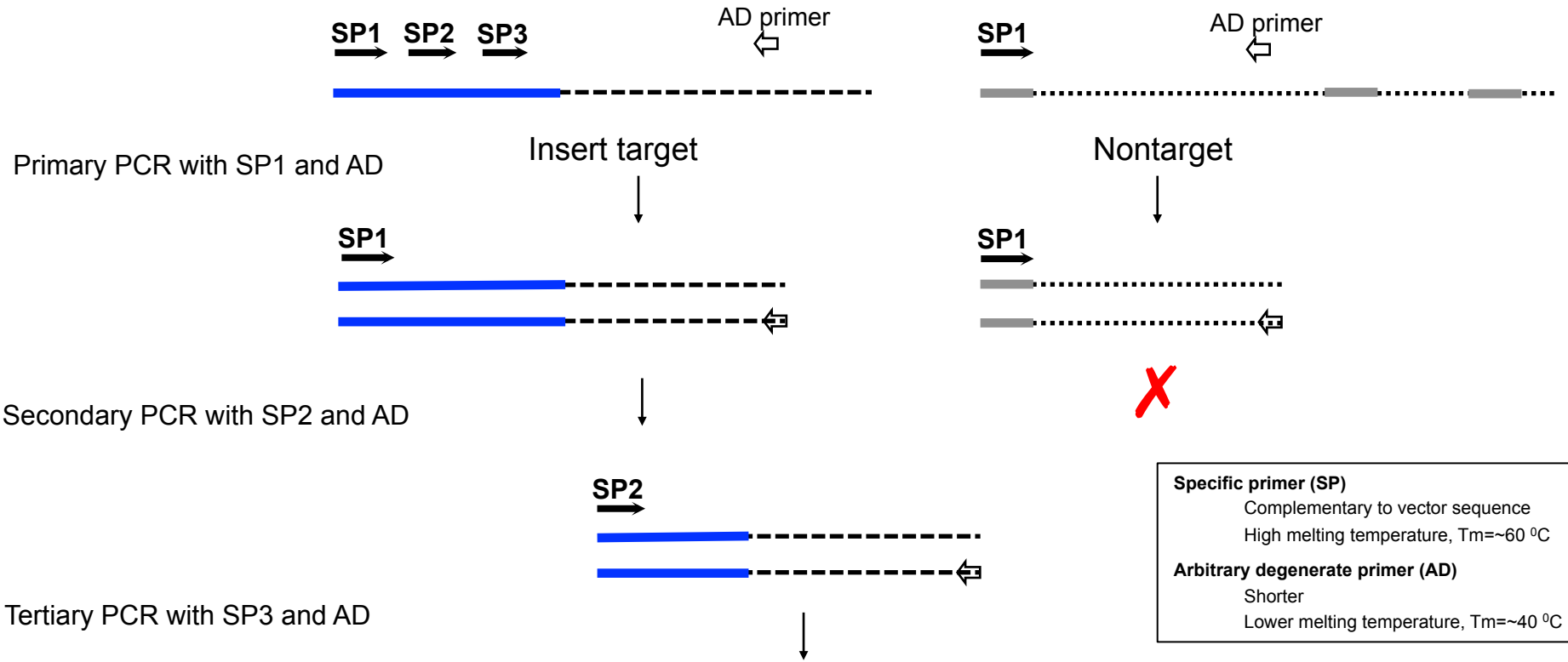
Gene-indexed catalogue of insertional mutants

- In traditional reverse genetics, PCR/Gel screening of thousands of insertional mutants has to be repeated for identifying a target mutant
- A gene-indexed catalogue, where all mutants are characterized, will allow scientists to easily find their target mutants
- A method to locate T-DNA inserts in all mutants is to get flanking sequences of the insertion site and then align the flanking sequences to reference genome



Thermal asymmetric interlaced PCR (TAIL-PCR) to obtain flanking sequences

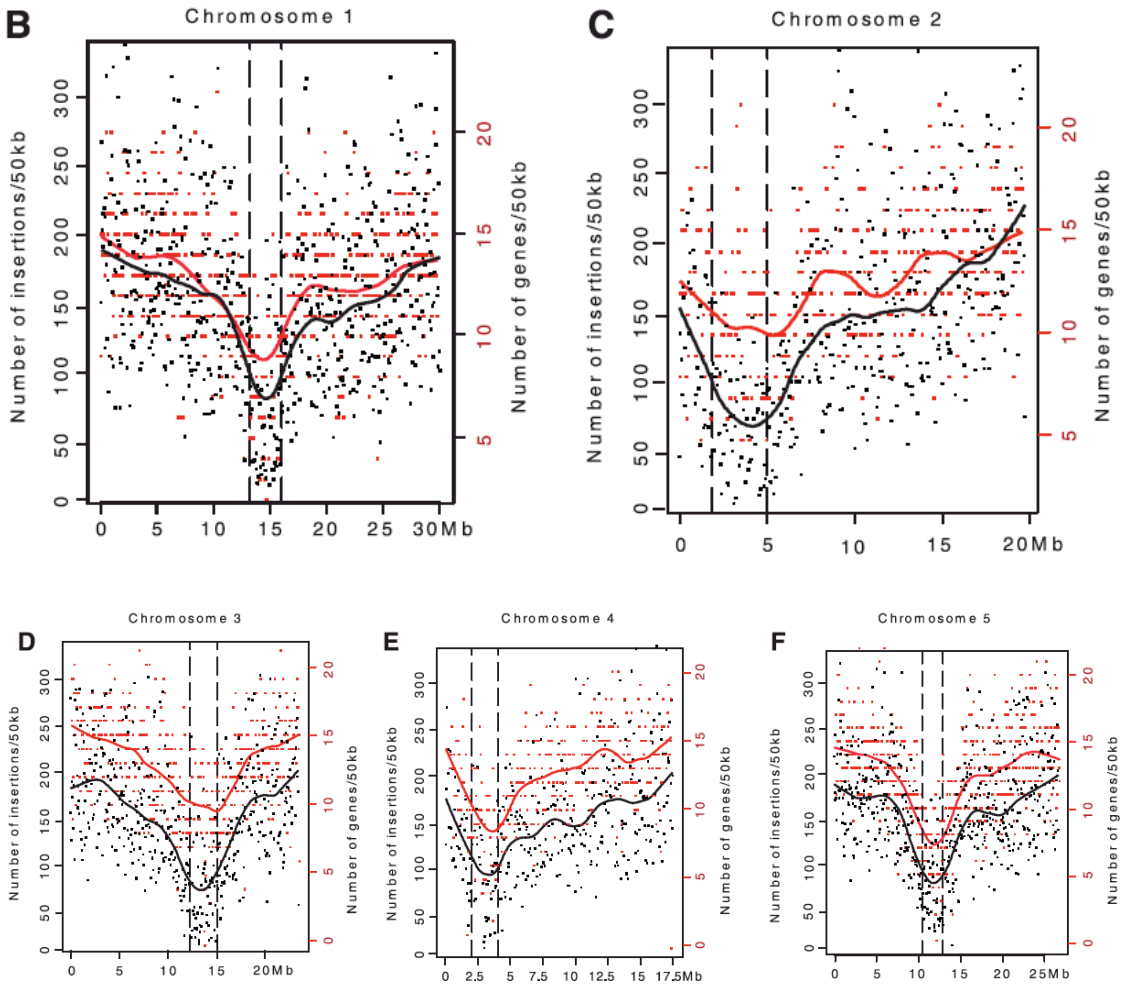
TAIL-PCR: Three PCR reactions are carried out sequentially to amplify target sequences using nested T-DNA-specific primers (SP1-3) on one side and an AD primer on the other side



Liu et al., Plant Journal 1995

Example: A gene-indexed catalogue of insertion mutants in Arabidopsis

- Screened 127,706 mutants using TAIL-PCR
- 88,122 insertion sites identified in 21,799 genes
- Fewer T-DNA integration events were consistently observed in regions surrounding each of the five centromeres



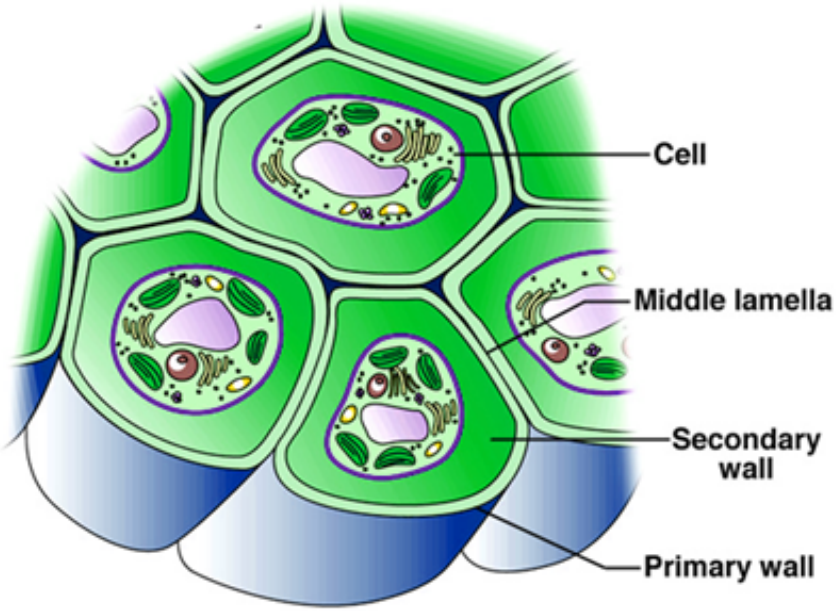
Alonso et al., Science 2003

Reverse genetics using induced insertional mutations to study a gene's function

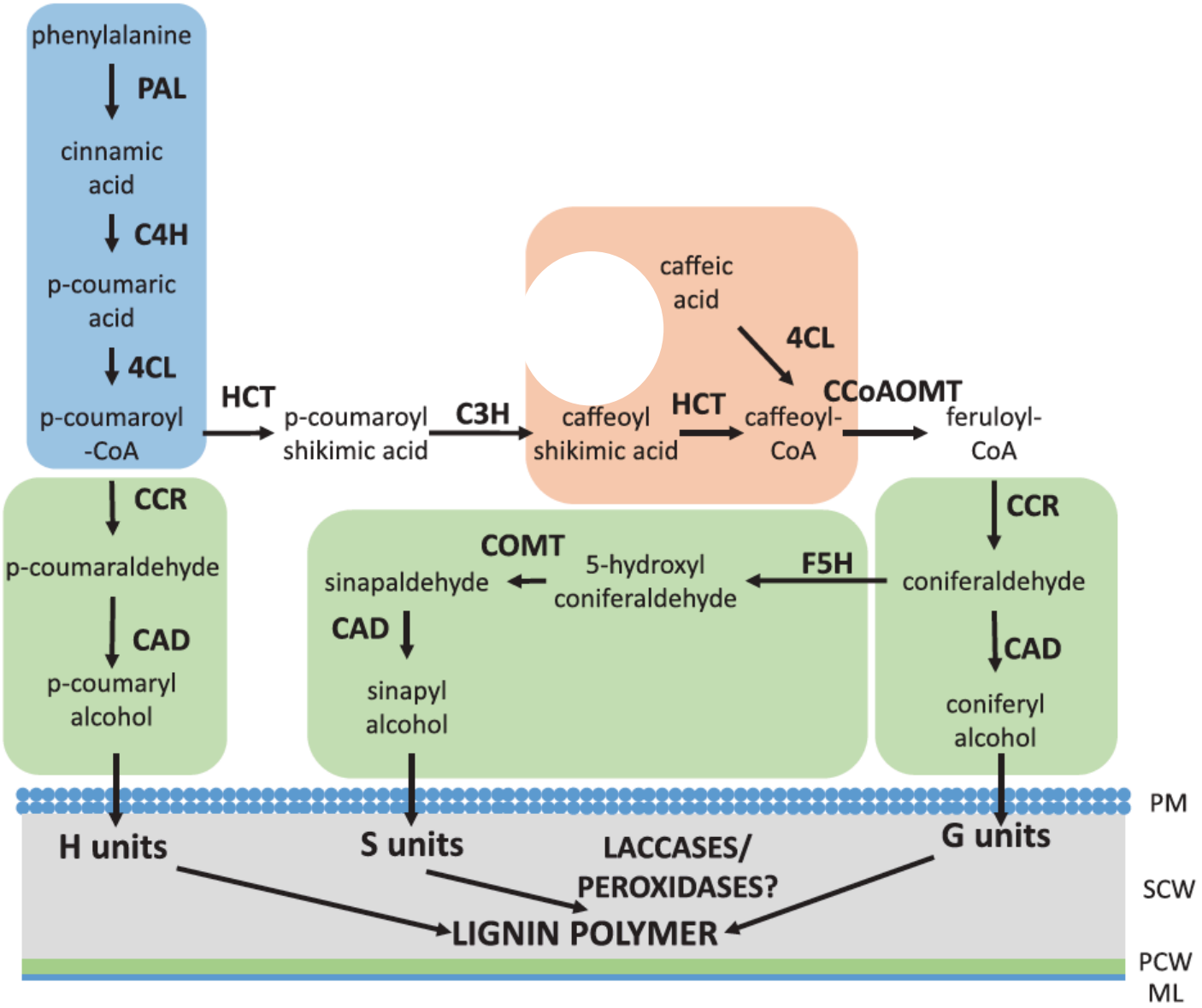
1. A candidate gene with known sequence → 2. Identify the mutant for the candidate gene → 3. Evaluate phenotype of the mutants → 4. Define gene function

An example of reverse genetics using induced insertional mutation

- Secondary cell walls found in some cells like stem, primary cell wall found in all plant cell
- Secondary cell walls provide additional protection to cells and rigidity and strength to the plants
- Primary cell wall: consist of cellulose, hemicellulose, and pectin
- Secondary cell wall: consist of cellulose, hemicellulose, and **lignin**
- **Lignin** makes the secondary cell wall less flexible and less permeable to water; **lignin** gives strength to stems and imparts hydrophobicity to vascular elements for water transport



Lignin biosynthetic pathway



Caffeoyl Shikimate Esterase (CSE) considered as a candidate gene

- CSE as a candidate for involvement in lignification, based on analyses designed to identify genes co-expressed with known components of the lignin biosynthetic pathway
- CSE was also identified as being co-expressed with lignin pathway genes in a set of lignin mutants

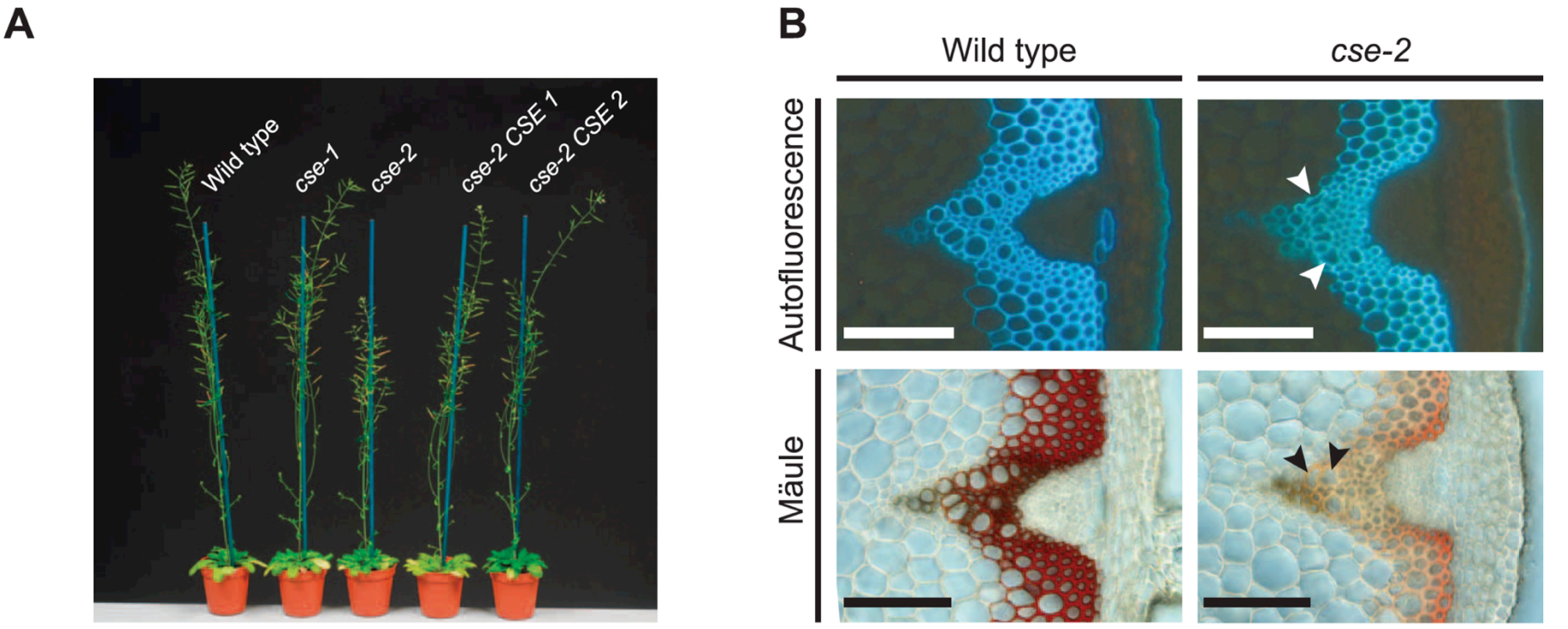
Vanholme et al., Science, 2013

Two T-DNA insertional mutants for the CSE

- **cse-1**, a knockdown mutant with an insertion in the promoter
- **cse-2**, a knockout mutant with an insertion in the second exon

Vanholme et al., Science, 2013

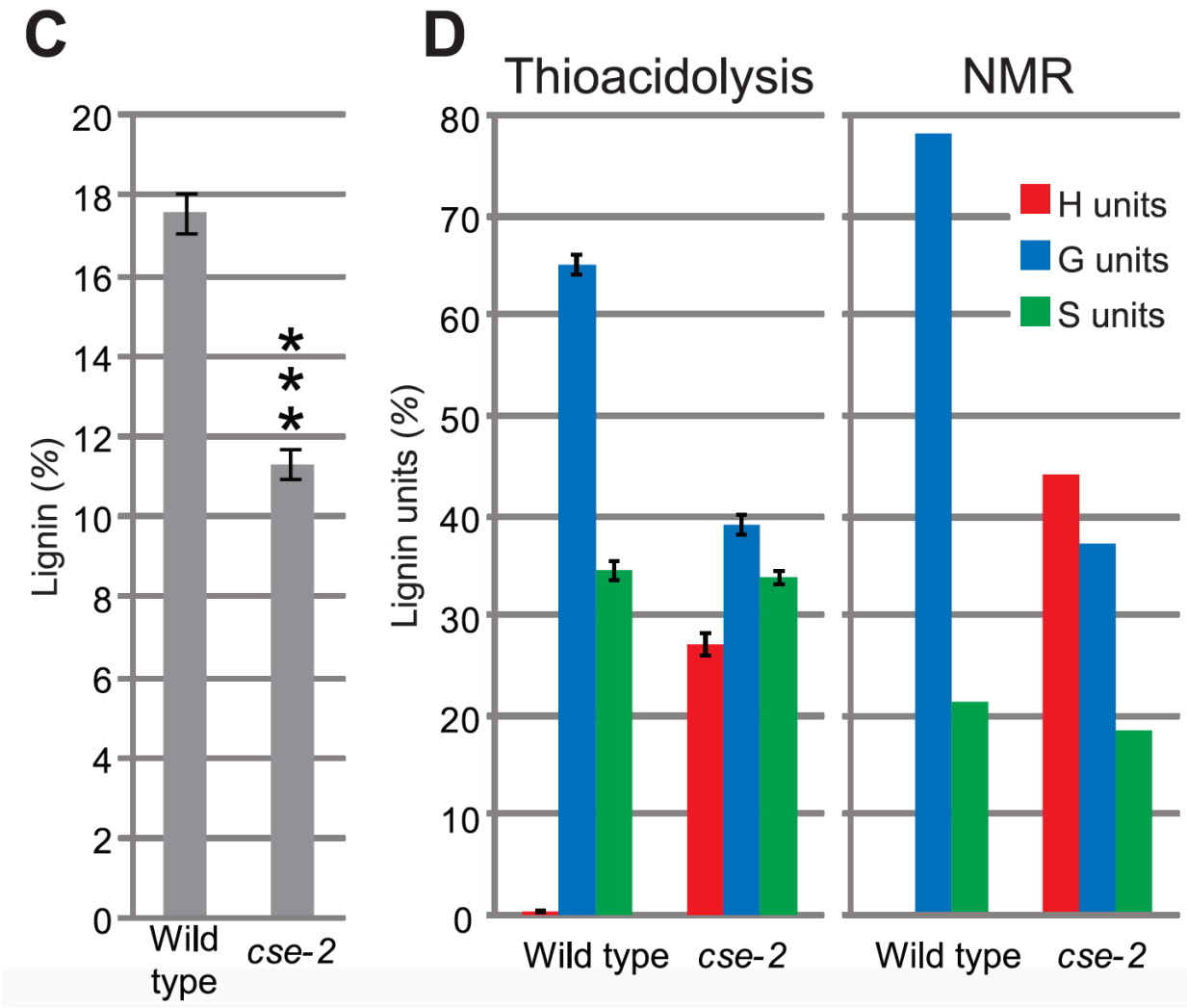
Phenotypic analysis of CSE mutants



Vanholme et al., Science, 2013

Compositional analysis of the CES mutants

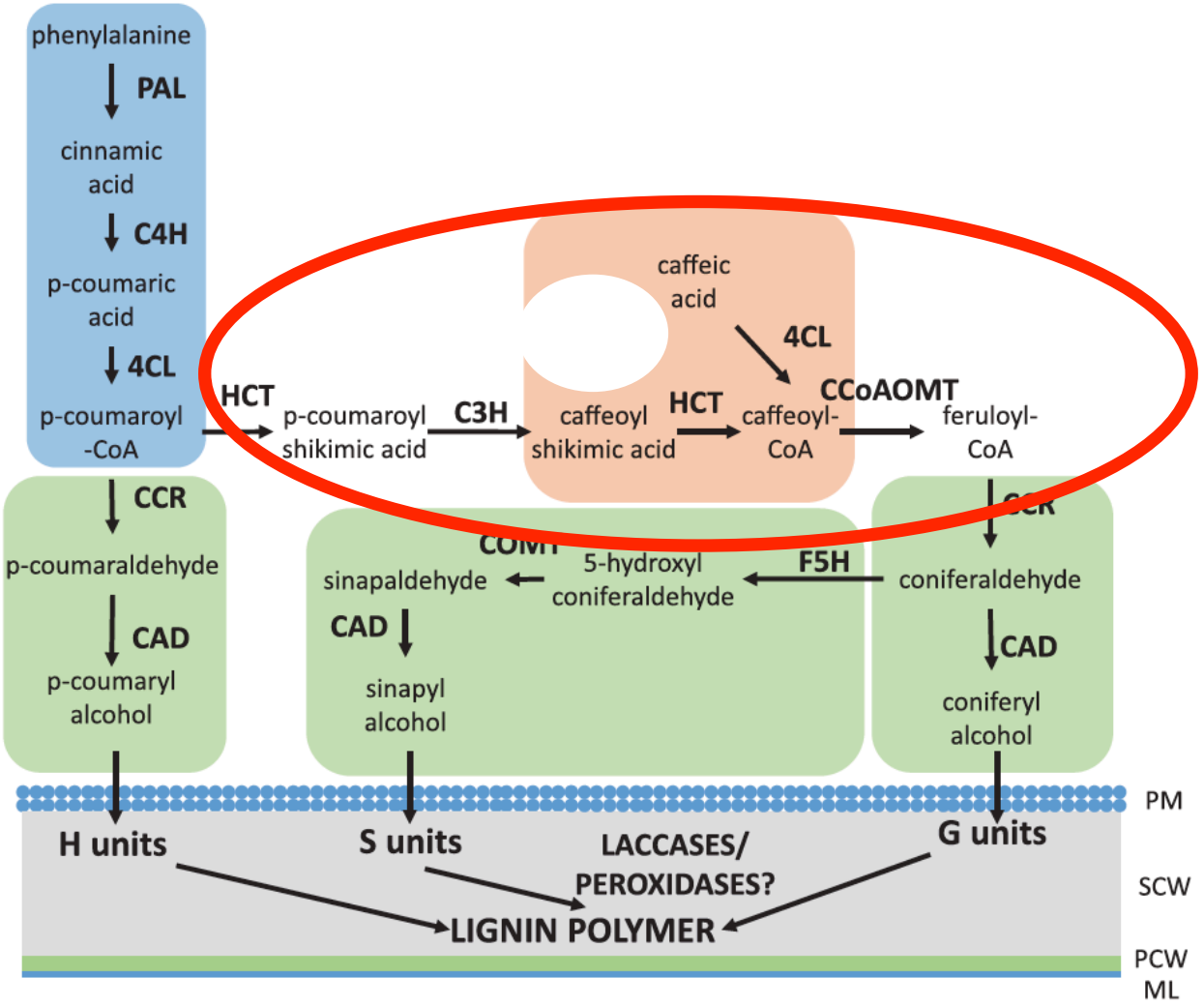
- Lignin was reduced by 36% in the *cse-2* mutant compared to wild-type (C)
- Proportion of H units increased over 30-fold in *cse-2* (D)



Vanholme et al., Science, 2013

CSE is active in the pathway after the branch leading to H units but before the pathways for G and S units diverge

At this part of the pathway, HCT, C3H, and CCoAOMT are also active and plants with reduced HCT and C3H activity also have lignin enriched in H units



Vanholme et al., Science, 2013

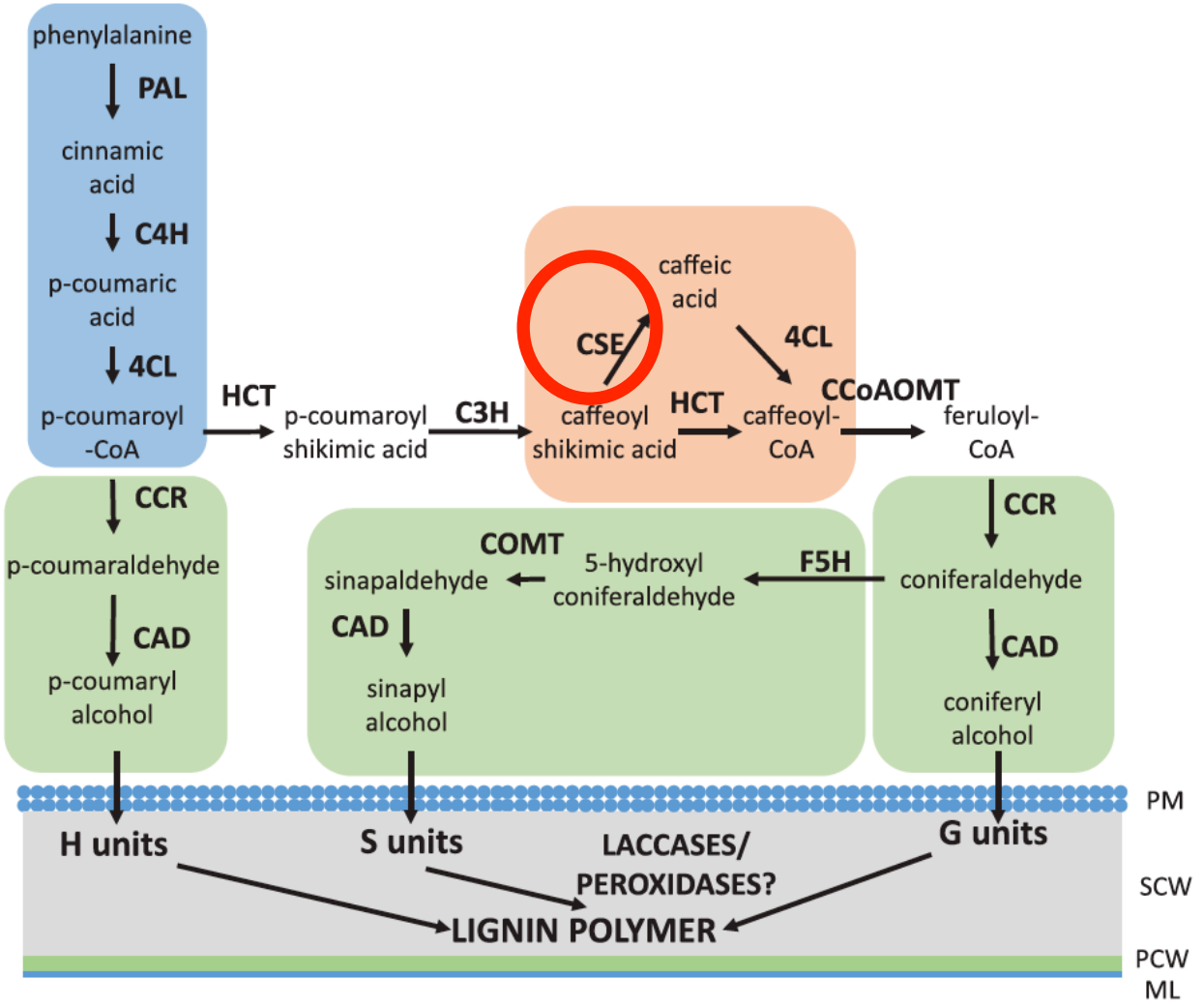
Phenolics analysis of the CSE mutants

- 27 compounds with increased abundance in the *cse* mutants, of which caffeoyl shikimate was most abundant
- Some of the compounds that accumulate in the *cse* mutants relative to the wild type might be substrates for CSE
- To test this possibility, CSE was purified; three compounds decreased in abundance upon treatment with CSE; **Caffeoyl shikimate** by was almost completely hydrolyzed CSE into **caffeic acid**
- Structure analysis of CSE revealed that **caffeoyl shikimate** could fit into the active site

Vanholme et al., Science, 2013

This study revised lignin biosynthetic pathway model

- Lignin pathway models before this study indicate that **caffeoyl shikimate** is converted to **caffeoyl-CoA**
- The authors tested whether this reaction could also be catalyzed by CSE with **caffeoyl shikimate** and **CoA** as substrates, only caffeate, but not caffeoyl-CoA, was produced



Vanholme et al., Science, 2013

Insertional mutagen

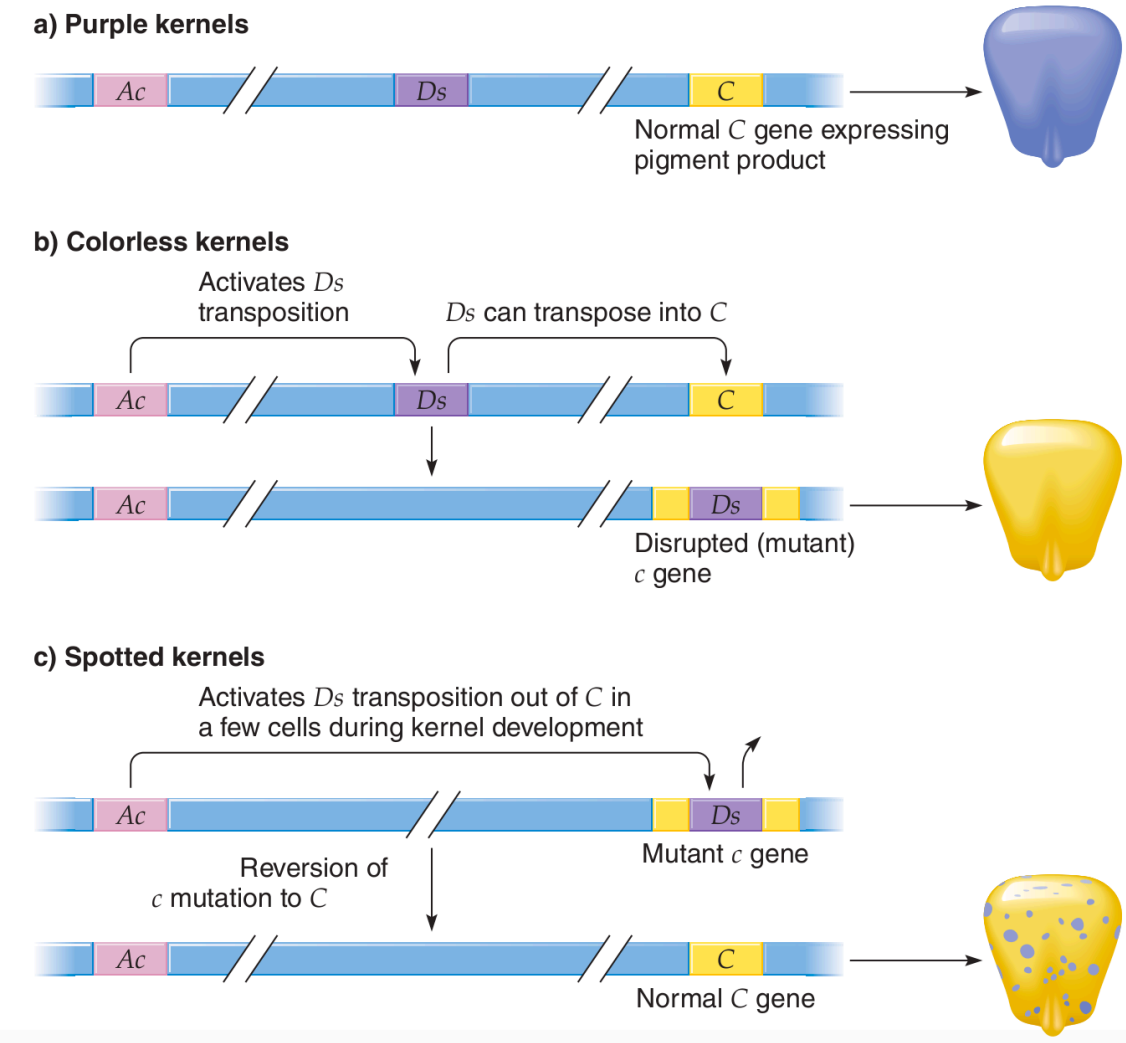
- **Insertional mutagen** induces mutations of DNA via incorporation of additional bases.
- Insertional mutations can mediated by bacteria, transposon, or virus.

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

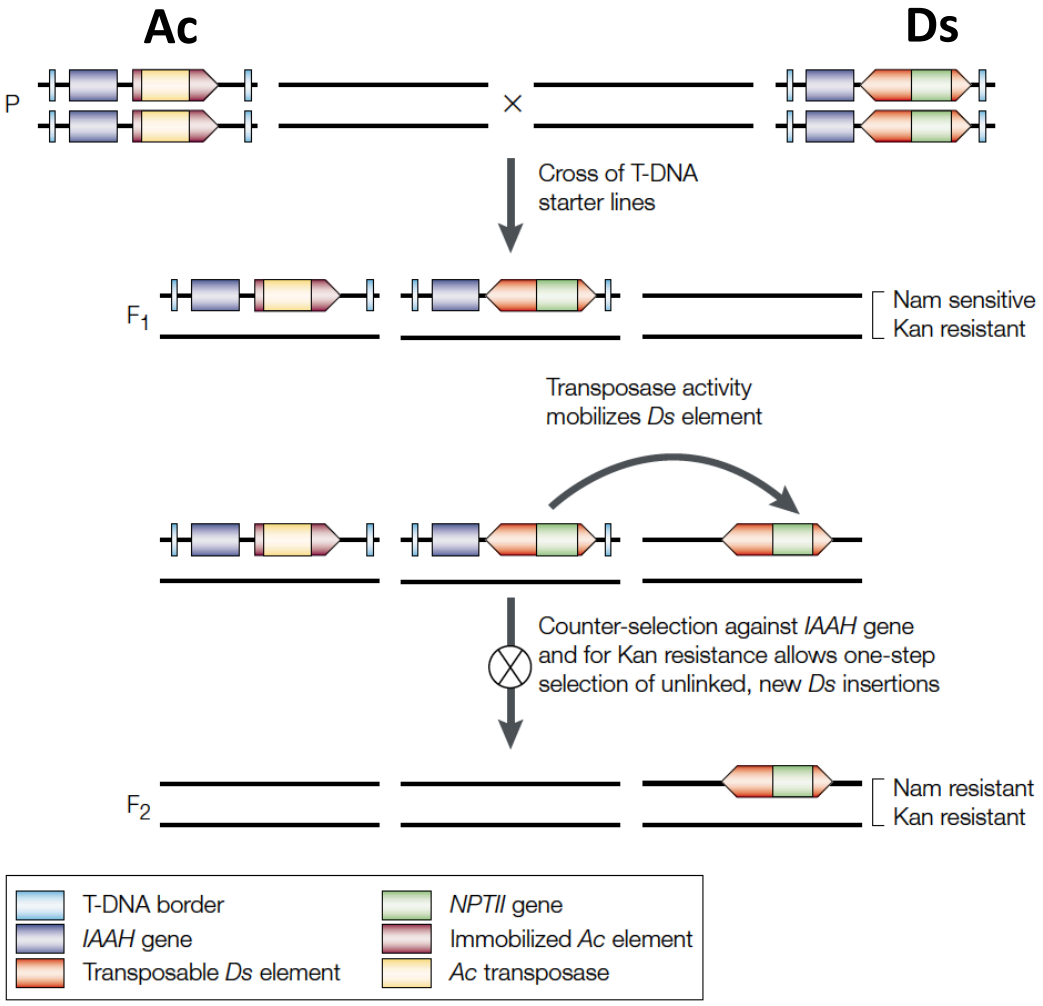


Silencing of transposons

- 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-polllinate 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

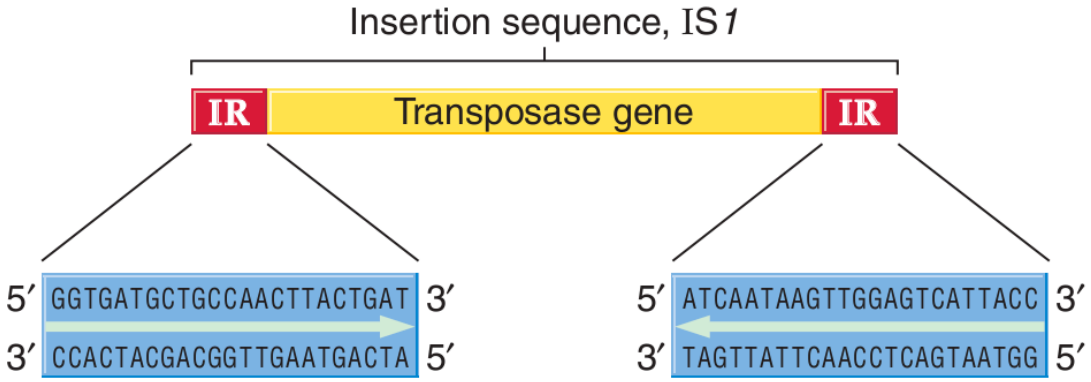
Genotypes of two loci Ac and Ds in F_2 population

- 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/-**	**-/- -/-**

Example 2. A mini-transposon for insertional mutagenesis in the bacteria *S. pneumoniae*

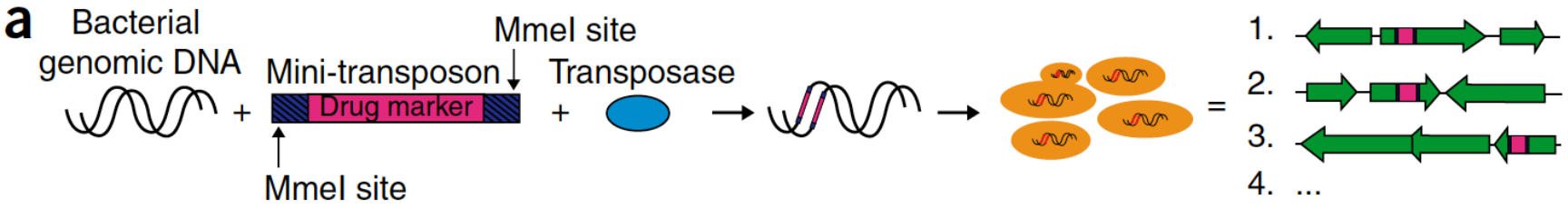
- The mini-transposon has ends with terminal inverted repeats (IRs) of 9 to 41 bp



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

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

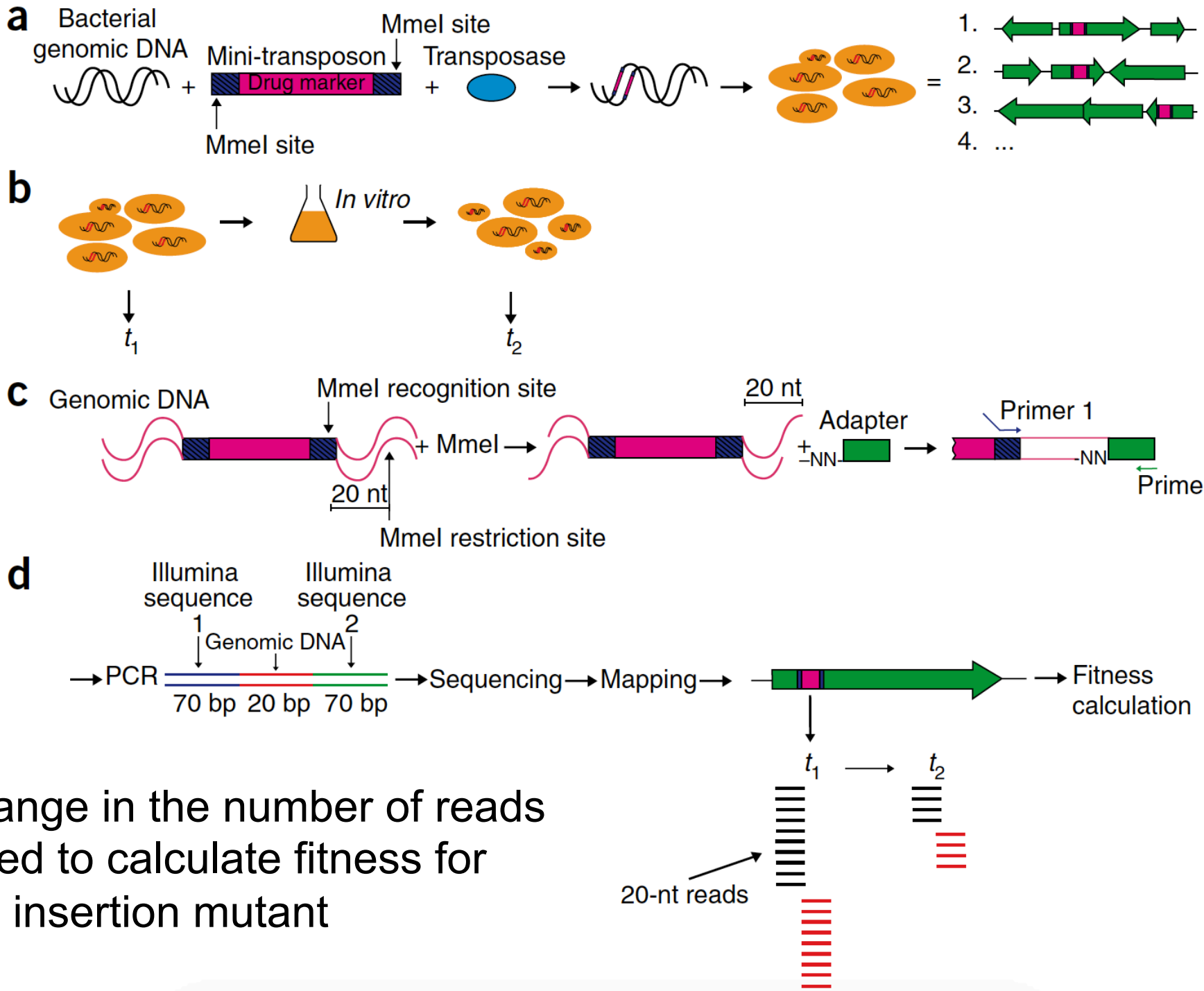
- A gene disruption library is constructed by first transposing the mini-transposon into bacterial genomic DNA **in vitro** and then transforming a bacterial population with the transposed DNA
- The transpose inserts randomly in the genome, requiring only a TA dinucleotide at the insertion site



van Opijnen et al., Nature methods 2009

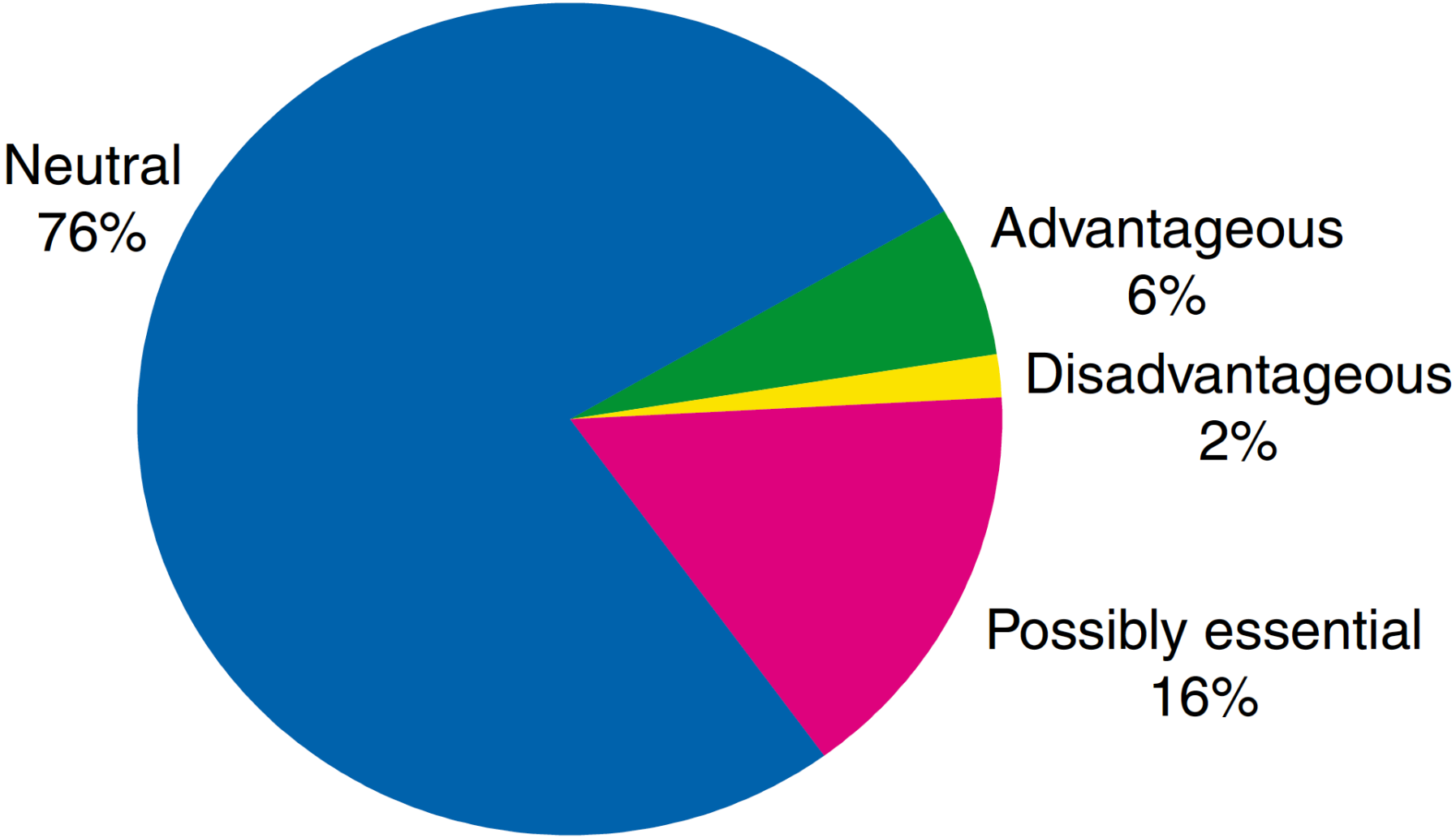
Using the mutants to identify genes affecting fitness of the bacteria *S. pneumoniae*

- *S. pneumoniae* is a bacteria causing pneumococcal disease in human
- **Fitness** in microbiology is the ability of microbes to thrive in a competitive environment. It is often determined by comparing the growth rate of a mutant strain in a given environment with that of its non-mutant isogenic relative (wild type)



A change in the number of reads is used to calculate fitness for each insertion mutant

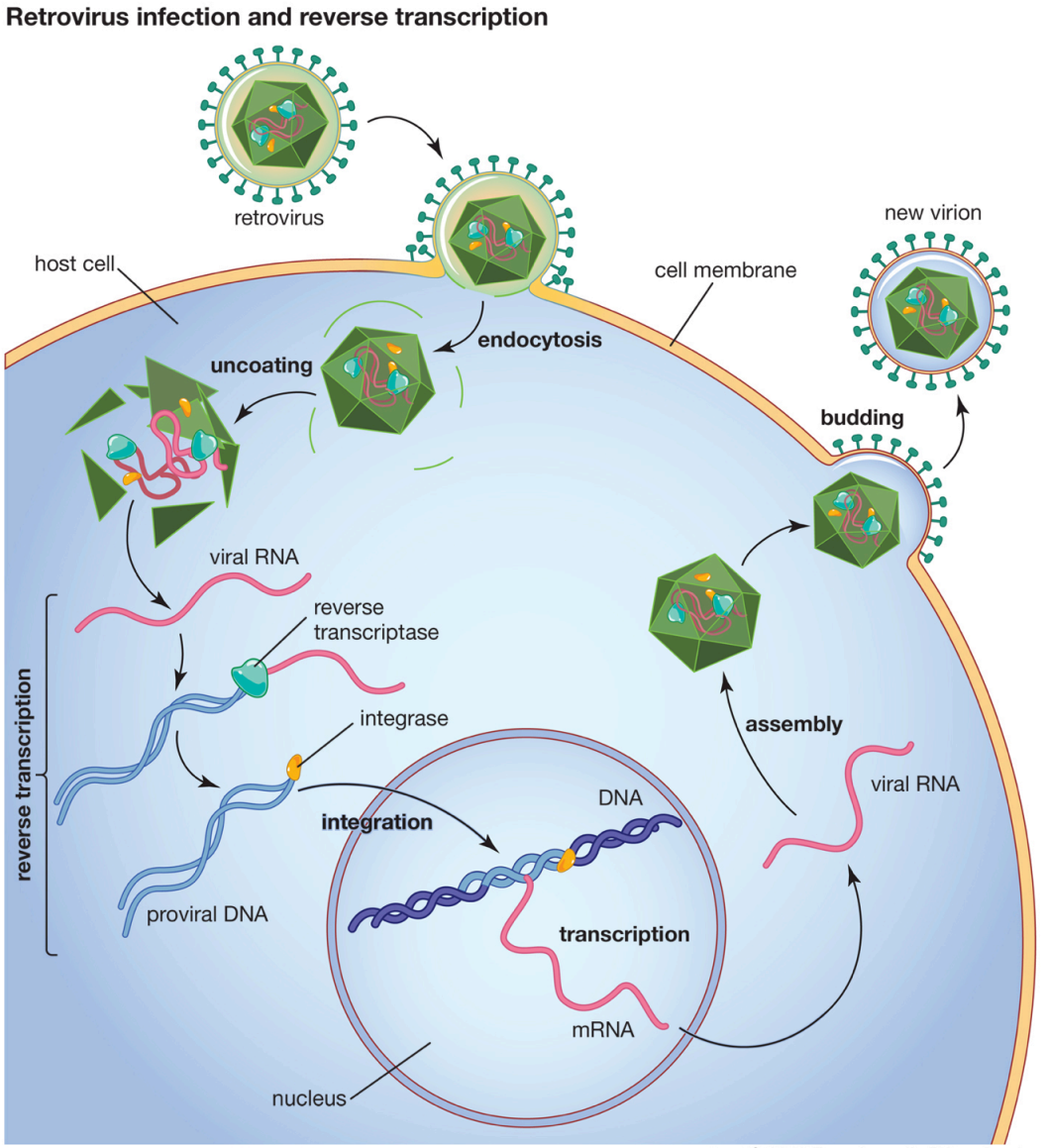
Categorization of every gene's fitness in the *S. pneumoniae* genome



Retrovirus

- Retrovirus: a type of virus that inserts a copy of its RNA into the DNA of a host cell that it invades
- HIV is a type of retrovirus and cause AIDs
  - https://www.biointeractive.org/classroom-resources/hiv-life-cycle

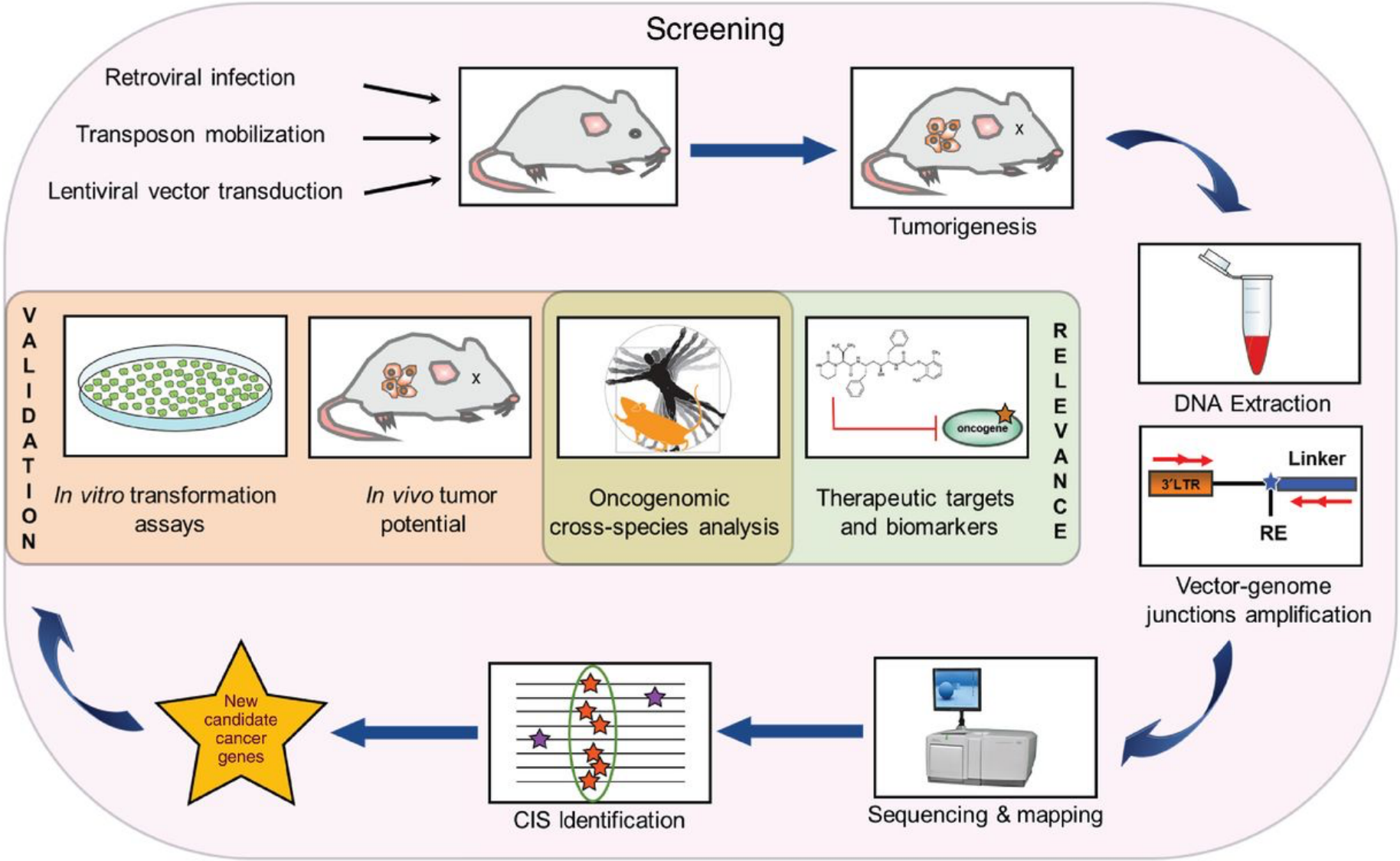
Retrovirus inserts its RNA into the DNA of a host that it invaded



Recombinant retrovirus system

- **Lentiviruses** are a subtype of retrovirus.
  lentiviruses are capable of infecting non-dividing and actively dividing cell types, whereas
  standard retroviruses can only infect mitotically active cell types
- Recombinant retrovirus system as tool for creating insertional mutations for gene function studies, gene delivery, or gene therapy
  - https://www.youtube.com/watch?v=kJSsZMdA8Sk

Cancer genes discovery by retroviral DNA integration



Forward genetics

- **Forward genetics** is a genetics approach for identifying gene(s) associated with for a specific trait. This was initially done by using naturally occurring mutations or inducing mutants with radiation, chemicals, or insertional mutagenesis. From individual(s) with abnormal phenotype for a trait to locate the causal gene(s) and then to characterize its function.

Three types of mutagens induce mutations randomly distributed on genome

- Mutagen types
  - Insertional mutagen
  - **Physical mutagen**
  - Chemical mutagen

✓ How to use mutagens to create mutations?

✓ Characters of induced mutations

✓ How to identify mutation site in a mutant?

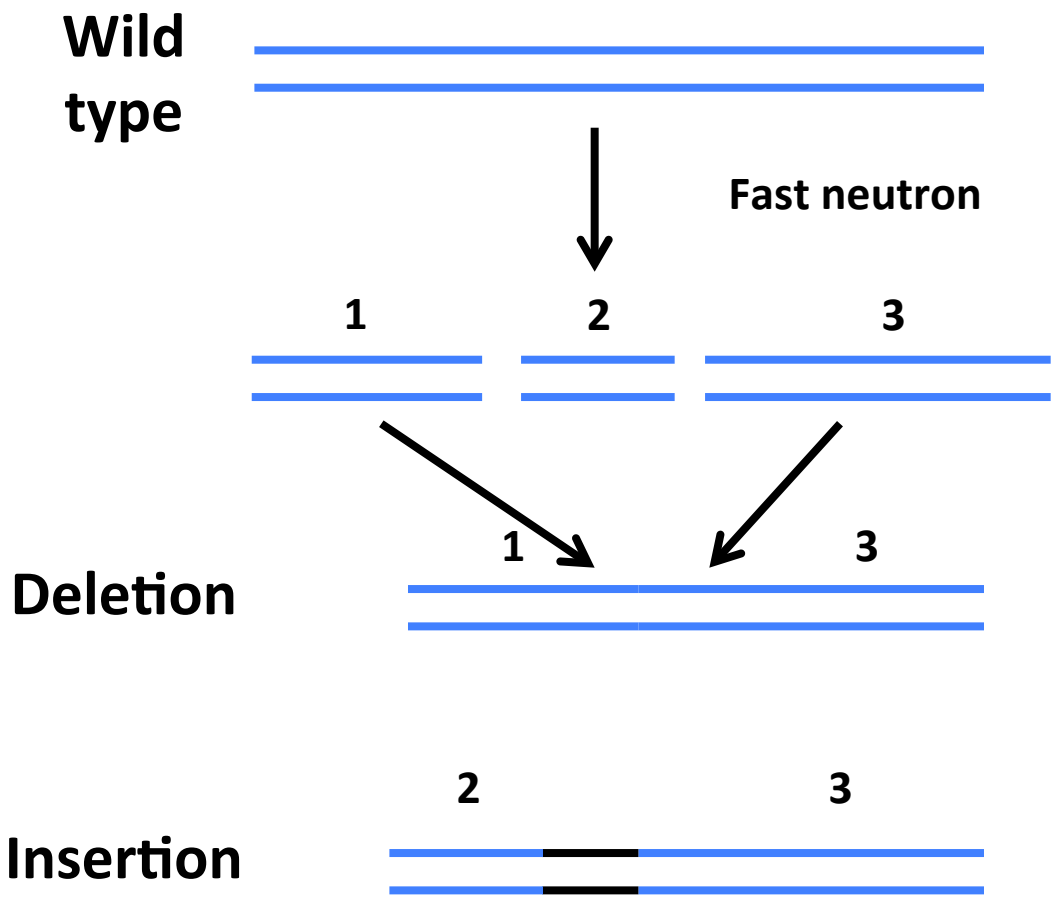
✓ How to use the induced mutations to study gene function

Physical mutagenesis

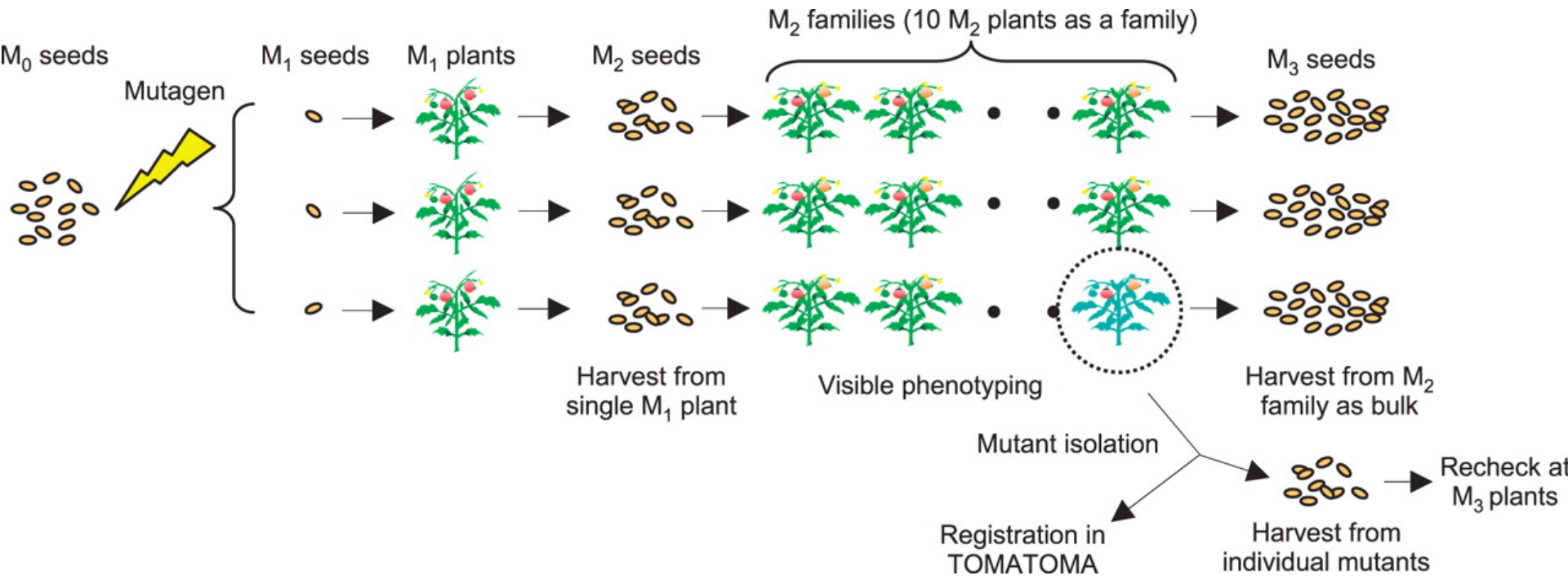
- Physical mutagens cause DNA breakage or other damages inducing mutations and work for all species.
- Physical mutagens
  - Ionizing radiation, such as Fast Neutron
  - Non-ionizing radiation, such as UV

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



Flow chart for constructing mutagenized population using physical mutagen



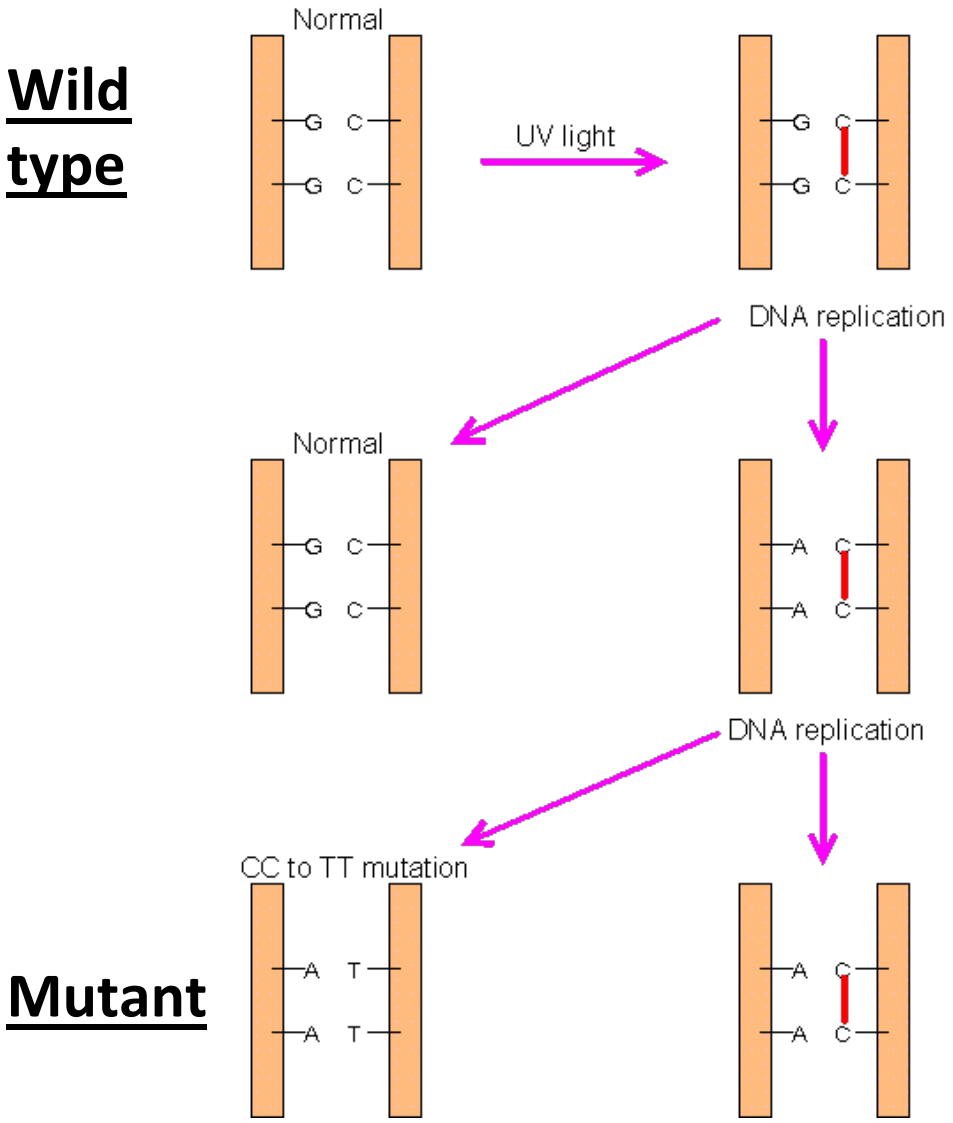
Sequencing 1504 rice mutants facilitate functional genomic studies

- Sequenced a fast-neutron-induced mutant population of 1504 lines in the model rice cultivar Kitaake
- Identified 91,513 mutations an average of 61 mutations per line
  - 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

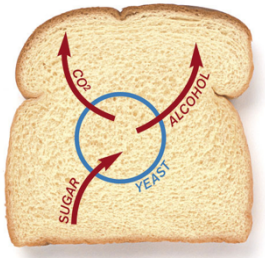
UV, non-ionizing radiation, induces pyrimidine dimers mutation

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



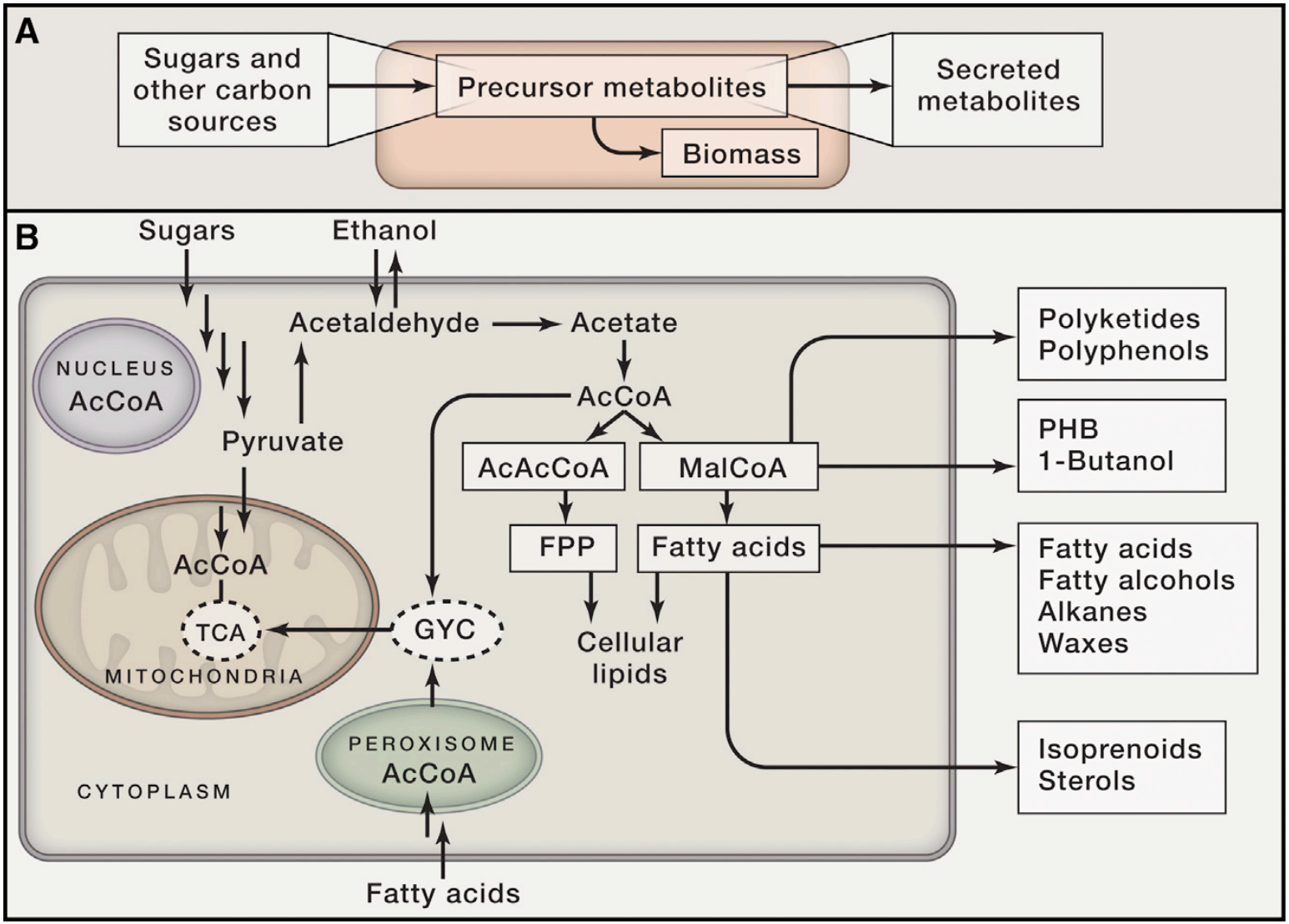
Yeast is a cell factory for production

- Food and beverages
  - Fermentation, converts sugar to carbon dioxide (CO_2) and alcohol
  - Beer, bread, yogurt, etc.
- Pharmaceuticals
  - Antibiotics, hormones, and anti-cancer drugs
  - 300 biopharmaceuticals have sales over $100 billion
- Fuel and chemicals
  - Bioethanol, citric acid, etc.
  - Production is $3000 billion in industry



Jens Nielsen 2014; Nielsen and Keasling 2016

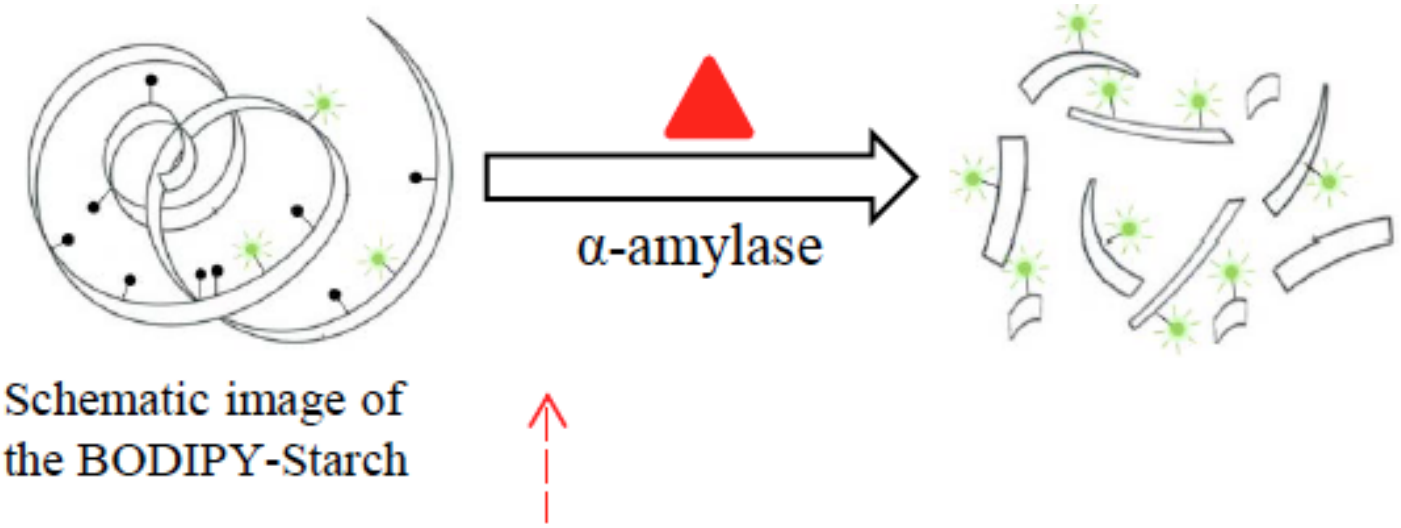
Yeast as cell factories



Nielsen and Keasling 2016

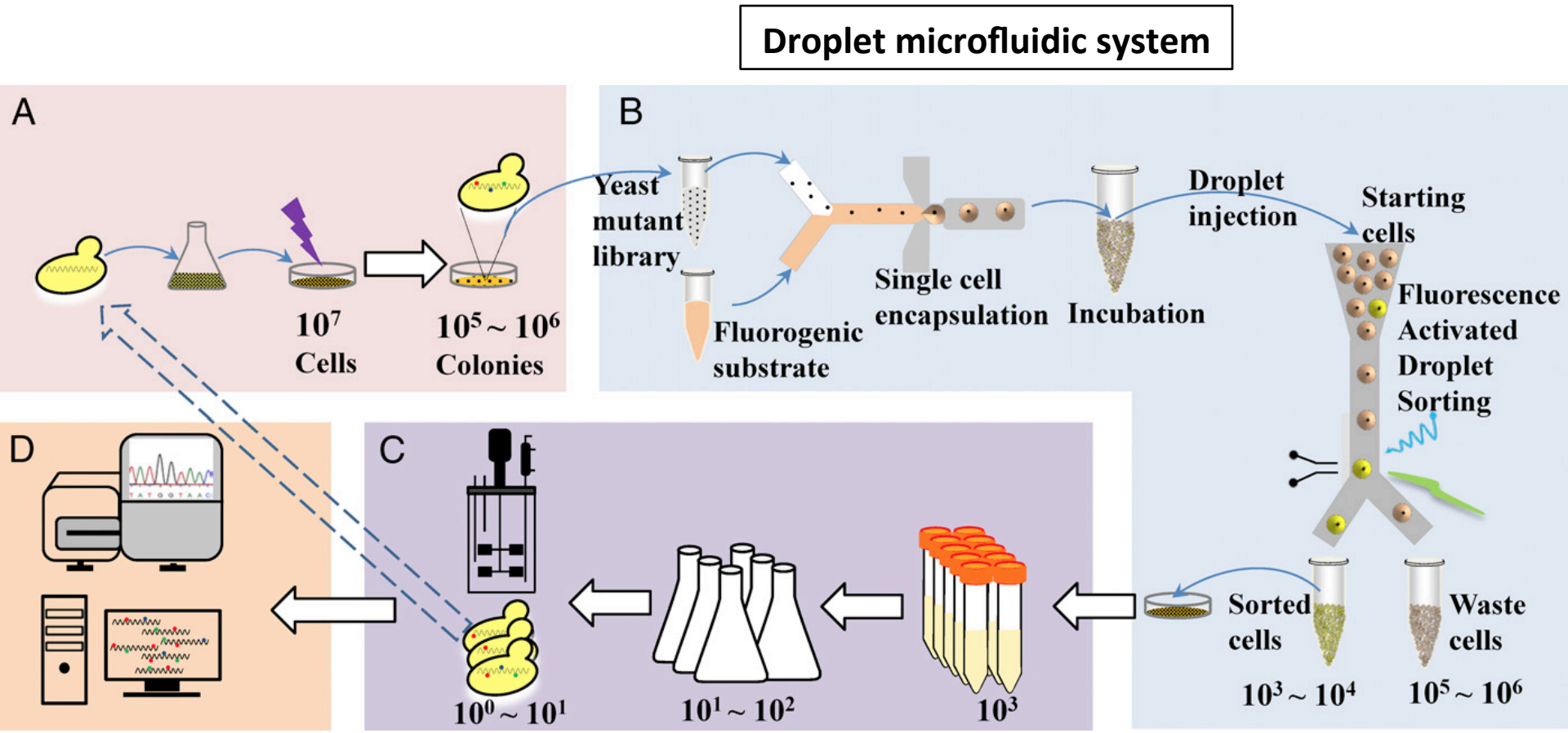
Example: UV induced mutations in Yeast to increase α-amylase production

- α-amylase catalyze a reaction that converts starch to simple sugars



Huang et al., 2015

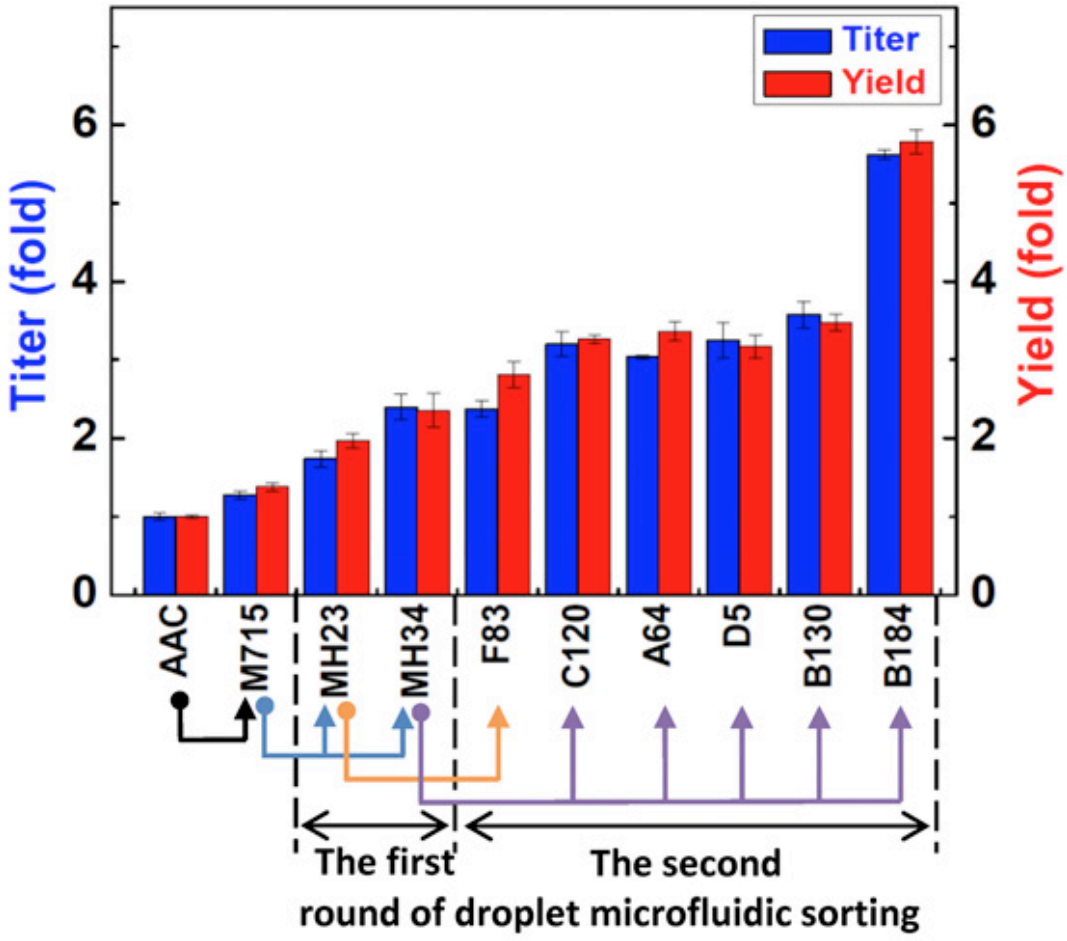
Example: UV induced mutations in Yeast to increase α-amylase production



Huang et al., 2015

Example: UV induced mutations in Yeast to increase α-amylase production

- Two cycles of selection of UV induced mutants leads to some yeast strains with improved α-amylase production



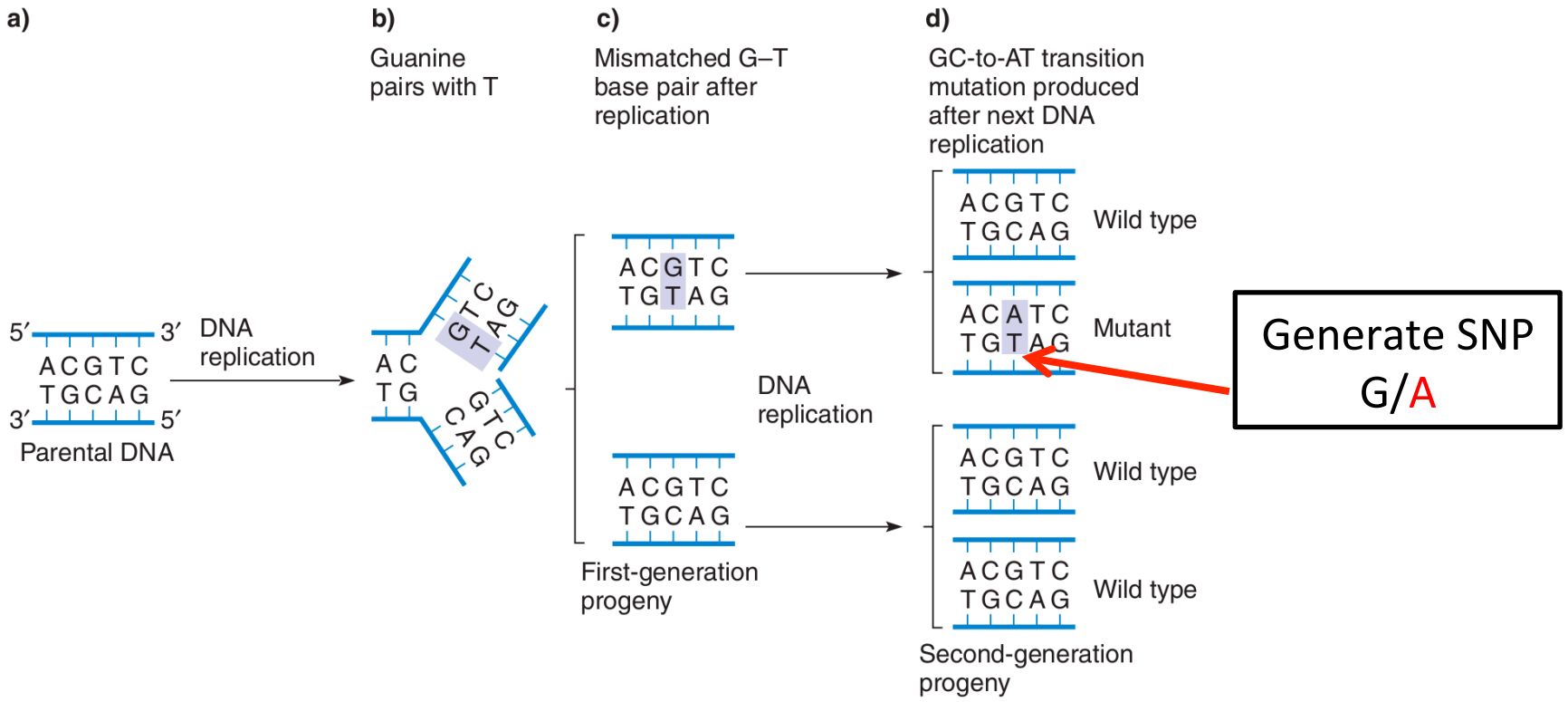
Huang et al., 2015

Point mutation

- Point mutation is a single nucleotide change of DNA
- It can be occurred naturally during DNA replication or induced by mutagens

Spontaneous point mutation --DNA replication error

- Spontaneous mutation rate is very low, 10^{-7}-10^{-11} per gene per generation
  - Natural DNA replication error is very low
  - Mismatch repair system can repair the mismatch

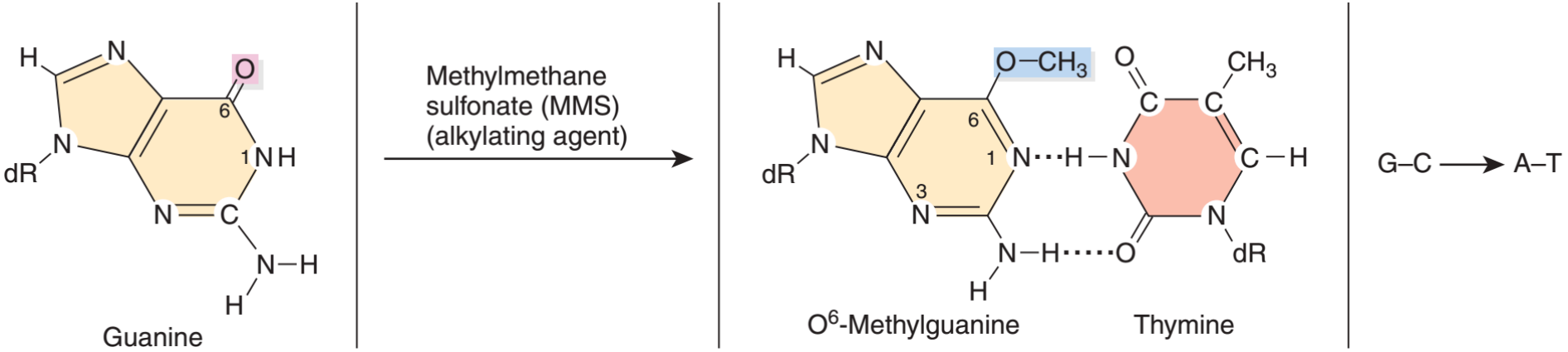


Chemical mutagens induce mutations

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

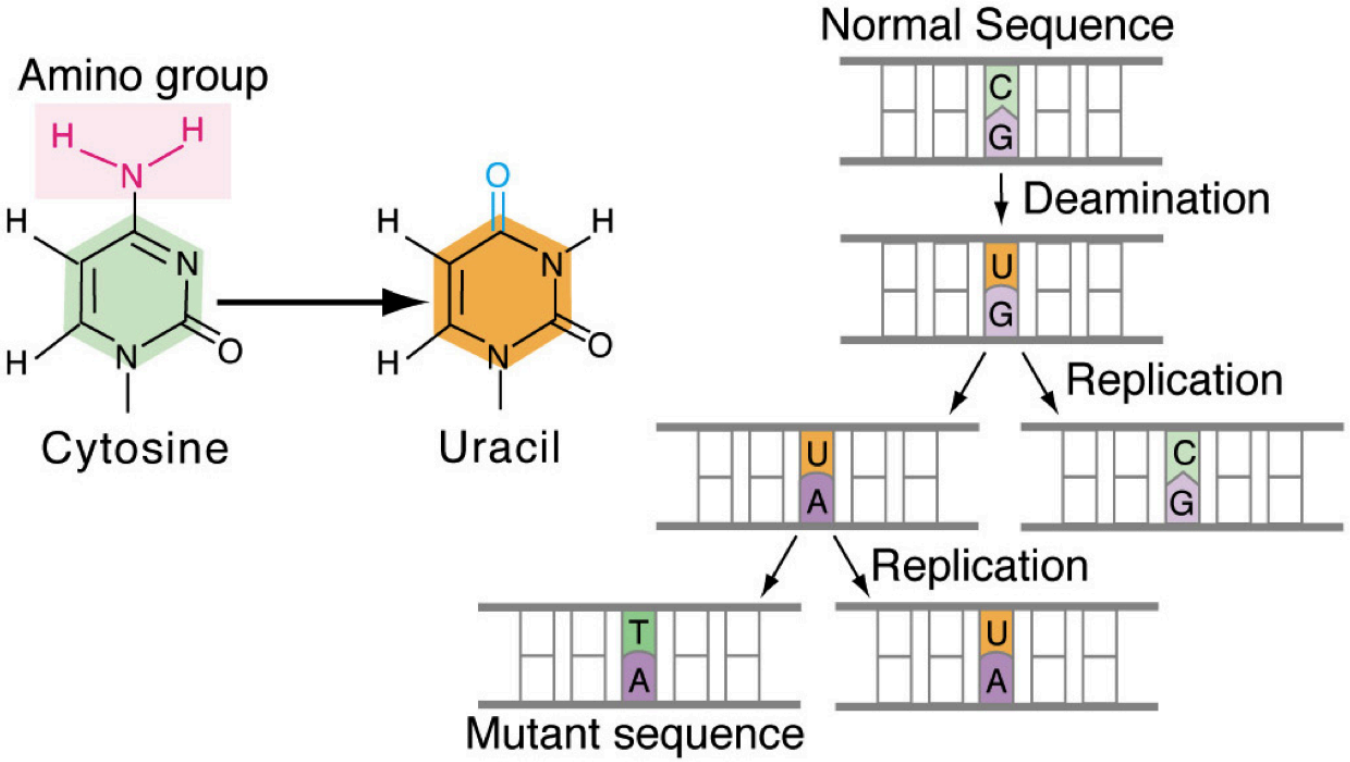
Chemical mutagen - Alkylation

- Ethyl-methanesulfonate (EMS) and Methylmethane sulfonate (MMS) add alkyl group (e.g., -CH_3) onto a base of guanine (G), producing O^6–alkylguanine
- The methylated guanine pairs with thymine (T) rather than cytosine (C), giving GC-to-AT transitions



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 during replication, resulting in a CG-to-TA transition mutation

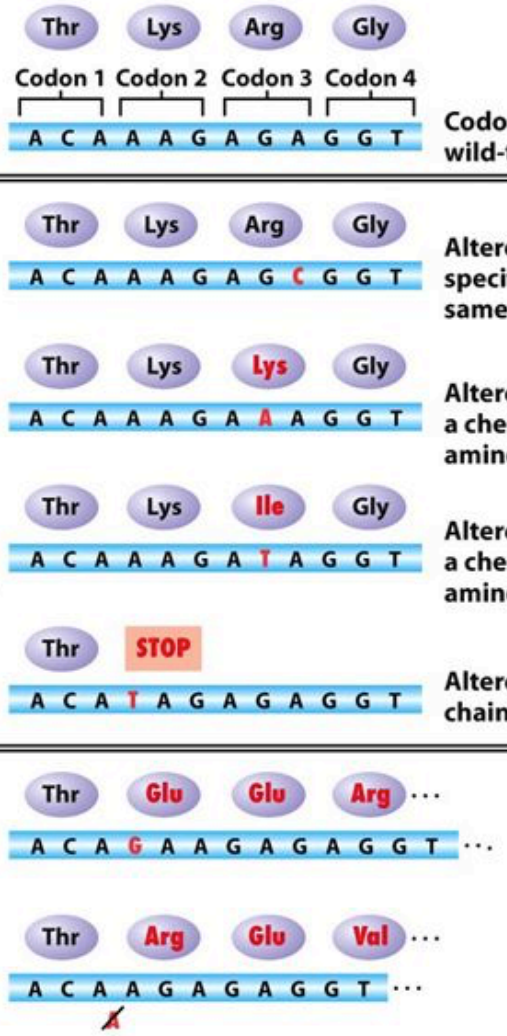


Types of point mutations

- **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.
  - Nonconservative 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

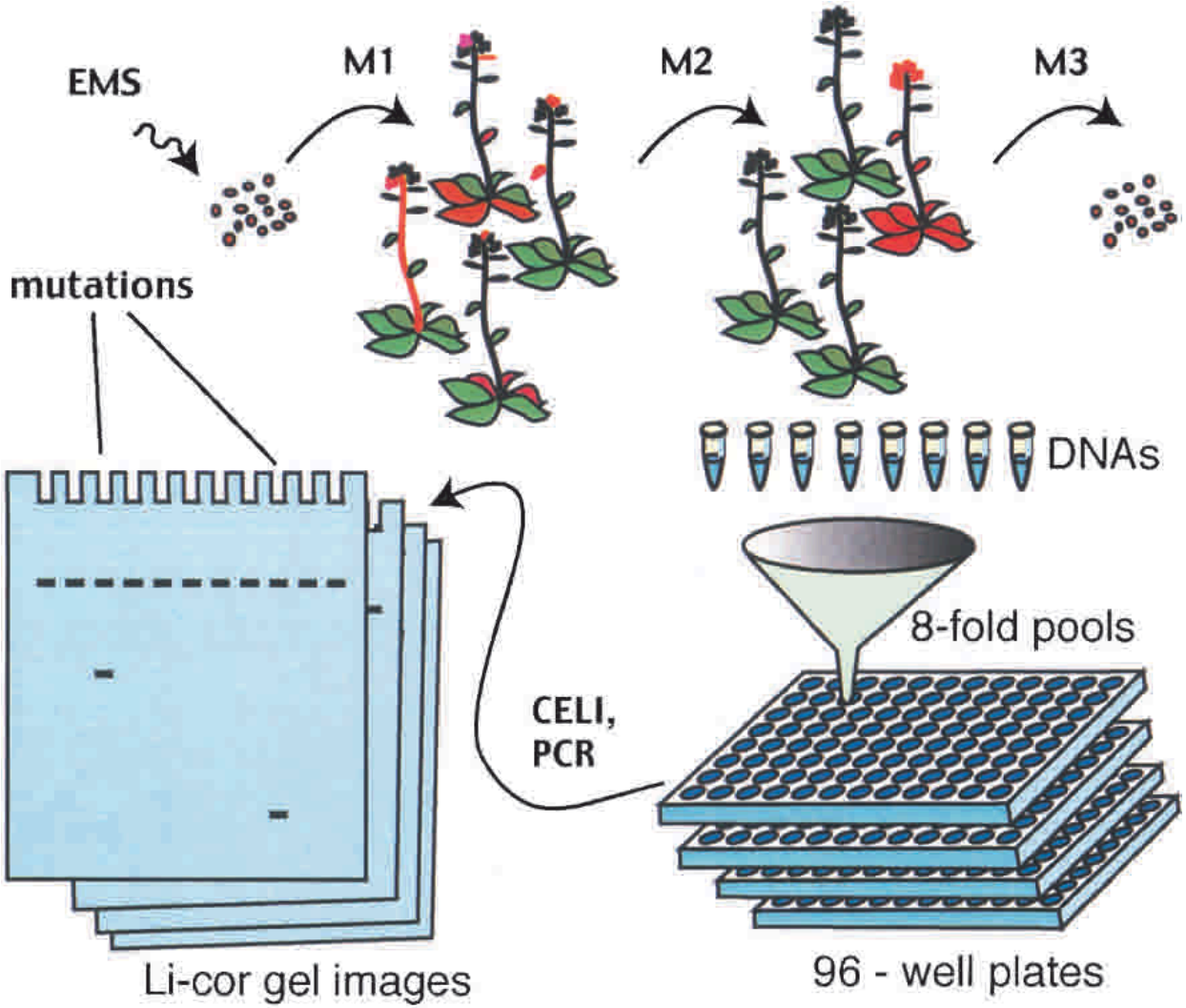
Amino acids biochemical properties | nonpolar | polar | basic | acidic

Termination: stop codon

Standard genetic code

1st base	2nd base								3rd base
	T		C		A		G		
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

Flow chart for constructing mutagenized populations using EMS



Characters of chemical mutagen induced mutations

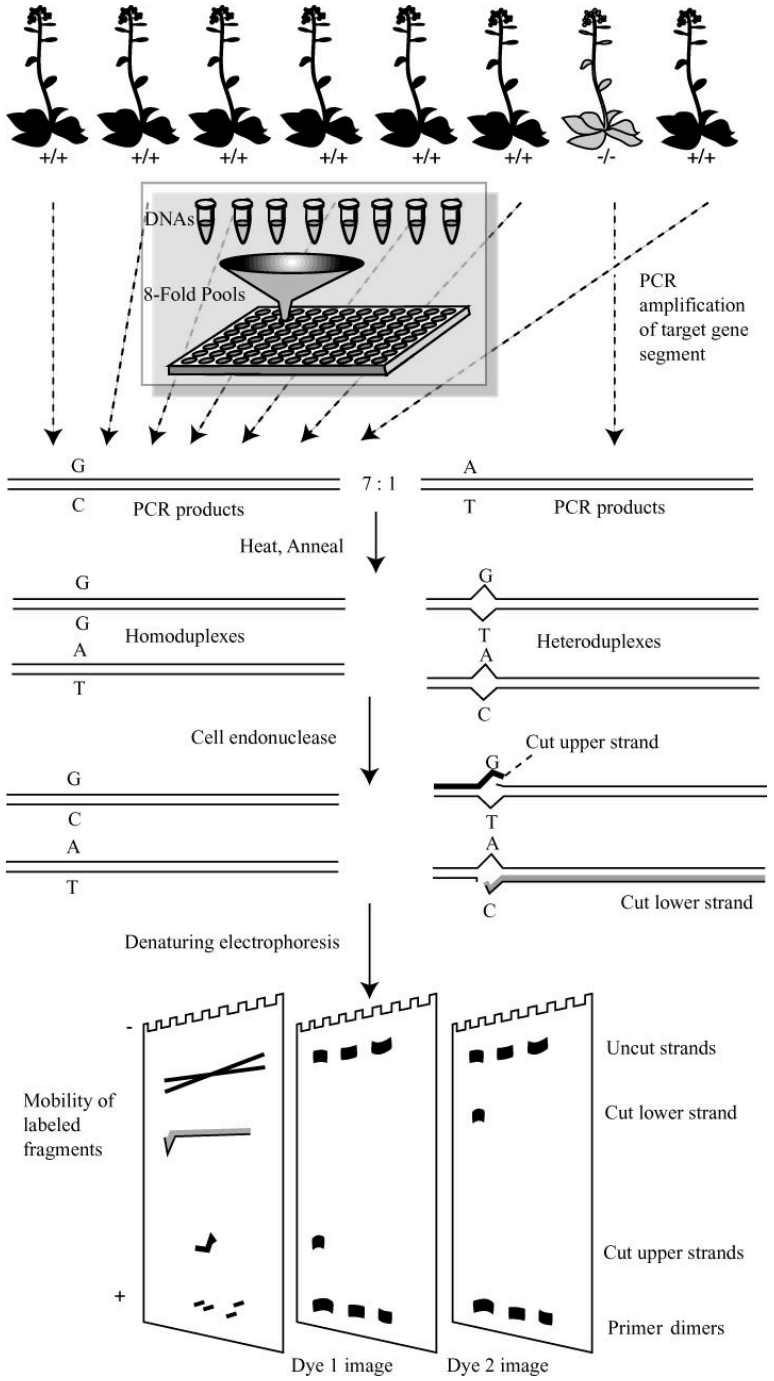
- Higher mutation rate allows for saturation to be achieved using relatively fewer individuals
  - Arabidopsis (120 Mbp): 240-1200 mutations/line
- Mutation rate (number of mutations per Mbp) similar among species, beneficial to large genomes
  - Wheat (17,000 Mbp): 34,000-170,000 mutations/line
- Provides allelic series, and not just knockouts, which can yield refined insights into gene function

Reverse genetics using induced point mutations

Identify an induced point mutation for a candidate gene using gene-specific PCR followed by

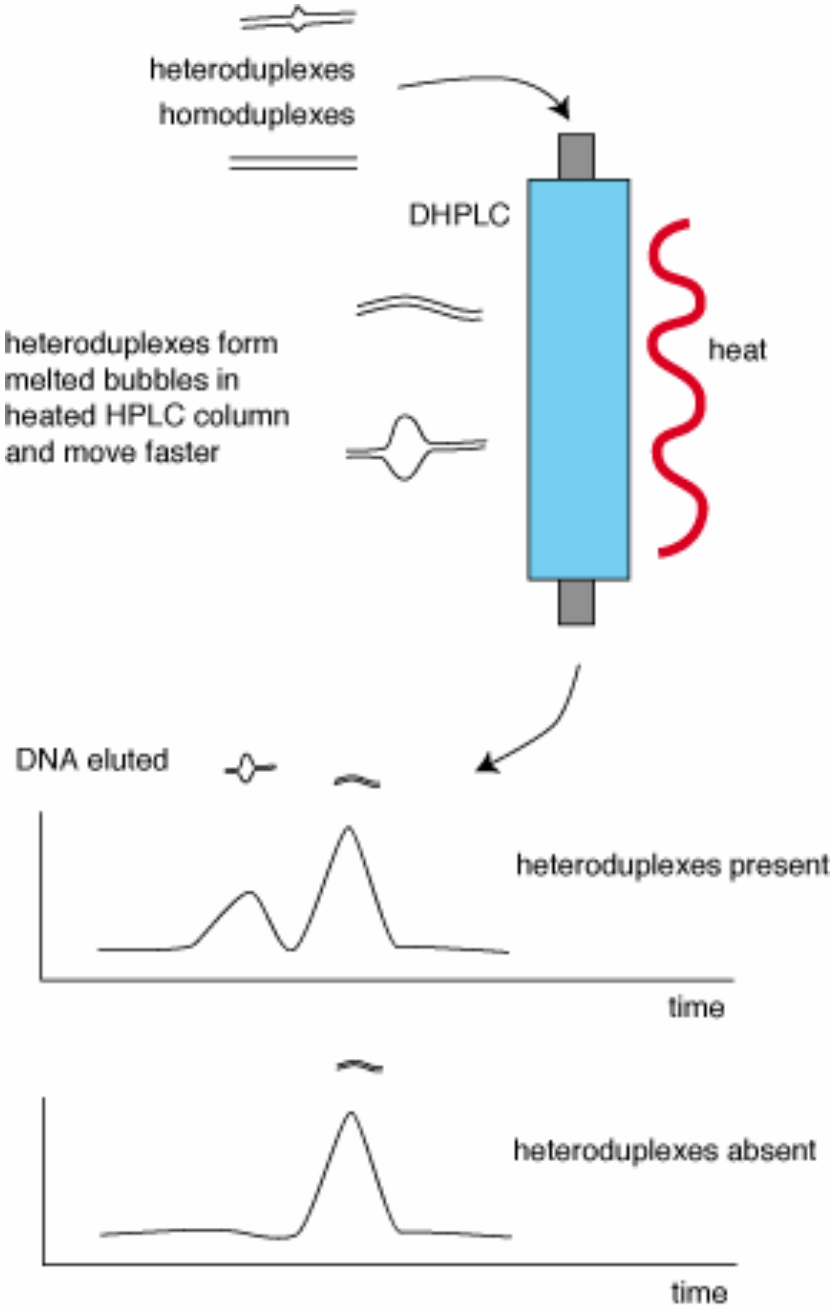
1. Heteroduplex analysis
2. Denaturing high performance liquid chromatograph
3. Sequencing

Identify an induced point mutation for a candidate gene using gene-specific PCR followed by heteroduplex analysis



Henikoff and Comai, 2003 ***Annual Review of Plant Biology***

Identify an induced point mutation for a candidate gene using gene-specific PCR followed by denaturing high performance liquid chromatography



McCallum et al. 2000 *Plant Physiol.*

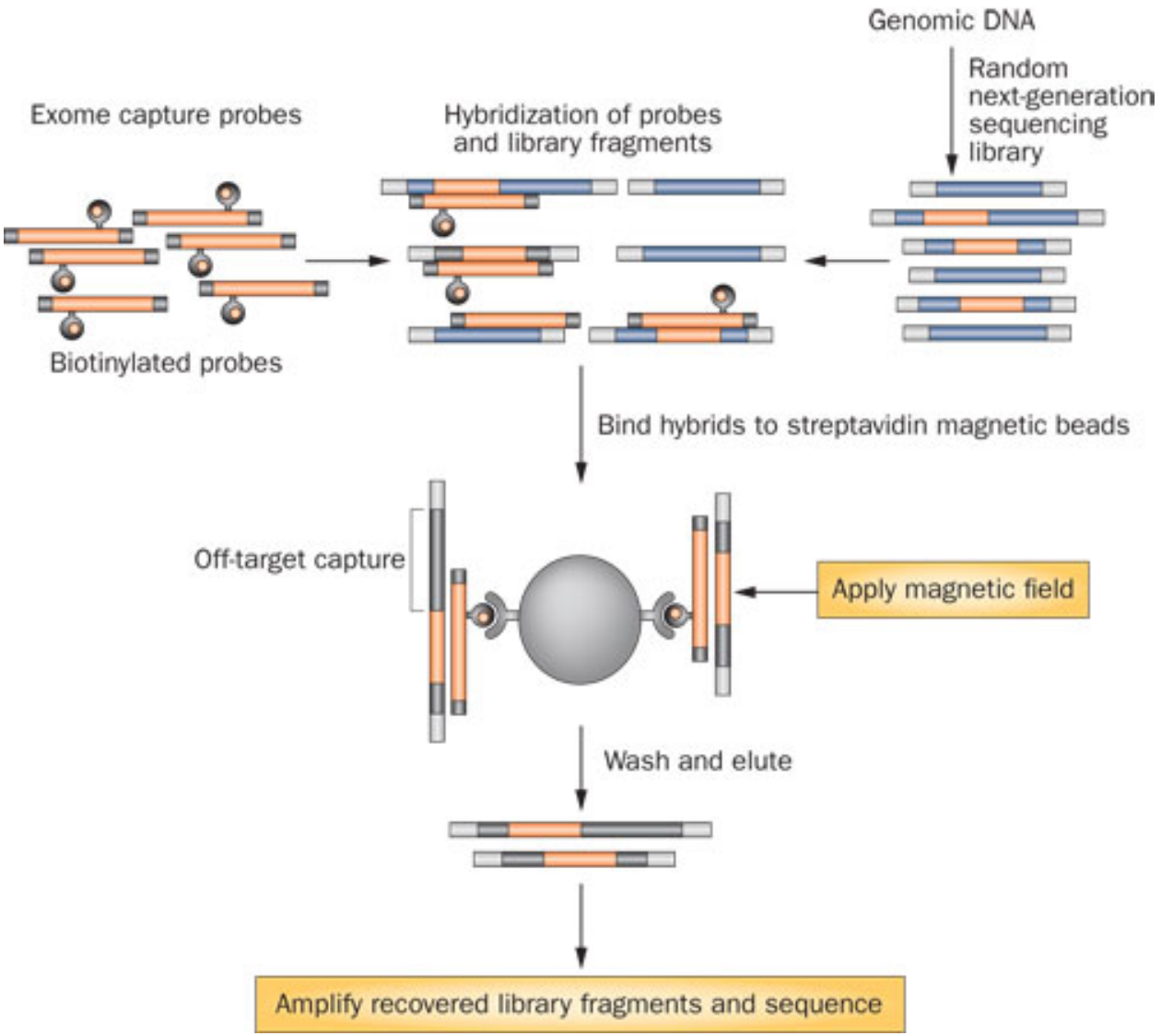
Identify an induced point mutation for a candidate gene using gene-specific PCR followed by next-generation sequencing

- Sequencing cost decreased that makes directly sequencing PCR product of a target gene possible to locate position of mutation and cost effectively
- How many mutants are included in one pool?
- ~10 million reads per run on MiSeq, cost about $1,000?
- Large pool allows screening of more individuals per sequencing run, but pool size is limited by the ability to detect a real single mutation from sequence error (sequence error rate is about 0.3-0.5%)
- A case study of effect of pooling size on point mutation detection: Helen Tsai et al. Plant Physiol. 2011;156:1257-1268

Gene-indexed catalogue chemical mutagens induced mutants

- Whole genome sequencing
- Exome capture sequencing
  - Exome is all exons of protein coding genes in a genome
  - Exome capture sequencing is a technique for sequencing all exons of the protein coding genes in a genome

Exome capturing



Whole-genome sequencing (WGS) or exome capture sequencing (ECS)

$2000/lane of Illumina HiSeq2000, 200 million reads of SE-100

Species	Analysis	Size of Target (Mb)	Reads for 20× Coverage (Million)	Capture Cost ($)	Total Cost ($)
Most plant species	ECS	40	20	40	307
Arabidopsis	WGS	120	12	0	180
Rice	WGS	380	38	0	507
Tomato	WGS	900	90	0	1,200
Maize	WGS	~2,300	230	0	3,067
wheat	WGS	~15,960	1,596	0	21,280

Henry et al., The Plant Cell, 2014

Whole genome sequencing of sorghum EMS-induced mutants

- Sorghum (~730 Mb) is a plant model species in family *Poaceae*
- 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
- **The vast majority (97.5%) of the induced mutations were distinct from natural variations**

Jiao et al., The Plant Cell, 2016

Exome capture sequencing of wheat EMS-induced mutants

- A wheat exome capture platform was developed and used to sequence the coding regions of 2,735 wheat mutants
- The targeted exon space in wheat was selected
  - An 84-Mb exome capture assay including overlapping probes covering 82,511 transcripts
- Over 10 million mutations were identified from the 2,735 mutants
- On average, 2,705-5,351 mutations per mutant, 35–40 mutations per kb

Krasileva et al., PNAS, 2017

What need to know for final exam

- Three types of mutagens used for inducing mutations
- Can briefly describe how to create insertional mutants using T-DNA/ agrobacterium
- T-DNA/agrobacterium induces mutants with large fragment insertion, which causes loss of the gene function
- T-DNA insertional sites are randomly distributed on genome
- 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.

What need to know for final exam

- What is reverse genetics?
- Can briefly describe how to use gene specific PCR followed by Gel analysis to identify a T-DNA insertional mutant for a candidate gene
- Can briefly descirbe how to use TAIL-PCR to obtain flanking sequences of inserts and locate T-DNA insertional sites for a large number of mutants
- What is transposon?
- What is retrovirus?
- What is forward genetics?
- What are synonymous mutation, missense mutation, and nonsense mutation?

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.