Precision Plant Breeding using Genome Editing Technologies

Alfred Huo

Mid-Florida Research and Education Center

UF|IFAS
UNIVERSITY of FLORIDA

MILESTONES IN PLANT BREEDING

CROP DOMESTICATION

Farmers select the best wild species to develop crops

10,000 BC

Domestication of wheat

FACTS

For **10,000** years, farmers and breeders have been developing and improving crops

For **150** years, plant scientists and breeders have improved plant breeding on a scientific basis

Today, farmers feed at least **10** times more people using the same amount of land as 100 years ago

By 2050, we will need **50%** more food to feed a population of 11 billion

PLANT BREEDING BASED ON CROSS BREEDING

Development of improved varieties by combining good characteristics from two parents

1865

Mendel's laws

Gregor Mendel describes the inheritance of traits from one generation to the next. His laws become the core of classical genetics

HYBRID BREEDING

Crossing two genetically different individuals to develop better performing hybrids

More vigorous hybrid corn

1926

MUTAGENESIS

Developing new genetic diversity by exposing crop plants to chemical agents or radiation

Blast-resistant rice

1940

1953

Understanding the structure of DNA

James Watson and Francis Crick identify the double helix of DNA

GMO

Introducing foreign genes into the DNA of a plant

Insect-resistant cotton

1994

MARKER-ASSISTED SELECTION

Locating desirable traits in a plant for efficient selection and breeding

Barley resistant to yellow dwarf virus

2000

TARGETED BREEDING

Using modern tools such as genome editing for more targeted breeding

Waxy corn

now

future

2016 CRISPR/WAX Corn

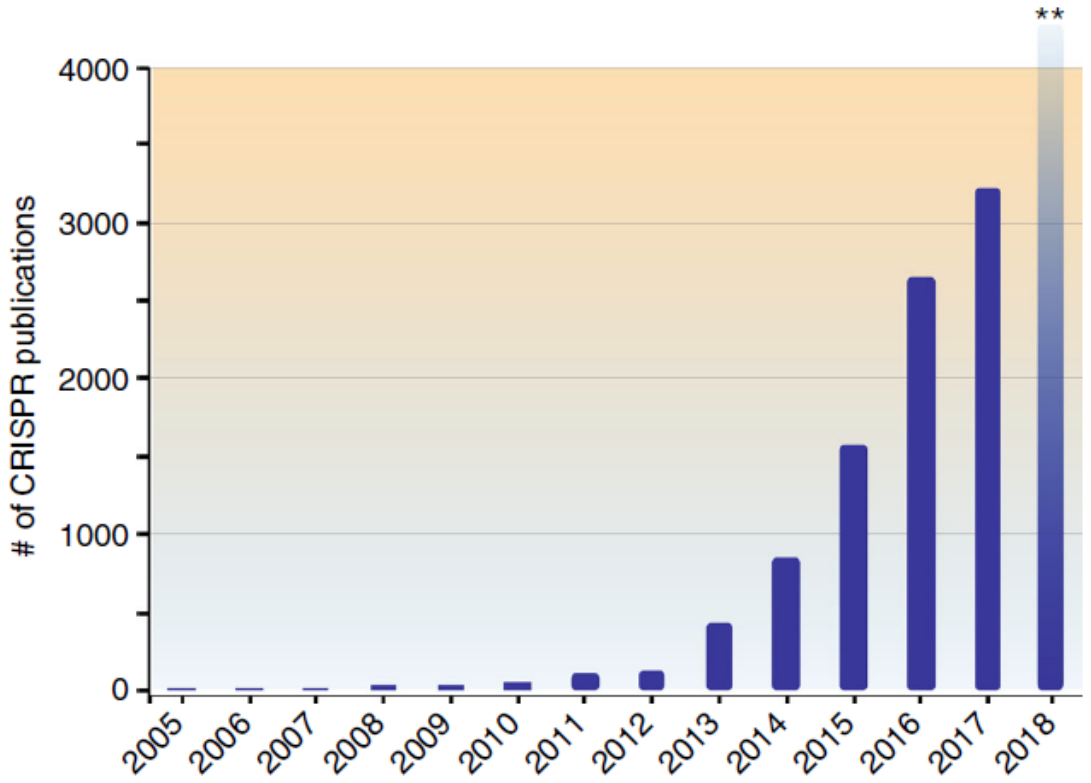
PLANT BREEDING BASED ON GENETIC INFORMATION

Development of improved varieties by working directly with the DNA

GENOME EDITING

A type of genetic engineering in which DNA is inserted, replaced, or removed from a genome using artificially engineered nucleases, or molecular scissors.

Genome editing was selected by Nature Methods as the 2011 Method of the Year



11053

CRISPR-based genome-targeting tools are widely used. Number of PubMed publications over the last 12 years that had the word "CRISPR" or "Cas9" in the abstract or title. **Number of publications in 2018 is projected to be more than 10,000

Nature Communication, (2018) 9:1911

Comparison between traditional and modern genome editing technologies

Mutagen	Chemical(e.g., EMS)	Physical (e.g., gamma, X-ray or fast neutron radiation)	Biological (ZFNs, TALENs or CRISPR/ Cas)	Biological- Transgenics (e.g., Agro or gene gun)
Characteristics of genetic variation	Substitution and Deletion	Deletion and chromosomal mutation	Substitution and Deletion and insertion	Insertions
	Loss of function	Loss of function	Loss of function and gain of function	Loss of function and gain of function
Advantages	Unnecessary of knowing gene function or sequences	Unnecessary of knowing gene function or sequences	Gene specific mutation	Insertion of genes of known functions into host plant genome
	Easy production of random mutation	Easy production of random mutation	Efficient production of desirable mutation	Efficient creation of plants with desirable traits

Comparison between traditional and modern genome editing technologies

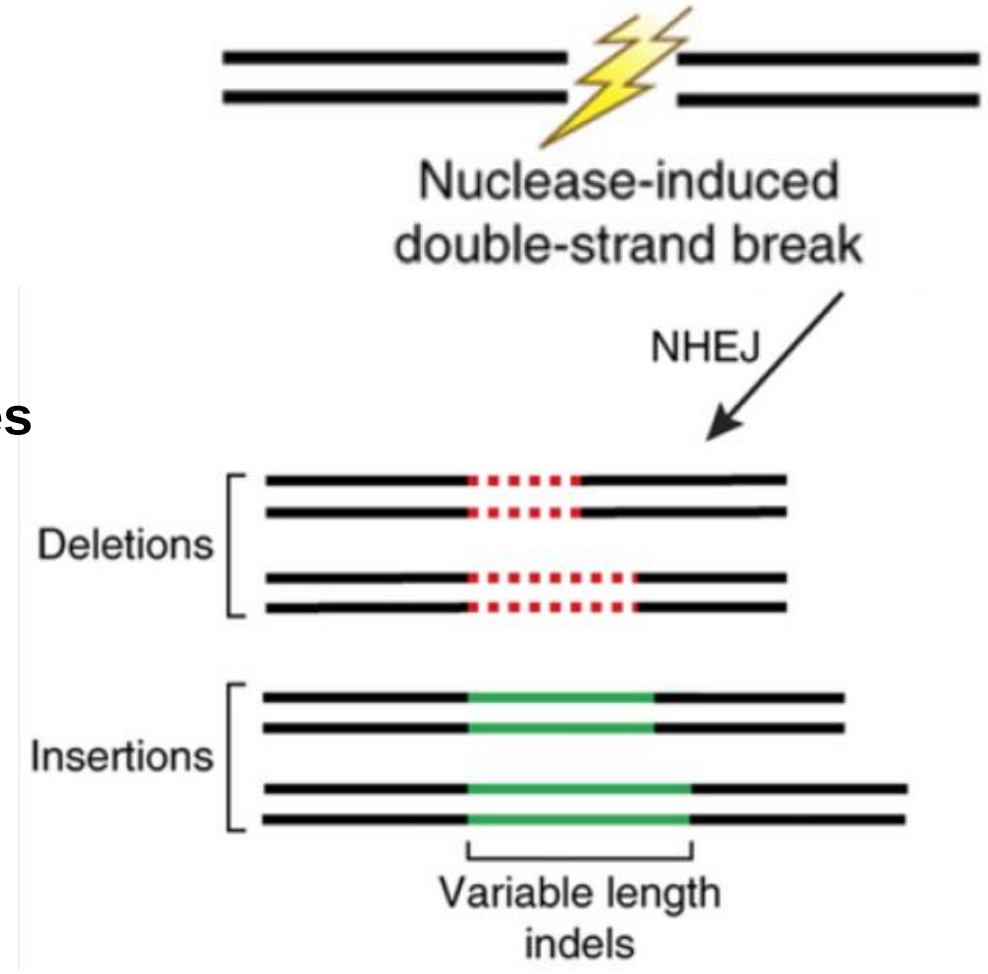
Mutagen	Chemical(e.g., EMS)	Physical (e.g., gamma, X-ray or fast neutron radiation)	Biological (ZFNs, TALENs or CRISPR/ Cas)	Biological- Transgenics (e.g., Agro or gene gun)
Disadvantages	Inefficient screening of desirable traits	Inefficient screening of desirable traits	Necessity of knowing gene function and sequences	Necessity of knowing gene function and sequences
	Non specific mutation	Non specific mutation	Prerequisite of efficient genetic transformation	Prerequisite of efficient genetic transformation
Other features	Non transgenic process and traits	Non transgenic process and traits	Transgenic process but non transgenic traits	Transgenic process and traits

General Principles

DNA repair system works when their will be **DNA double strand breaks** (DSBs)

DSB is repaired by Non-homologous end-joining (NHEJ)

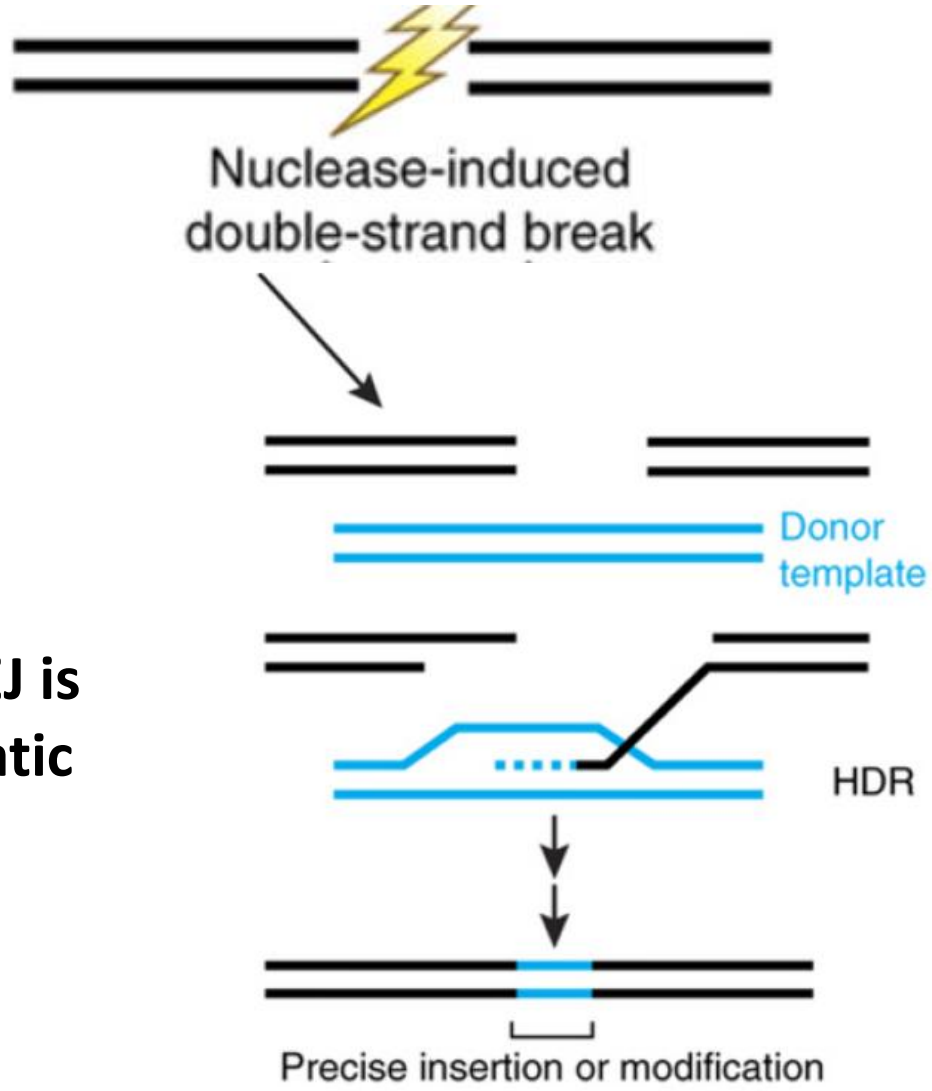
- **Rejoins the broken ends and is often accompanied by loss/gain of some nucleotides**
- **Thus the outcome of NHEJ is variable**



Refer Trends in Plant Science (2017) 22:38-52

General Principles

DSB is repaired by Homology directed repair (HDR)

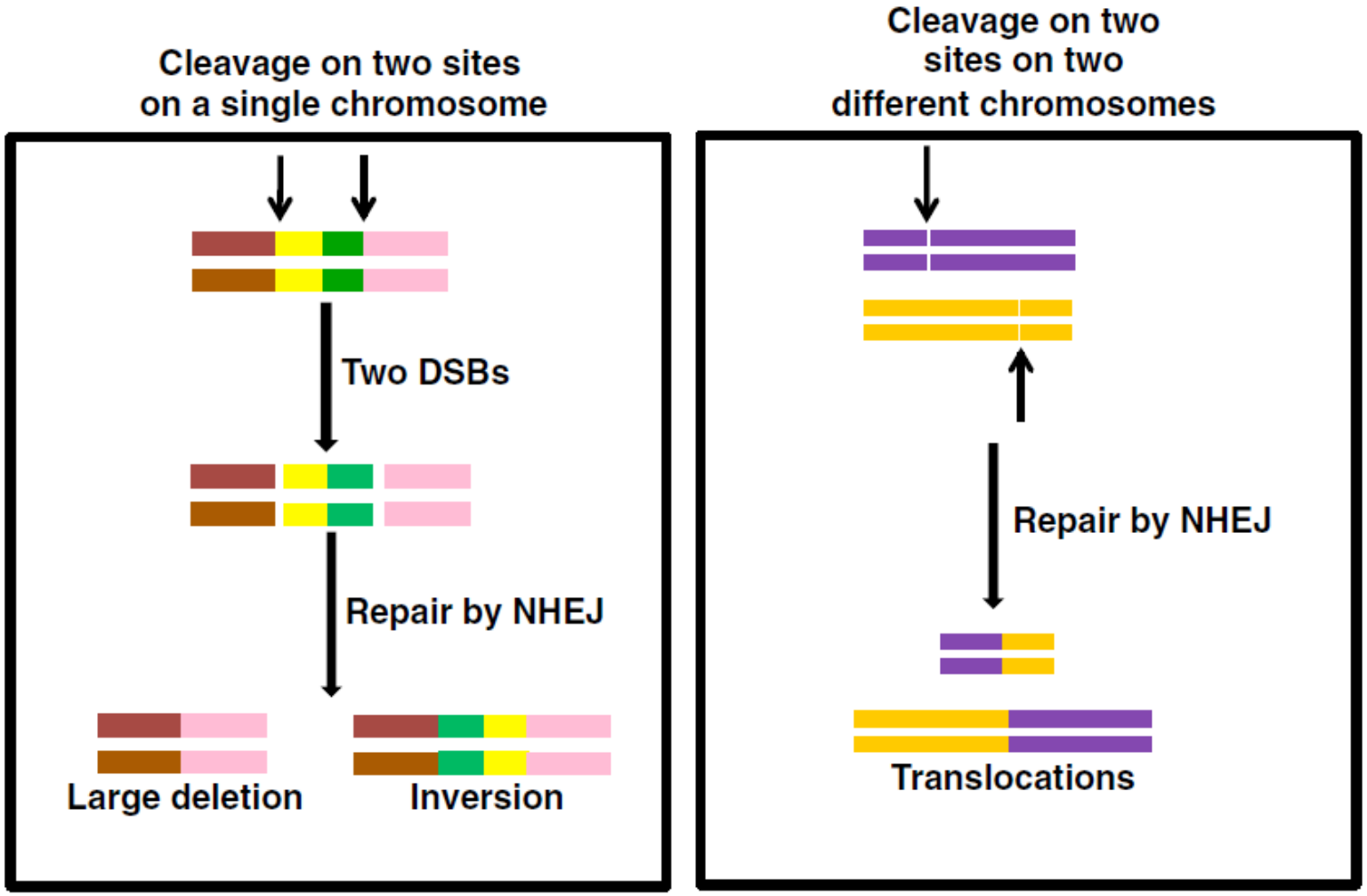


- **Repair DNA as a template to restore the DSBs**
- **Outcome of this kind of repair is precise and controllable**
- **Mutation induction by NHEJ is very high, as NHEJ is the preferred mechanism for DSB repair in somatic plant cells.**
- **HDR efficiency is very low**

Refer Trends in Plant Science (2017) 22:38-52

The Potential Outcomes of Genome Editing

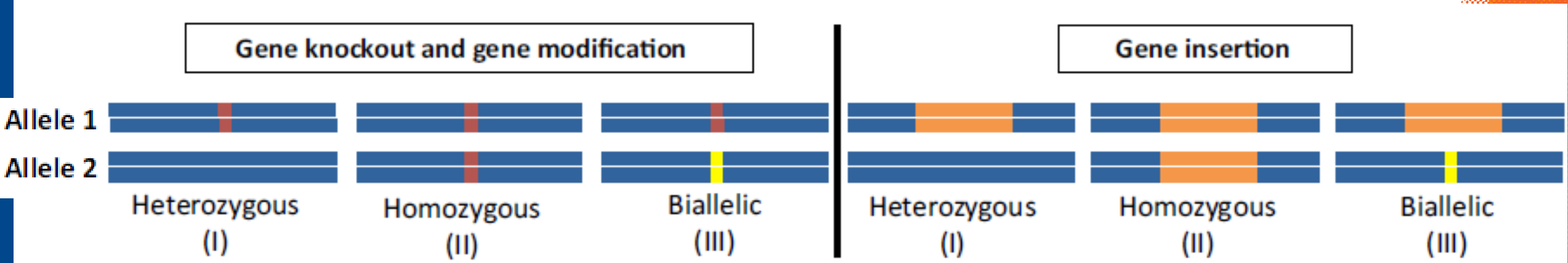
Transgenic Res (2016) 25:561–573



If the DSBs are generated on a single chromosome, the region between the two breaks can be deleted or inverted

If the DSBs are induced on two different chromosomes, a chromosomal translocation can occur

The Potential Outcomes of Genome Editing



potential consequences in **diploid** plants – the results of gene editing can be

(I) heterozygous (single allelic change)

(II) homozygous (identical changes to both alleles)

(III)biallelic (different changes at each allele) depending onwhich repair pathway is in operation.

Trends in Plant Science (2017) 22:38-52

Novel Tools for Genome Editing

- Mega-Nuclease: not novel
- Zinc-Finger Nucleases (ZFNs) :Zinc Finger technology was presented in 1991 in the journal *Science*.

TALEN Transcription Activator-like Effector Nucleases (TALENs)

- In 2009 the genome targeting abilities of TAL effectors was published and was used for genome editing and thus TALEN, emerged.
- Potential target sites and simple method of building TAL effector arrays, it was named "Method of the Year 2011" by journal *Nature*.

- CRISPR-Cas9

In 2013 CRISPR was demonstrated as a new genome editing tool.

https://www.broadinstitute.org/what-broad/areas-focus/project-spotlight/crispr-timeline

Mega-Nuclease

- First tool used for double strand break-induced genome manipulation
- Mega-nucleases are endonucleases characterized by a large recognition site i.e. of 12 to 40 base pairs, as a result this site generally occurs only once in any given genome.
- Meganucleases are considered to be most specific naturally occurring restriction enzymes
- Limitation: In these enzymes binding site and cleaveage site occur within same unit, hence difficult to modify the protein

Trends in Plant Science (2017) 22:38-52

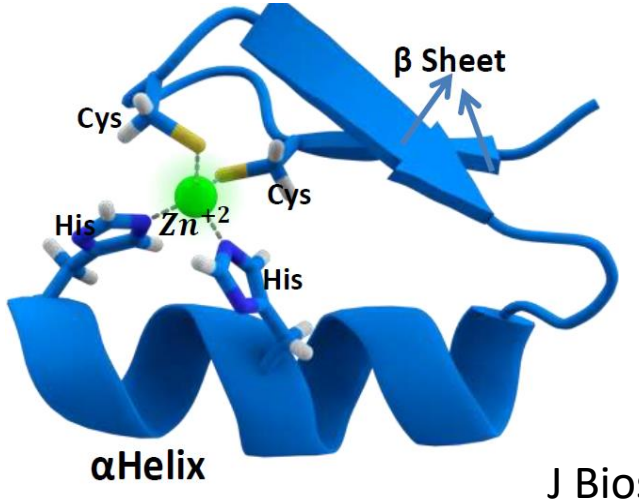
Zinc-Finger Nucleases (ZFNs)

These are hybrid restriction enzymes that contain two components:

N-terminal binding domain and a non-specific DNA cleavage domain at the C-terminal

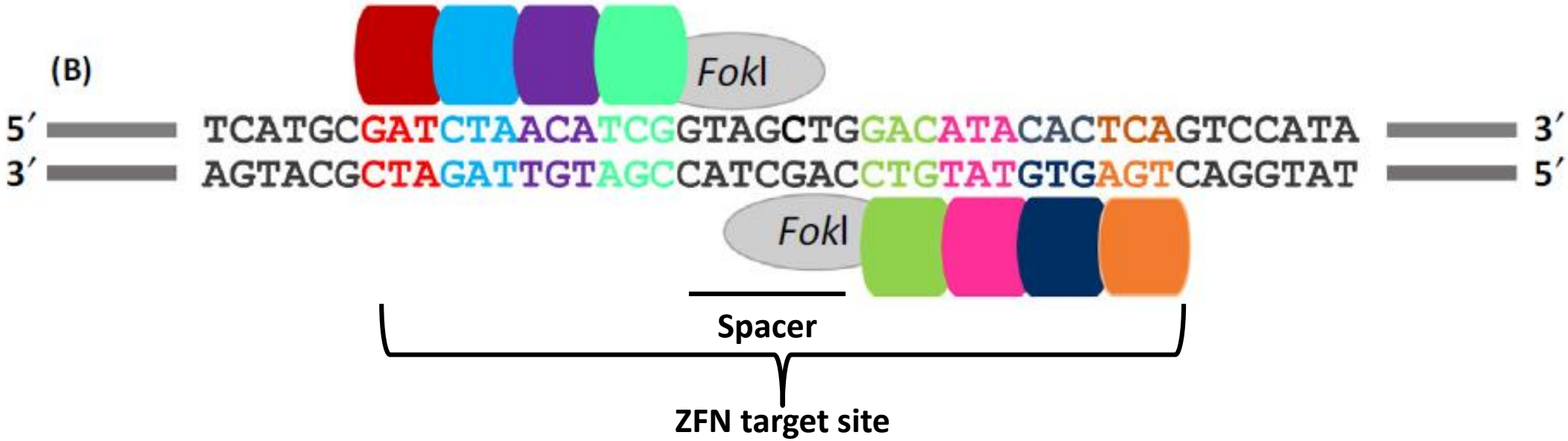
- **DNA binding domain contains a array of zinc-finger proteins**
  - They were first identified as a DNA-binding motif in Transcription factor TFIIIA from **African clawed frog**
  - Small protein structural motif that is characterized by the coordination of one or more zinc ions in order to stabilize the three dimensional structure
  - The C-terminal part of each finger is responsible for the specific recognition of the DNA sequence.

- **DNA-cleavage domain**
  - **FokI , a type IIS restriction enzyme**



J Biosci. 2012 Jul;37(3):483-91

Zinc-Finger Nucleases (ZFNs)



1 finger -3bp; 8 fingers -24bp

- Only limited sites in the genome can be targeted
- Construction is cumbersome and time consuming
- High Off-target mutagenesis 1:10, because some zinc-finger modules can bind degenerate sequences.

PLoS ONE(2009) 4(2): e4348

Trends in Plant Science (2017) 22:38-52

TALEN = Transcription Activator-like Effector Nucleases

- TAL-effectors: Bacteria: *Xanthomonas sp.*
- crop plant (rice, pepper & tomato) pathogens
- TAL-effectors mimicking eukaryotic transcription factors secreted into plants
- TALEs are transcriptional activators that specifically bind and regulate plant genes during infection

A plague for pepper and tomatoes but a boon for life scientists trying to edit genes: *Xanthomonas* transcription activator-like effectors.

Science. 2009, 326(5959):1509-12

Molecular Structure of TALE effectors

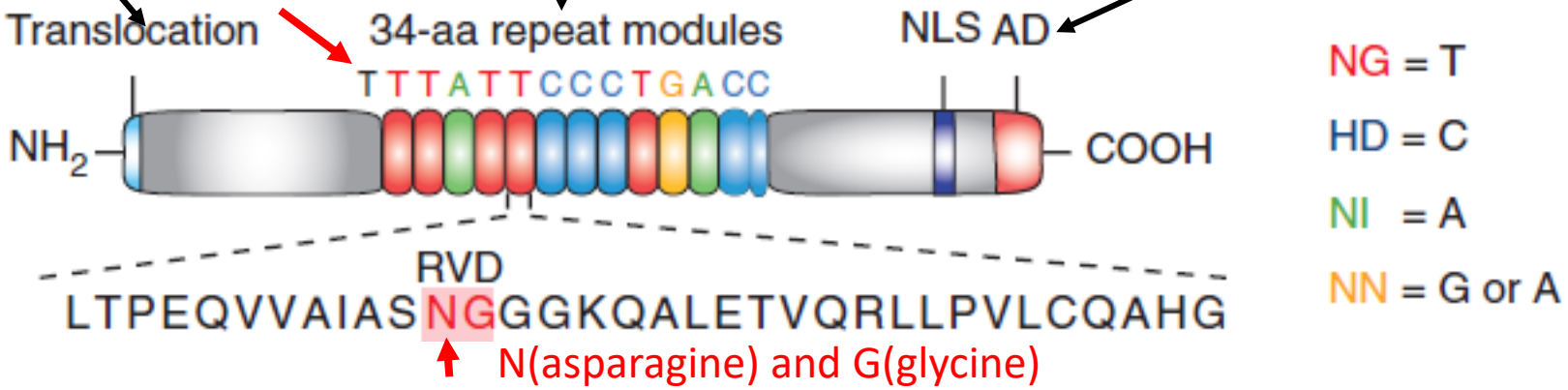
TALEs are organized into three sections

An N-terminal domain containing translocation signal

A central repeat domain that determines DNA binding specificity

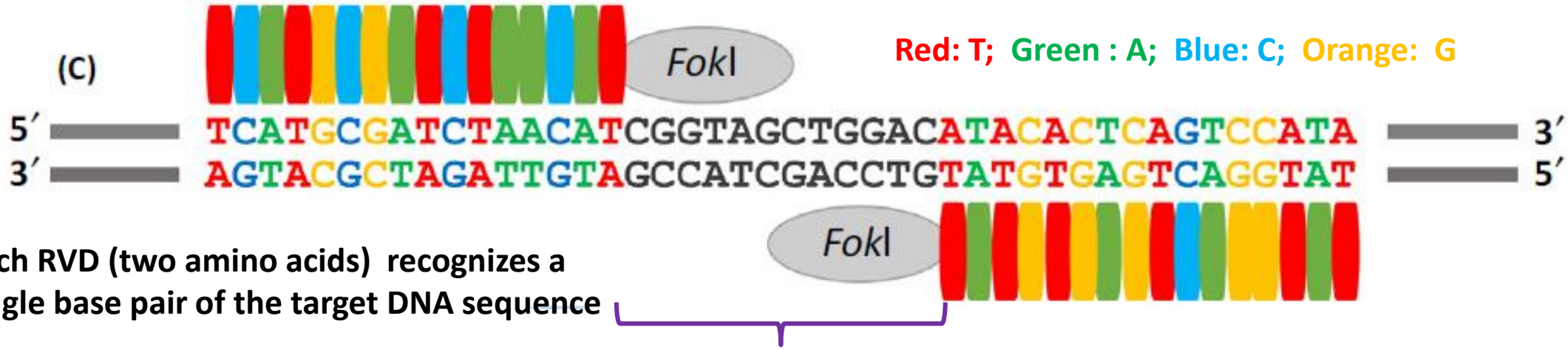
a C-terminal domain containing a nuclear localization signal and an activation domain

A stretch of 34 amino acid repeated at 1.5 - 33.5 times
The aa at the position of 12 and 13 form a Repeat variable diresidues (RVDs)



The amino acid identity of the RVD (Repeat variable di-residues) is responsible for DNA nucleotide recognition, enabling the design of TALENs to target unique DNA sequences

Transcription Activator-like Effector Nucleases



- TALEN contains DNA binding domain and catalytic domain of restriction endonuclease FokI.
- Compare to ZFNs: DNA binding specificity is higher, off-target effects are lower, and construction of DNA-binding domains is relatively easier
- Limitation: One limitation for the selection of TALEN targets is the requirement for thymidine “T” at the first position. Site selection may be made in most cases by varying the spacer sequence length

Trends in Plant Science (2017) 22:38-52

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

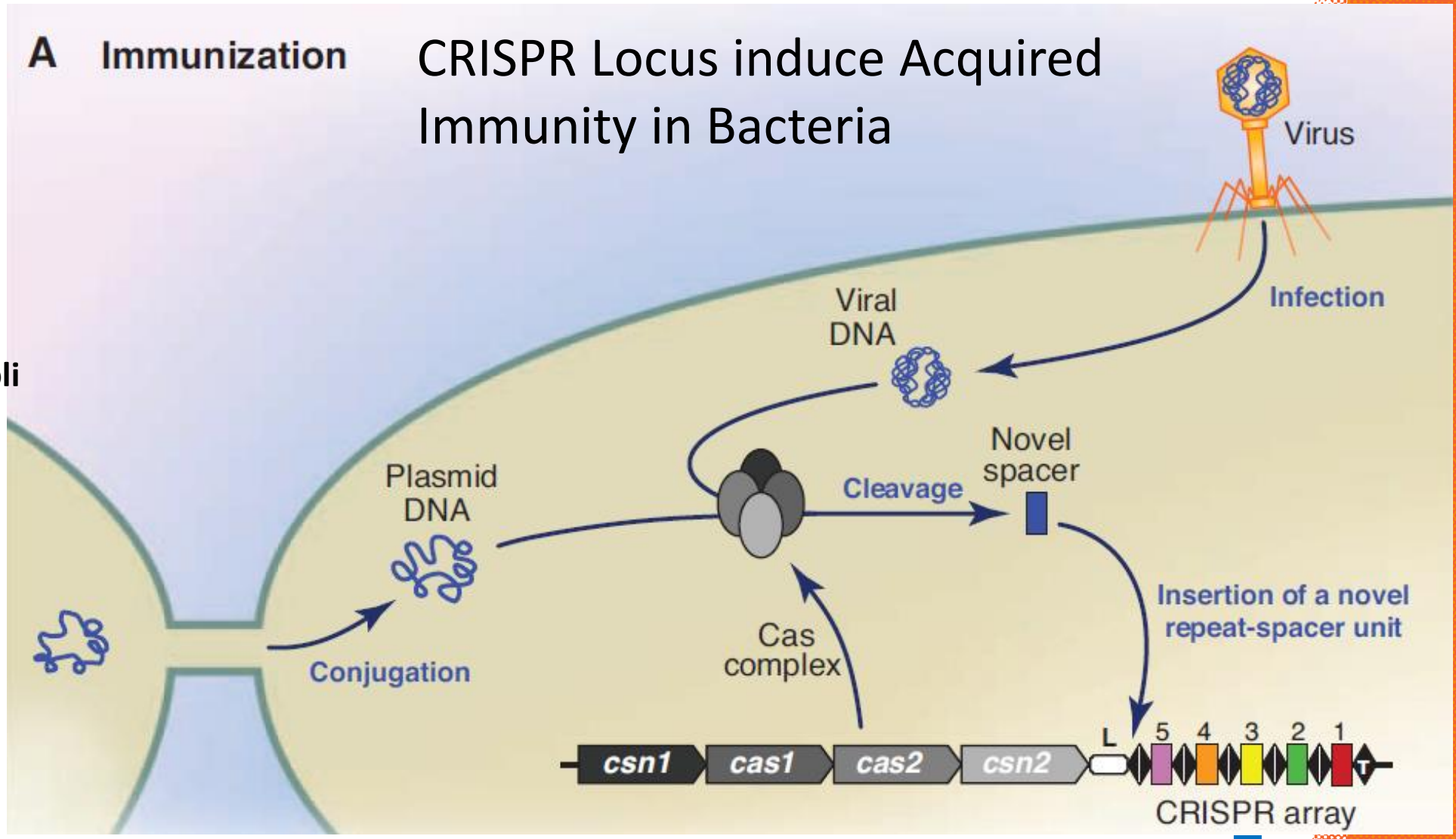
- Mechanism of adaptive immunity in bacteria and archaea
- Evolved to adapt and defend against foreign genetic material (e.g. phage)
- Several different types of CRISPR pathways in bacteria and archaea
- CRISPR-Cas9 can generate a **Double Strand Break** in the targeted DNA

Mechanism of CRISPR/Cas Based Immunity in Bacteria

CRISPR:
Clustered Regularly Interspaced Short Palindromic Repeats

Ishino et al (1987) first discovered the CRISPR sequences in Escherichia coli but cannot explain their function.

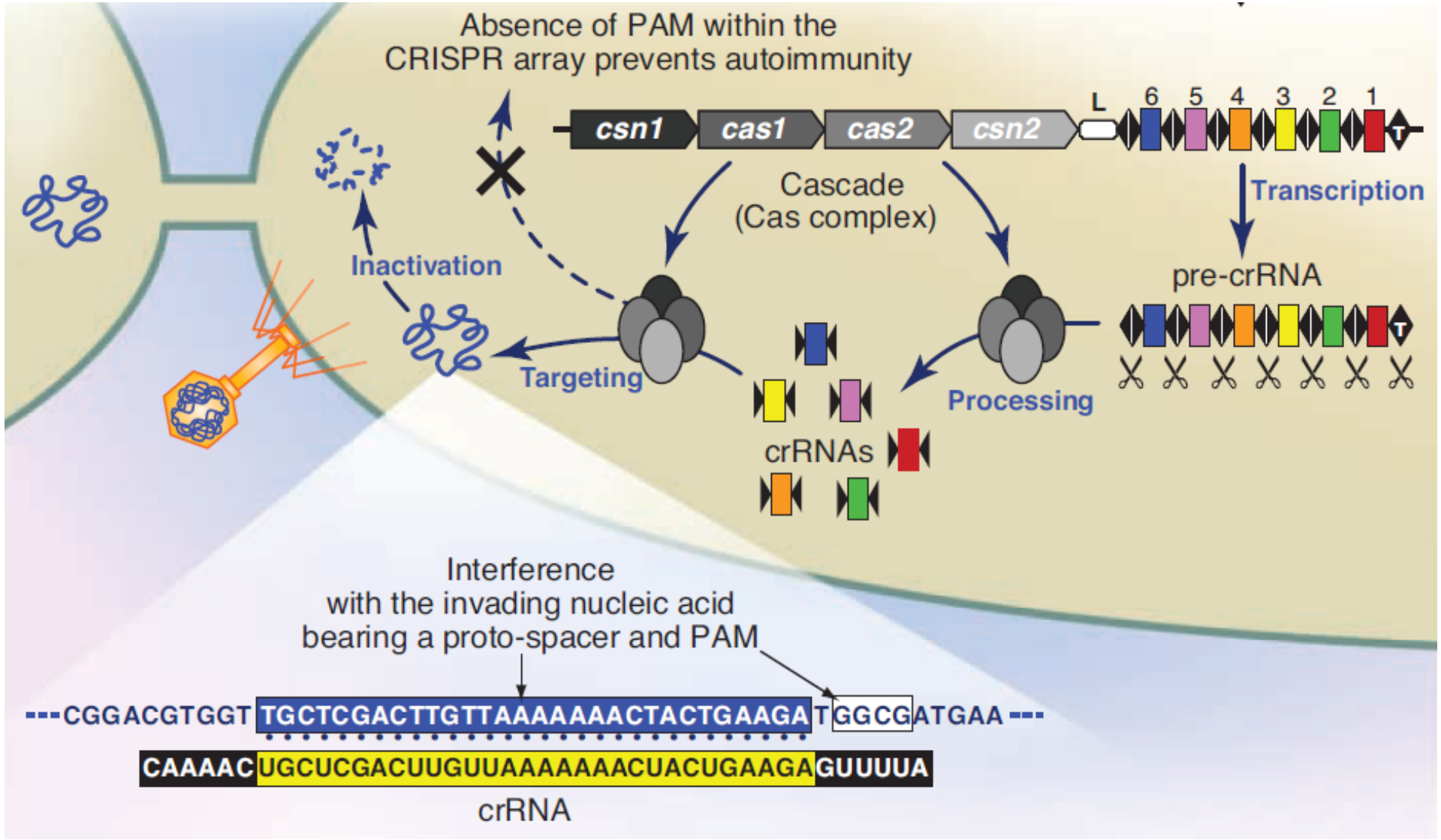
- **Adaptation**
- **CrRNA biogenesis**
- **Interference**



Horvath and Barrangou, 2010, Science: 167-170

Acquires immunity against the virus infection or plasmid transfer

Mechanism of CRISPR/Cas Based Immunity in Bacteria

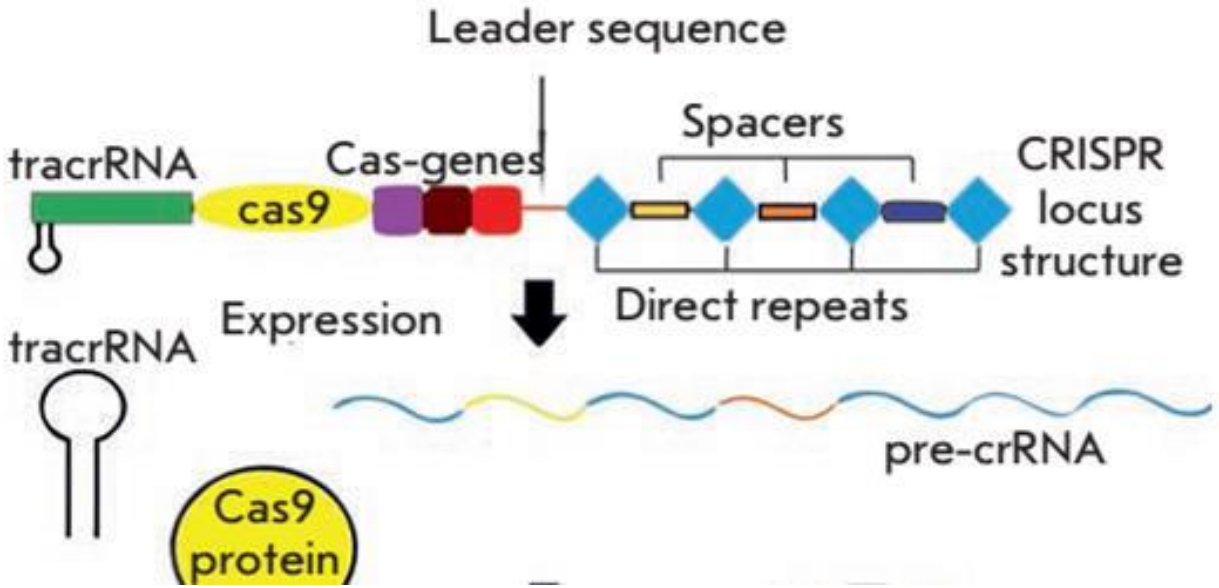


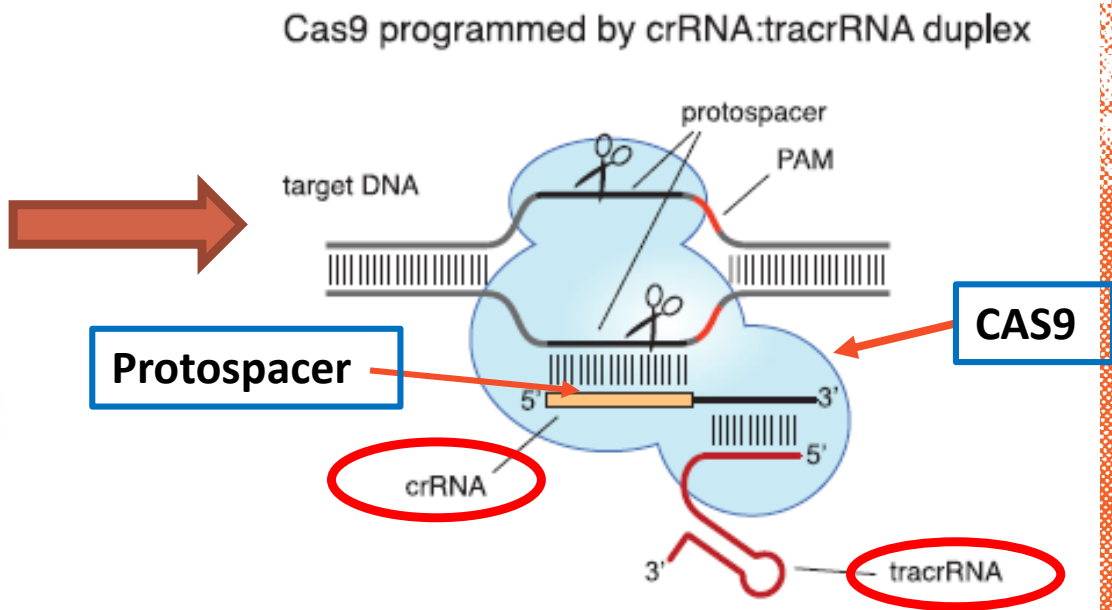
Horvath and Barrangou, 2010, Science: 167-170

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

1.Protospacer with adjacent motif (PAM); 2. CRISPR-RNA (crRNA)

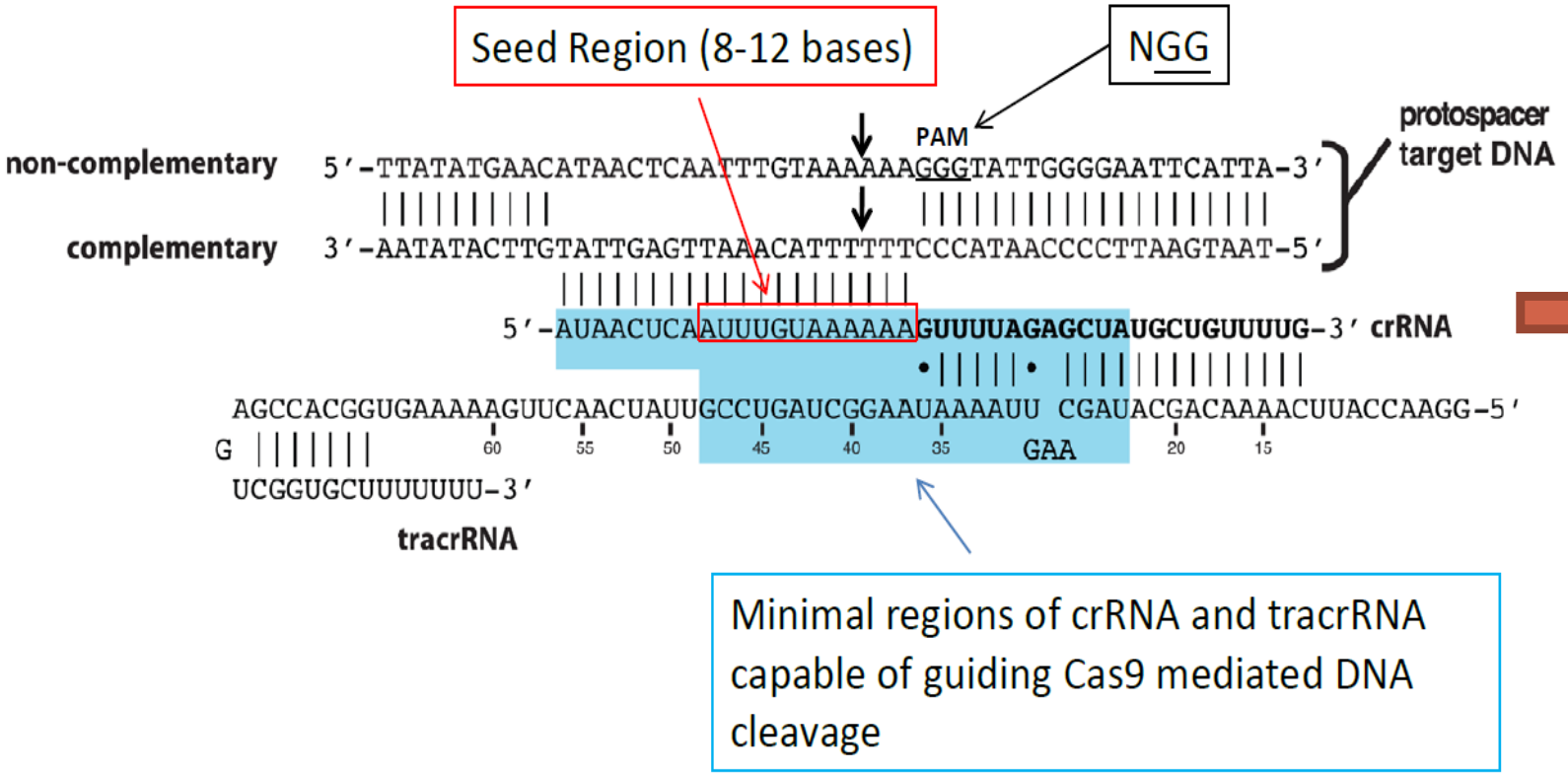
3. trans-activating crRNA (tracrRNA); 4. Cas9 protein

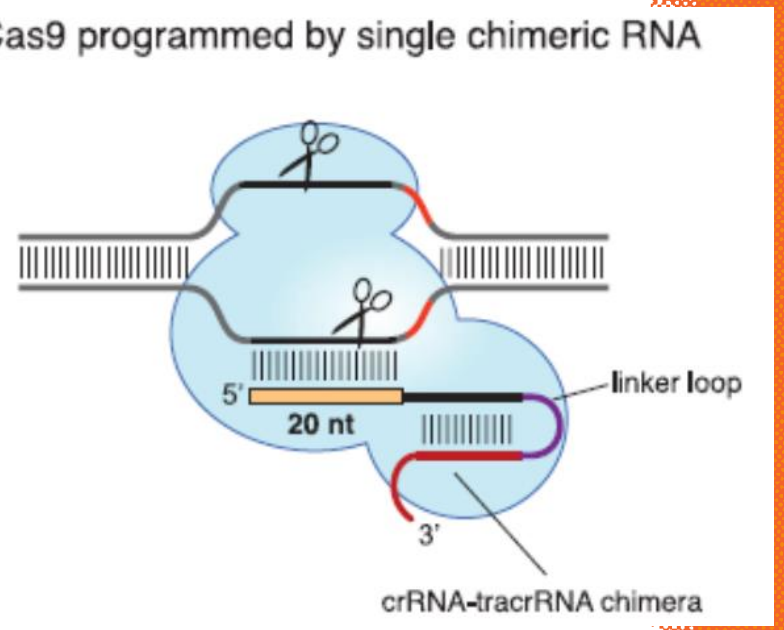




Science (2012, 337(6096):816-21

Clustered Regularly Interspersed Short Palindromic Repeats (CRISPR)





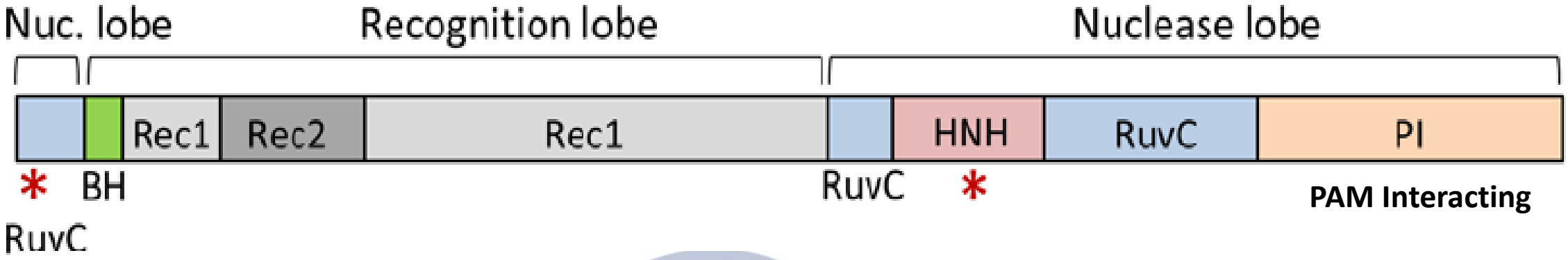
Jennifer Doudna

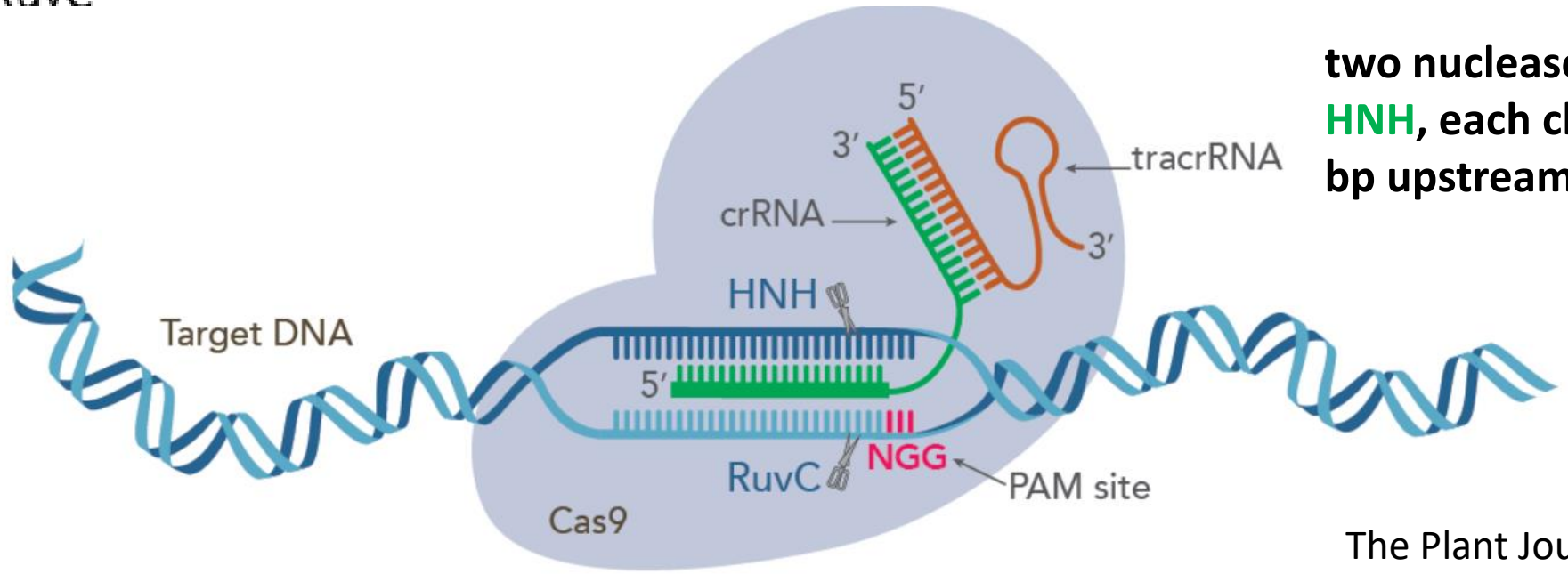
Science (2012, 337(6096):816-21

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

Red asterisks indicate catalytic sites for target DNA

CAS9 Protein Structure





two nuclease domains, RuvC and HNH, each cleaving one strand 3 bp upstream of the PAM

The Plant Journal, (2018), 94, 767–775

Types of CRISPR-Associated Proteins

Class I system:

- type I (Cas3), III (Cas10) and IV: employ several Cas proteins and crRNA to form multisubunit effector complexes
- the most abundant CRISPR/Cas systems found in both bacteria and Archaea

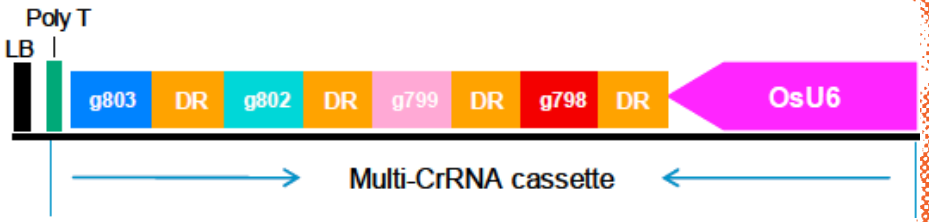
Class 2 system:

- including type II, V and VI: employ a single, large, multidomain protein
- much less common and largely restricted to bacteria
- SpCAS9 was the first identified Cas protein, naturally a complex of the large multi-domain protein Cas9 and two short RNAs, tracrRNA and crRNA

Types of CRISPR-associated proteins

Class 2 type V: Cas12/Cpf1

- tracrRNA is not required, only a small crRNA
- 21-35nt Direct repeat (crRNA)
- **PAM: TTTN- or TTTV** guided RNA , DSB occur at 13-23nt from PAM
- gRNA sequence: 23nt, more specific than 20nt for Cas9
- Generate staggered-ended DSB distal from the PAM
- Amenable to edit multiple genes
- Three Cas12a were identified from different bacte
  - Francisella novicida (FnCas12a) 5'-TTV(A/C/G)-3'
  - Lachnospiraceae bacterium ND2006 (LbCas12a) 5'-TTTN-3'
  - Acidaminococcus sp. BV3L6 (AsCas12a) 5'-TTTV-3'
- Not work well in my lab, require >37°C for higher enzyme activity



Types of CRISPR-associated proteins

Class 2 type VI: Cas13a

- Cas13a is a newly identified CRISPR effector that specifically cleaves **single-stranded RNA** in eukaryotic cells, opening up a wide range of new possibilities

- The RNA cleavage ability can be used for downregulation of specific transcripts with improved specificity over RNAi

- This enables diverse RNA manipulations, from base editing to base modification to visualization

- Just like Cas12, Cas13 proteins exhibit the ability to process precrRNA on their own without the involvement of tracrRNA

Types of CRISPR-associated proteins

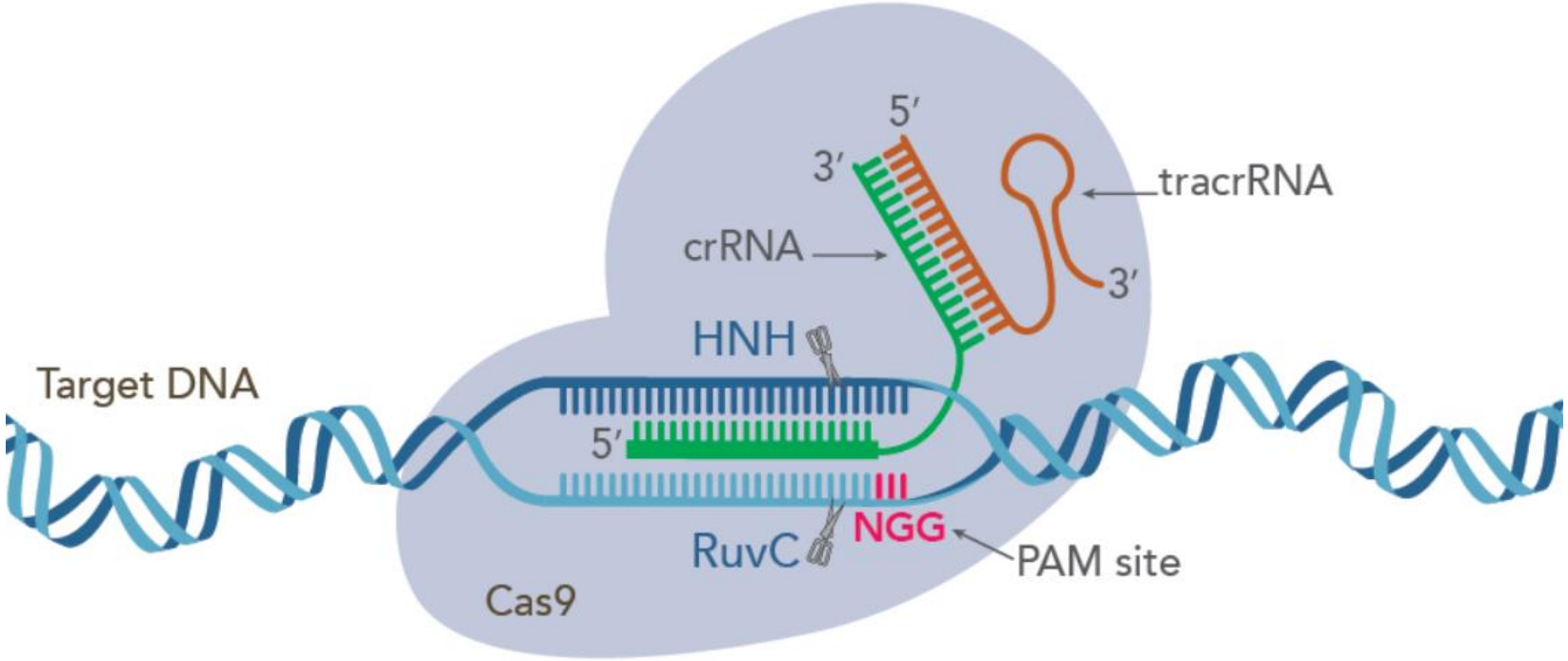
SpCAS9 (*Streptococcus pyogenes*): 1368 amino acids, 4.1bp
SaCAS9(*Staphylococcus aureus*): 1053 amino acids, 3.16kb
CjCas9 (*Campylobacter jejuni*): 984 amino acids, 2.5kb
SpCas9s: 880aa, under development by Cold-Spring Harbor laboratory

Table 1 Naturally occurring major CRISPR-Cas enzymes

	Size	PAM sequence	Size of sgRNA guiding sequence	Cutting site
spCas9	1368	NGG	20 bp	~ 3 bp 5′ of PAM
FnCas9	1629	NGG	20 bp	~ 3 pb 5′ of PAM
SaCas9	1053	NNGR RT	21 bp	~ 3 pb 5′ of PAM
NmCas9	1082	NNNNG ATT	24 bp	~ 3 bp 5′ of PAM
St1Cas9	1121	NNAGA AW	20 bp	~ 3 bp 5′ of PAM
St3Cas9	1409	NGGNG	20 bp	~ 3 bp 5′ of PAM
CjCas9	984	NNNNACAC	22 bp	~ 3 bp 5′ of PAM
AsCPf1	1307	TTTV	24 bp	19/24 bp 3′ of PAM
LbCpf1	1228	TTTV	24 bp	19/24 bp 3′ of PAM
Cas13	Multiple orthologs	RNA targeting	28 bp	

Nature Communication, (2018) 9:1911

Variants of CRISPR-SpCas9



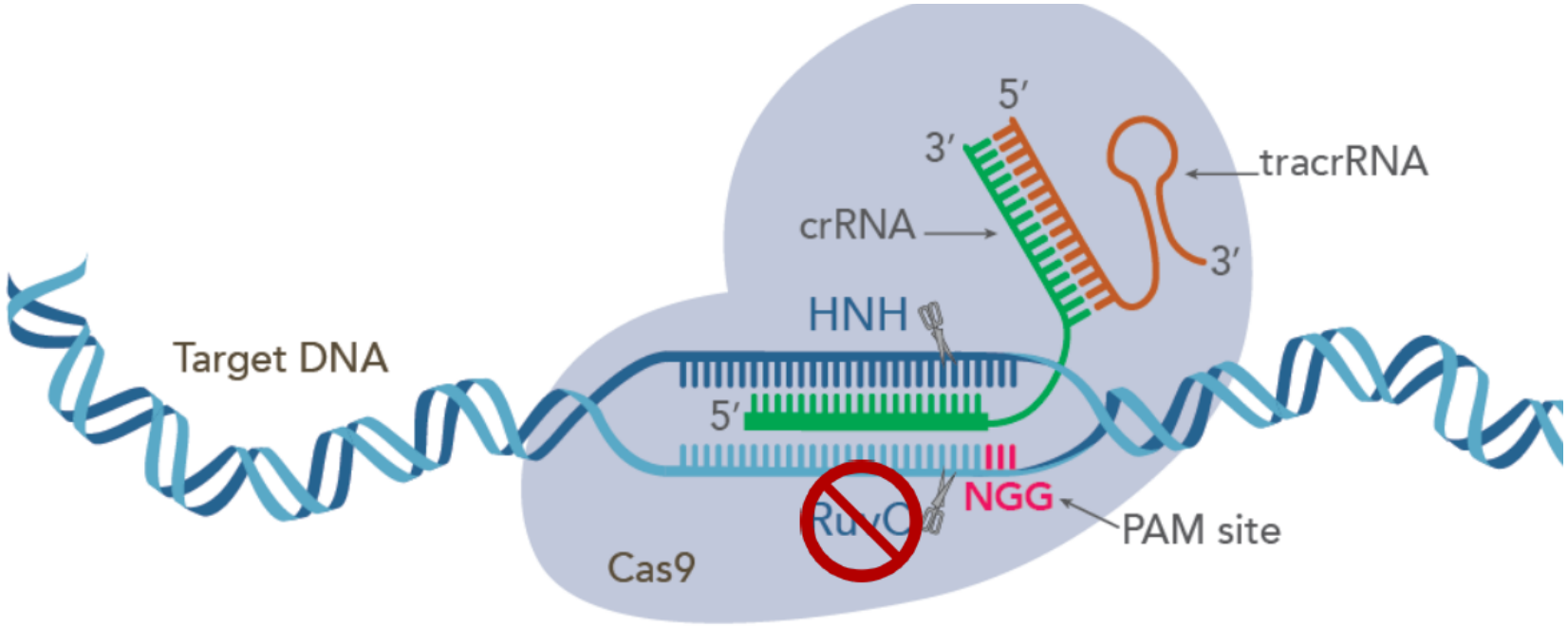
Wild type Cas9 → Cleaves both strands of target DNA

https://blog.addgene.org/crispr-101-cas9-nickase-design-and-homology-directed-repair

Variants of CRISPR-SpCas9

D10A Cas9 nickase cleaves the target strand

Position 10
Aspartic acid (D)
↓
Alanine (A)



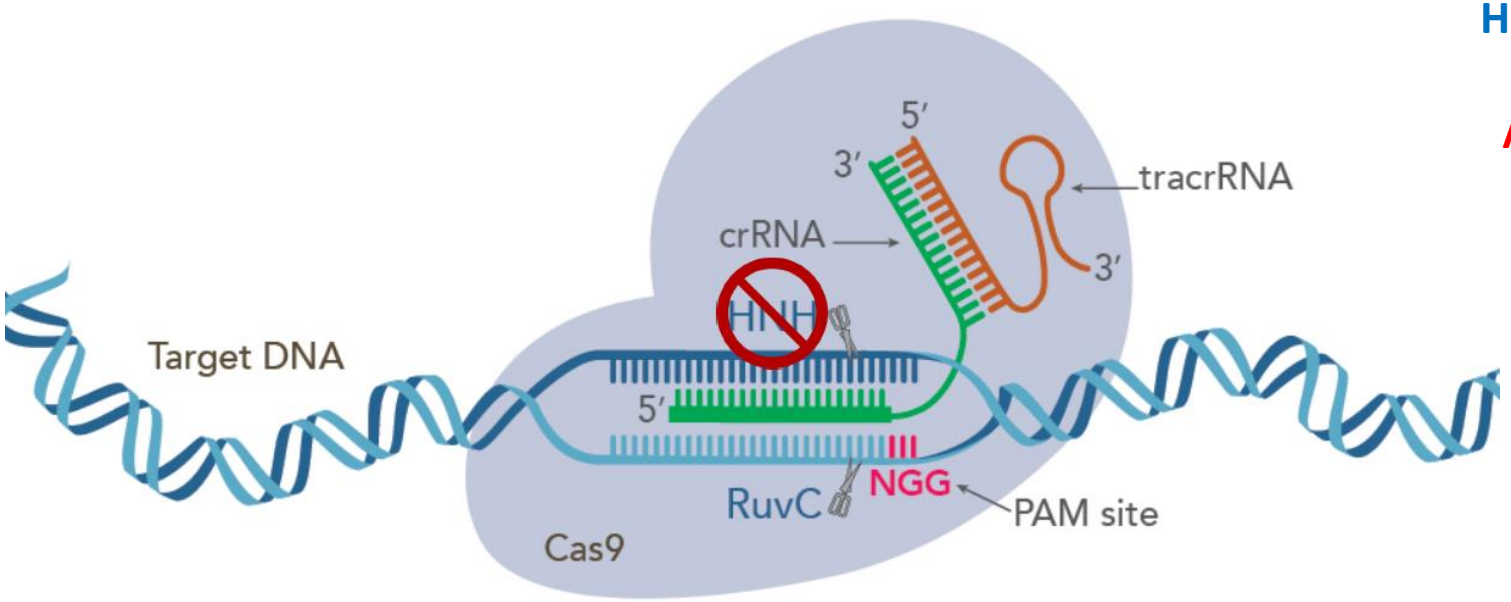
Cas9 D10A nickase has an inactivated RuvC domain → Cleaves the target strand

https://blog.addgene.org/crispr-101-cas9-nickase-design-and-homology-directed-repair

Variants of CRISPR-Cas9

H840A Cas9 nickase cleaves the non-target strand

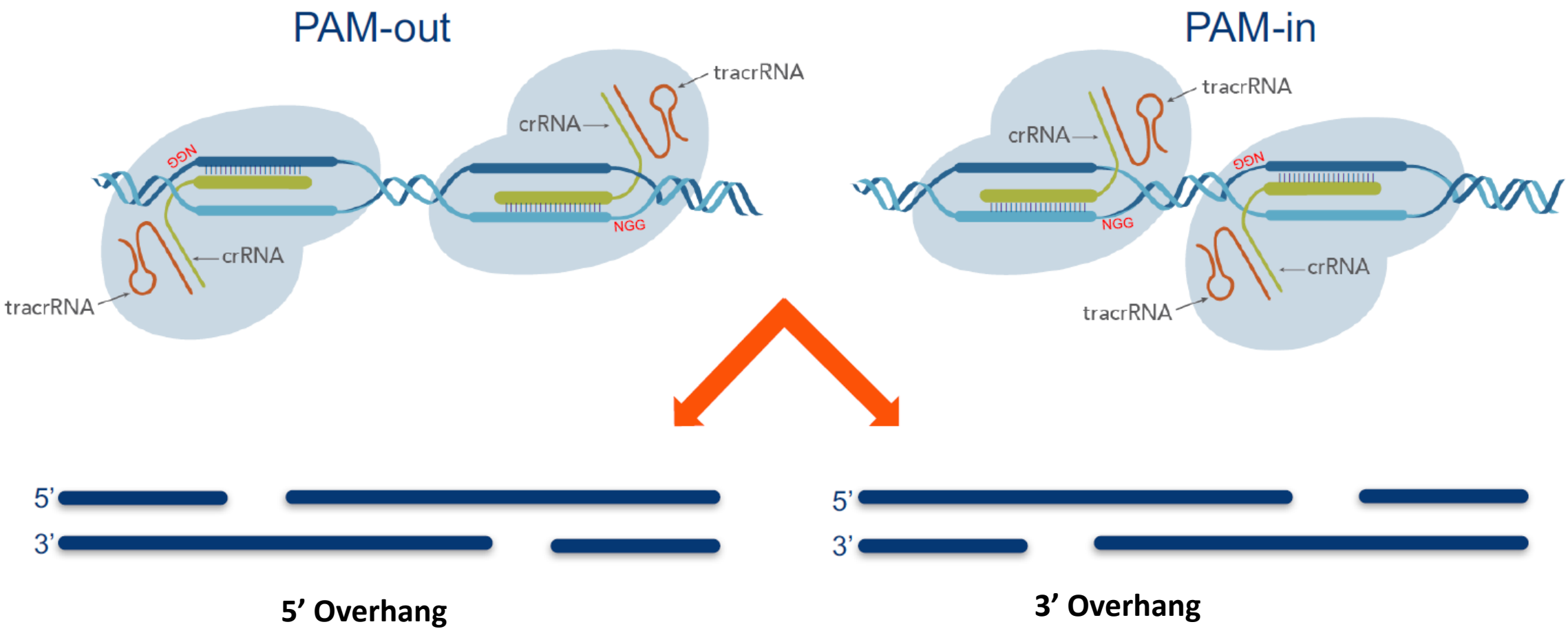
Position 840
Histidine(H)
↓
Alanine (A)



Cas9 H840A nickase has an inactivated HNH domain → Cleaves the non-target strand

https://blog.addgene.org/crispr-101-cas9-nickase-design-and-homology-directed-repair

Variants of CRISPR-Cas9



5' Overhang

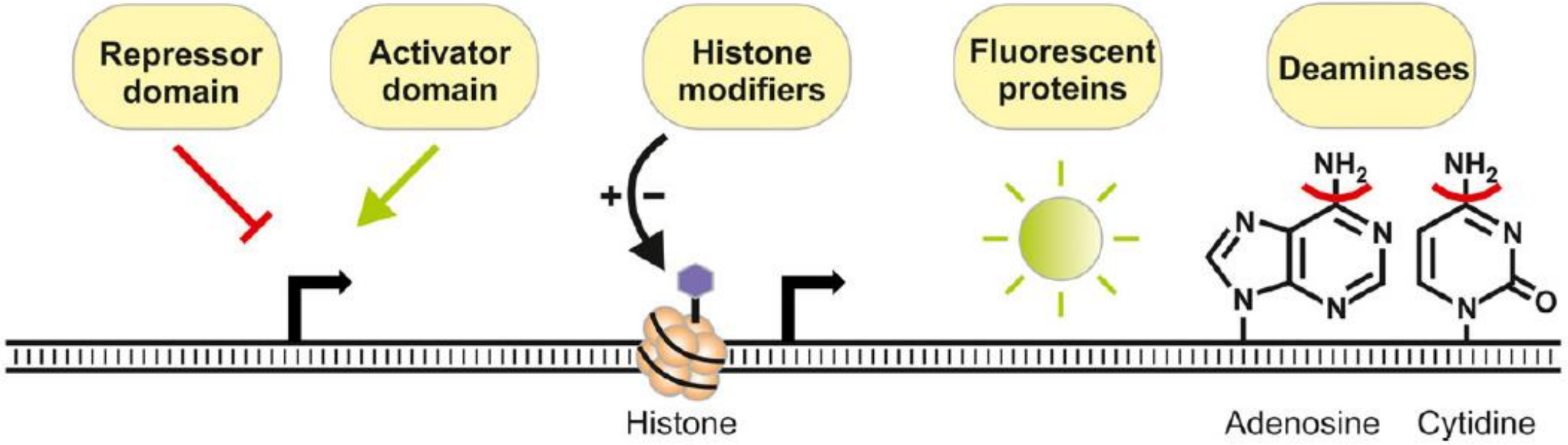
3' Overhang

Different staggered DSB

https://blog.addgene.org/crispr-101-cas9-nickase-design-and-homology-directed-repair

Applications of CRISPR/Cas system

- dead Cas9 (dCas9) can mediate transcriptional down-regulation or activation, particularly when fused to activator (VP64) or repressor domains (SRDX)
- Fusion with chromatin or DNA modification domains
- Fusion with GFP for live-cell imaging of chromosomal loci
- Fuse deaminases with nickase for single-base editing (we are doing this)



Variants of CRISPR-Cas9

- SpCas9-WT (mid off-target, very high activity)
- SpCas9-HF1 (very low off-target, low activity)
- eSpCas9(1.1) (low off-target, high activity)
- hypaCas9 (very low off-target, high activity)
- xCAS9 (NGG, NGA, NGT) (low off-target, high activity)
- SpCas9-NG (NGG, NGA, NGT) (mid/low off-target, high activity)

We have developed plant-codon-optimized hypa Cas9 and xCas9

Breeding with CRISPR-Genome Editing

4 stages:

1.Selection of target nt sequence in the genome

- off target effects
- avoid repeated and homologous sequence
- http://crispor.tefor.net/

2.Generation of a nuclease construct directed at the selected target

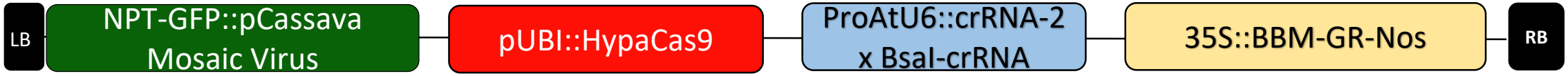
3.Delivery of this construct to the cell nucleus

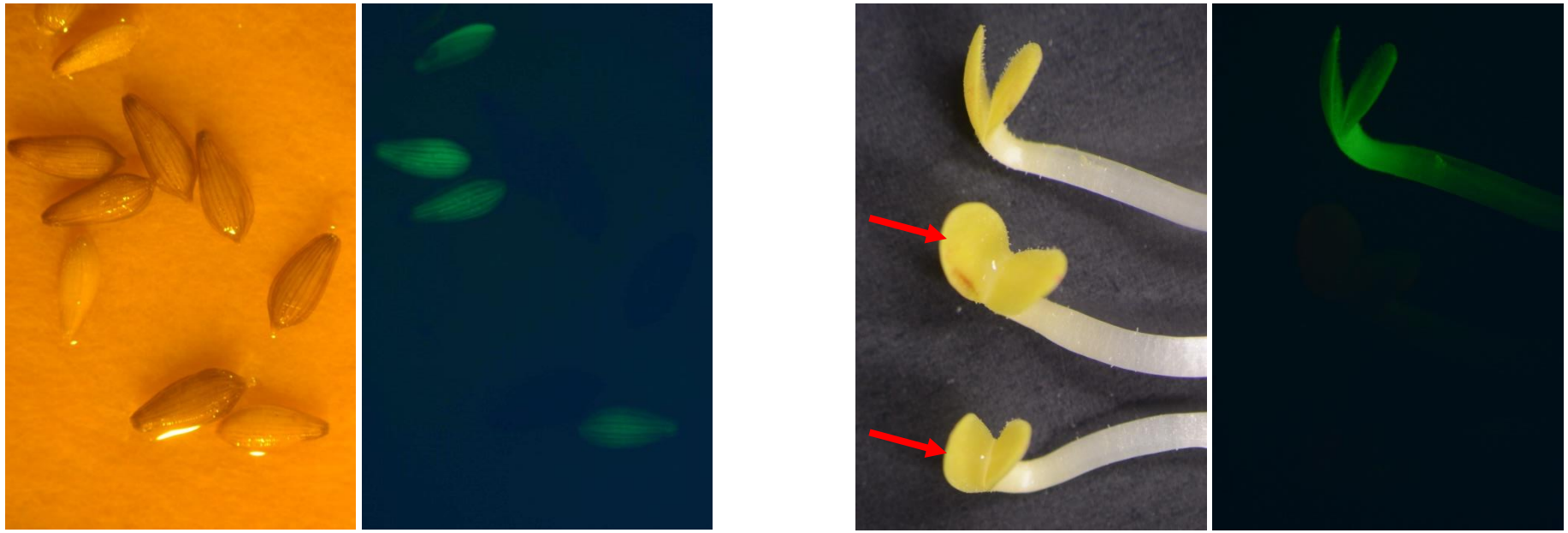
Agrobacterium Transformation/ Bombardement

4.Analysis of produced mutants

Breeding with CRISPR-Genome Editing

2.Generation of a nuclease construct directed at the selected target

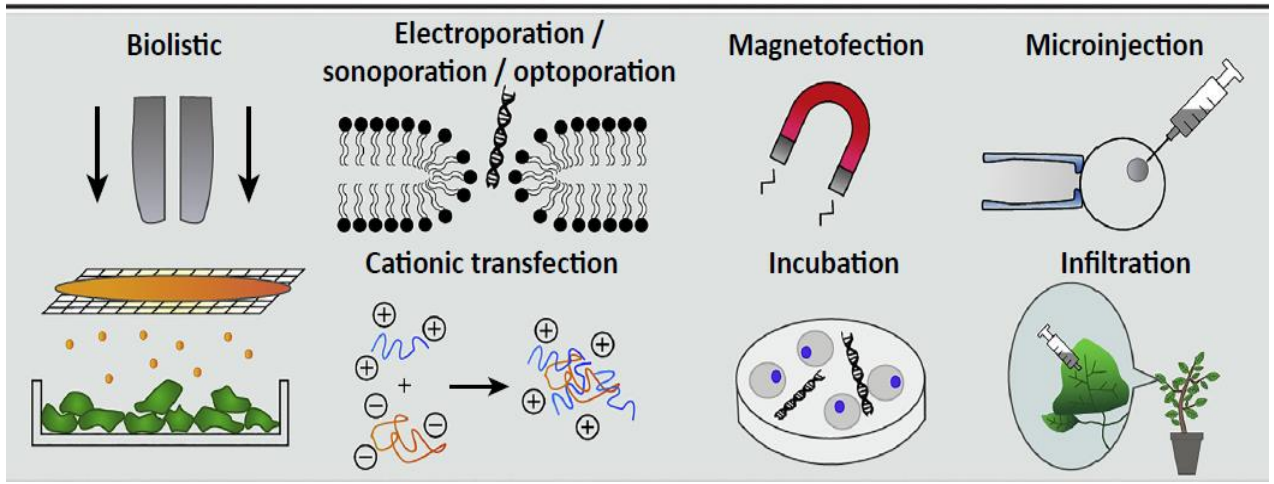


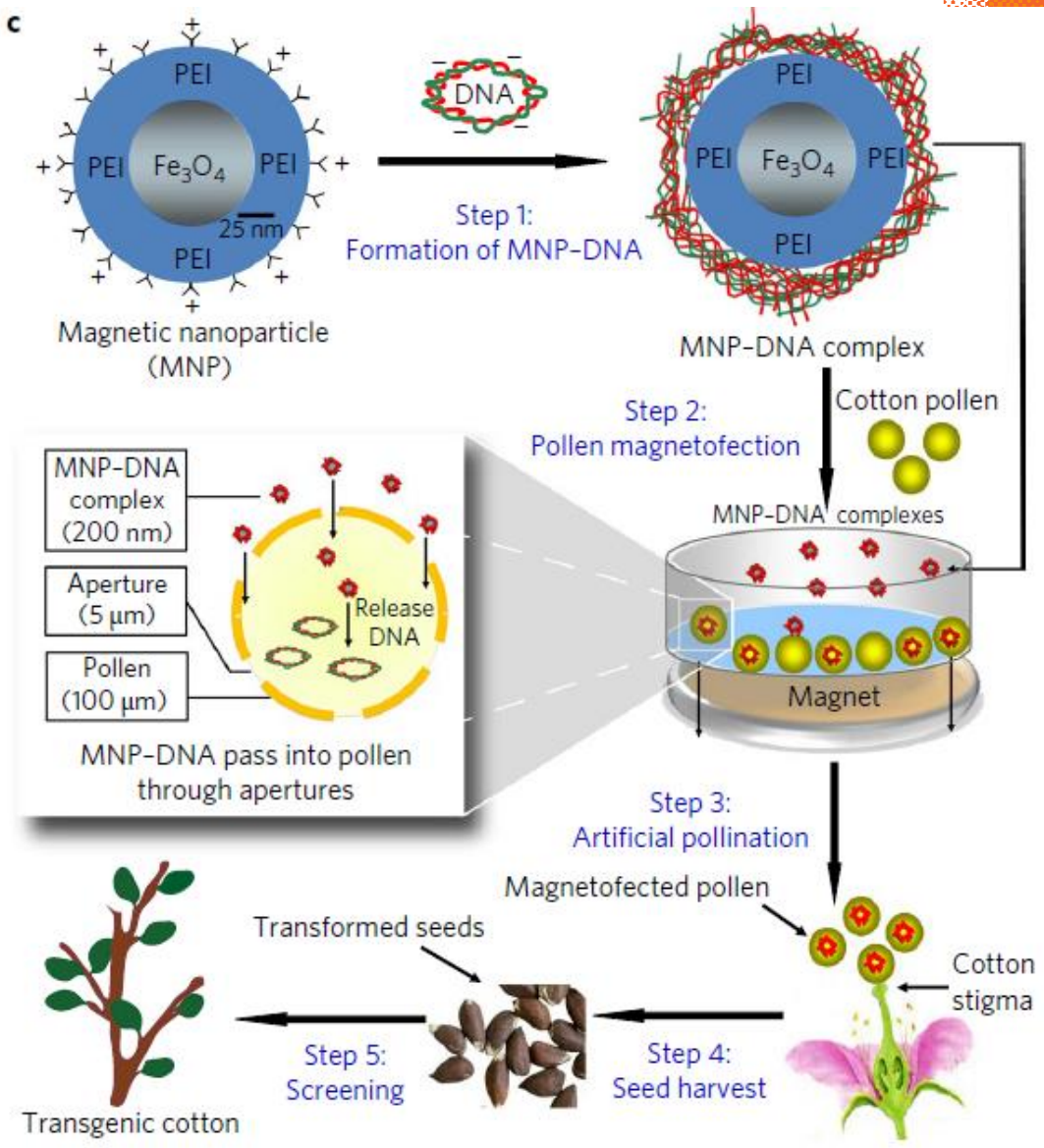


F2 Segregated Seeds

DNA-Free Gene-editing

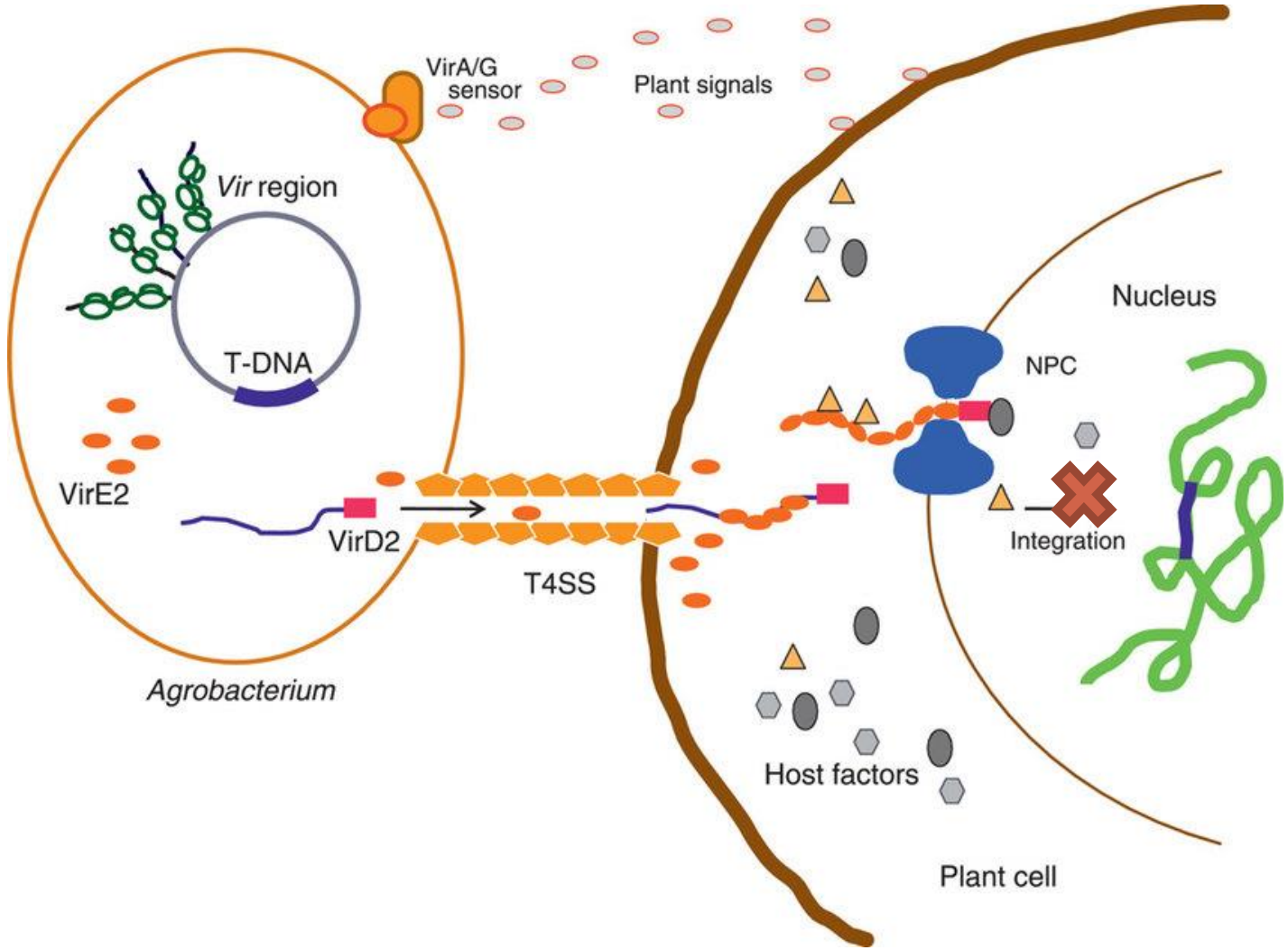
- Tobacco rattle virus (TRV)-based expression system for indirect transient delivery of CRISPR/CAS9 into plant tissues
- Direct delivery of CAS9/sgRNA complex into plant cells such as protoplast
- Bombardment of Cas9/sgRNA complex into plant tissues
- Nanoparticles: mixture nanoparticles with CRISPR/CAS9 Complex with pollen



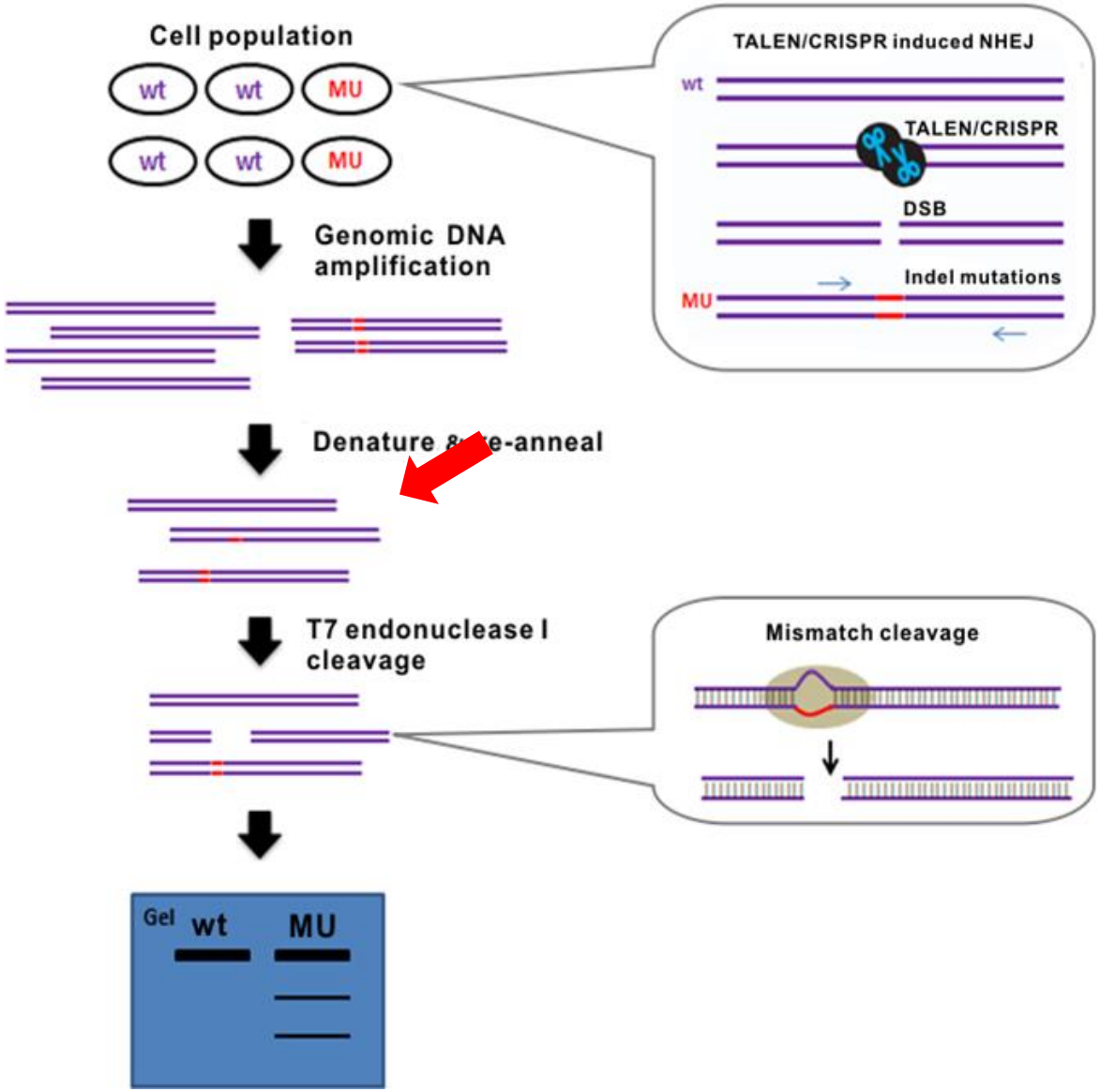


DNA-Free Gene-editing

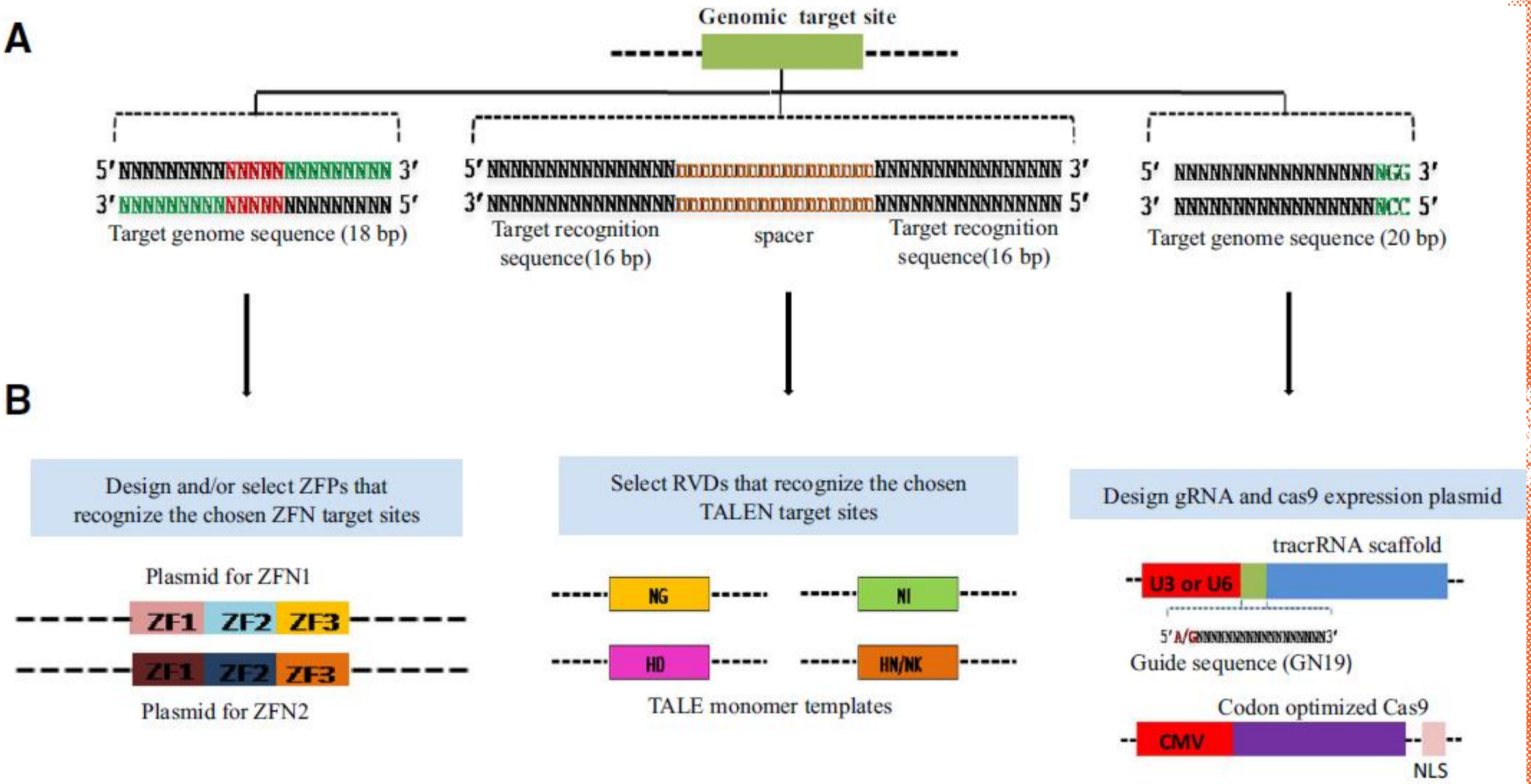
Our own DNA-free gene-editing method (under development)

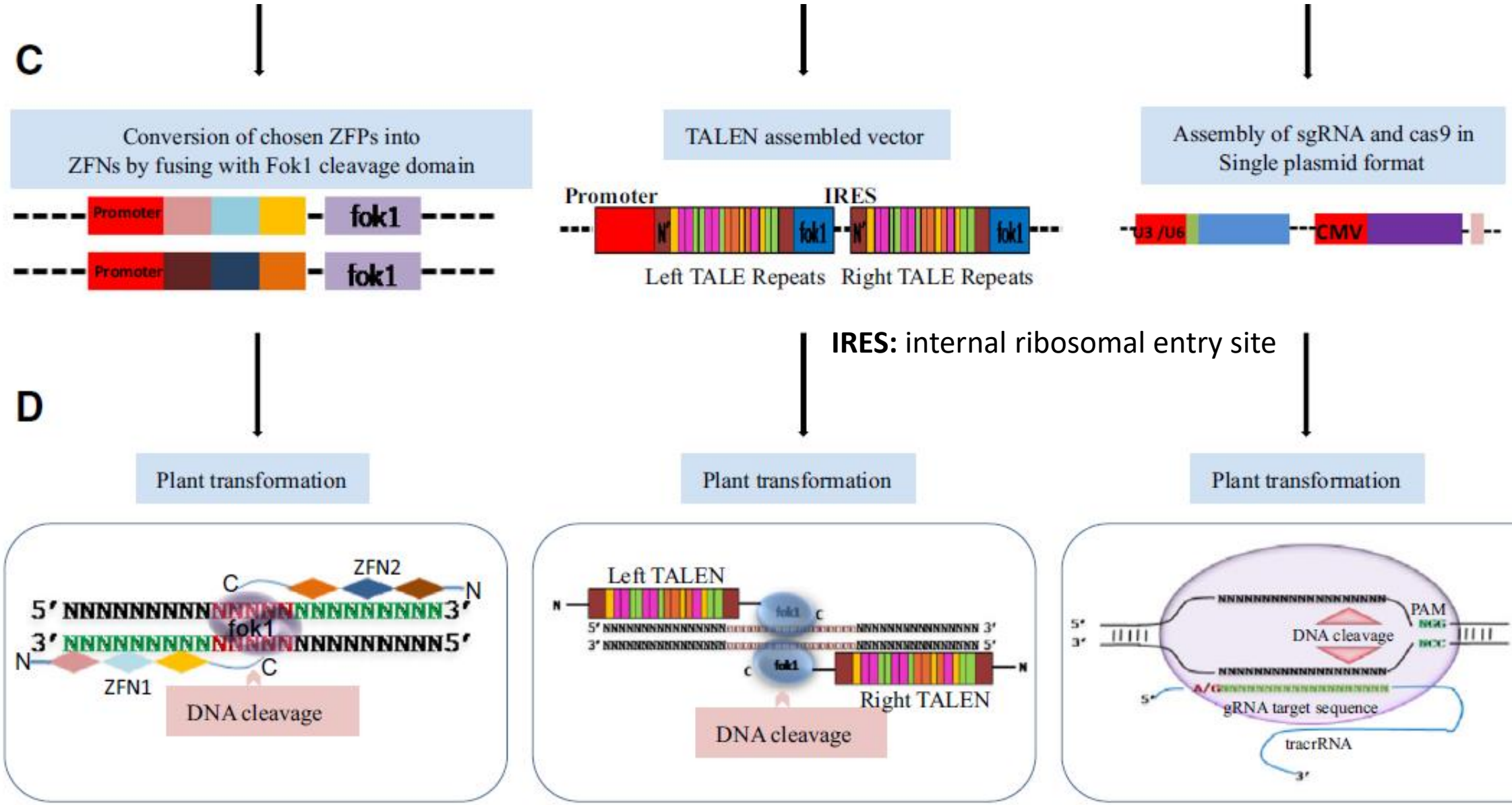


4. Analysis of produced mutants
T7E1 assay, DNA sequencing, melt curve (Qpcr)



T7 Endonuclease I can only identify and cut mismatch sequence


A
Genomic target site
5′ NNNNNNNNNNNNNNNNNNNN 3′
3′ NNNNNNNNNNNNNNNNNNNN 5′
Target genome sequence (18 bp)
5′ NNNNNNNNNNNNNNNNNNNN 3′
3′ NNNNNNNNNNNNNNNNNNNN 5′
Target recognition sequence(16 bp)
spacer
Target recognition sequence(16 bp)
5′ NNNNNNNNNNNNNNNNNNNNGG 3′
3′ NNNNNNNNNNNNNNNNNNNNCC 5′
Target genome sequence (20 bp)
B
Design and/or select ZFPs that recognize the chosen ZFN target sites
Plasmid for ZFN1
ZF1 ZF2 ZF3
ZF1 ZF2 ZF3
Plasmid for ZFN2
Select RVDs that recognize the chosen TALEN target sites
NG
NI
HD
HN/NK
TALE monomer templates
Design gRNA and cas9 expression plasmid
tracrRNA scaffold
U3 or U6
5′A/GNNNNNNNNNNNNNNNNNN3′
Guide sequence (GN19)
Codon optimized Cas9
CMV
NLS



IRES: internal ribosomal entry site

Advantages of CRISPR/Cas9 system

- Targeting of both ZFNs and TALENs relies on DNA sequence recognition by DNA-binding proteins; CRISPR/ Cas9 technology utilizes a novel RNA-guided targeting approach for sequence identification and binding
- The complexity and difficulties of protein design and synthesis is one of the primary barriers to the scientific community for routine utilization of ZFNs and TALENs as genome editing tools; For CRISPR/CAS9 only 20-nt in the sgRNA needs to be modified to recognize a different target
- Generating plants carrying multiple mutated genes using ZFNs or TALENs has proven to be inefficient, time consuming and labor intensive. It is highly efficient for editing multiple targets.
- Simpler to design and implement and less expensive

Thank You!