

# CRISPR

Clustered Regularly Interspaced Short Palindromic Repeats



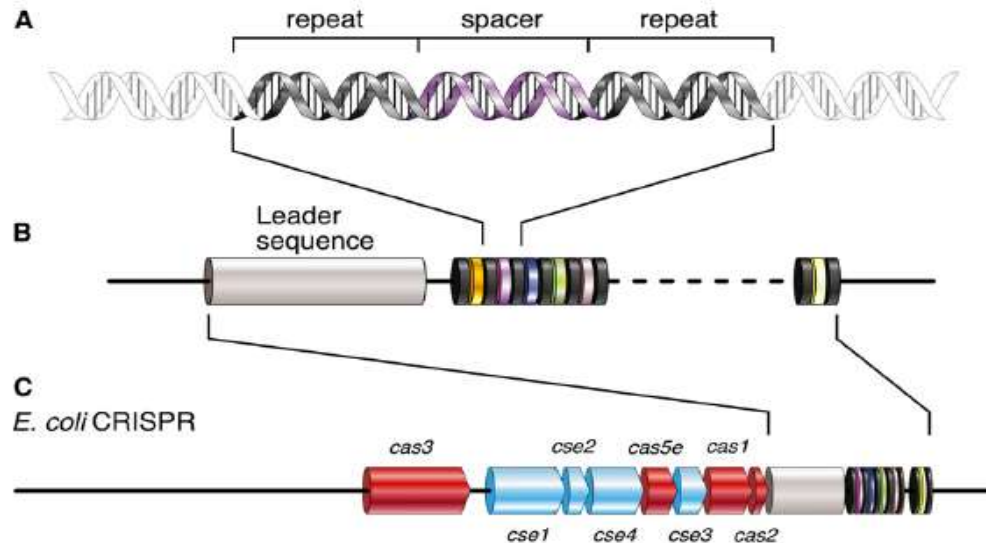
Le regioni CRISPR sono essenzialmente delle banche di memoria di sequenze fagiche ostili.

Una CRISPR è costituita dalla ripetizione di diversi frammenti fagici (SPACERS) alternate con sequenze ripetute identiche (REPEATS).

Il sistema CRISPR conferisce resistenza ai fagi che contengono nel proprio genoma una sequenza identica o strettamente correlata a quella contenuta nei CRISPR.

Sono quindi regioni presenti nel genoma della gran parte dei batteri ( 70% nei Batteri, 90% degli Archea)

# Elementi caratteristici di un locus CRISPR



In genere si ritrova un locus CRISPR per genoma anche se non è raro trovare genomi con diversi CRISPR loci.

Un locus CRISPR è caratterizzato dall'allineamento di sequenze ripetute **REPEATS** praticamente identiche all'interno di un locus separate da sequenze **SPACERS** di lunghezza identica ma di contenuto in basi molto diverso.

I **REPEATS** variano in lunghezza da specie a specie dalle 21 alle 47 basi (con una media di 32 basi) mentre la lunghezza degli **SPACERS** varia tra i 20 e 72 bp. I **REPEATS** sono caratterizzati dalla capacità di formare un'ansa (sono sequenze parzialmente palindromiche). Il numero di unità ripetute di **SPACERS/REPEATS** per ogni locus è di circa 25 (con oscillazioni da 1 a 300!)

Oltre ai REPEATS and SPACERS il locus CRISPR è caratterizzato:

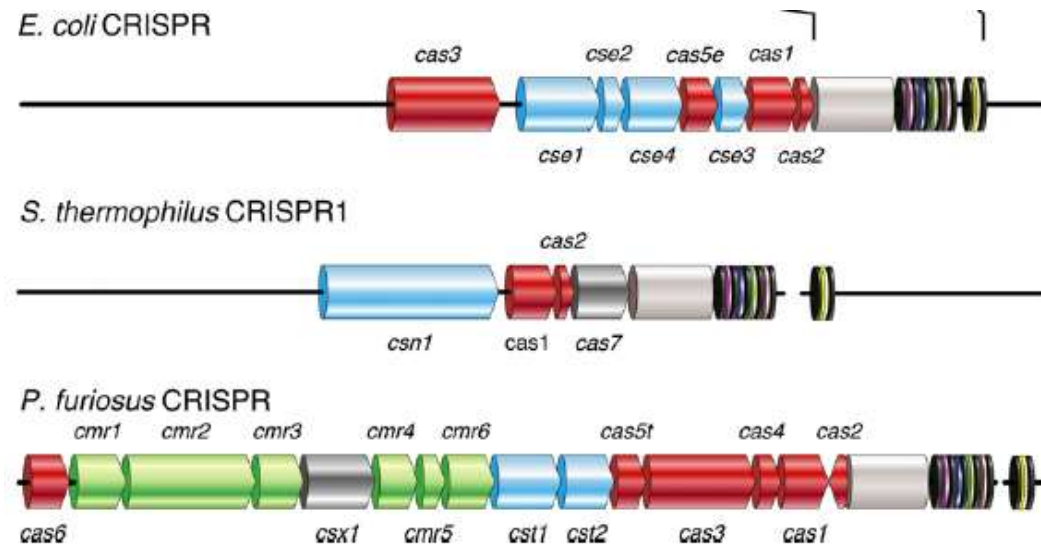
- dalla presenza di una regione definita LEADER lunga diverse centinaia di basi e non codificante posizionata sempre e solo ad una estremità del locus.



- Dalle proteine associate ai CRISPR (CAS) che sono codificate da geni localizzati a monte delle sequenze CRISPR



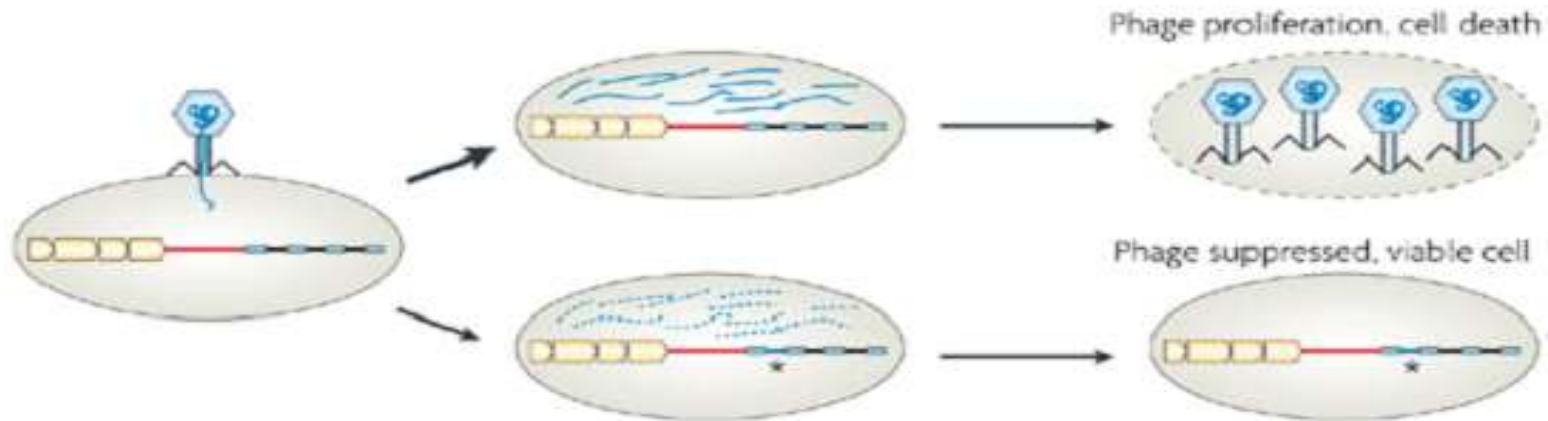
# Le proteine CAS (CRISPR Associated Proteins)



I CRISPR sono caratterizzati dalla presenza di una serie di proteine CAS che variano in numero e disposizione da locus a locus.

L'analisi comparativa di diversi loci ha permesso di identificare diversi tipi di proteine CAS che costituiscono il core del sistema anche se spesso non sono tutte presenti.

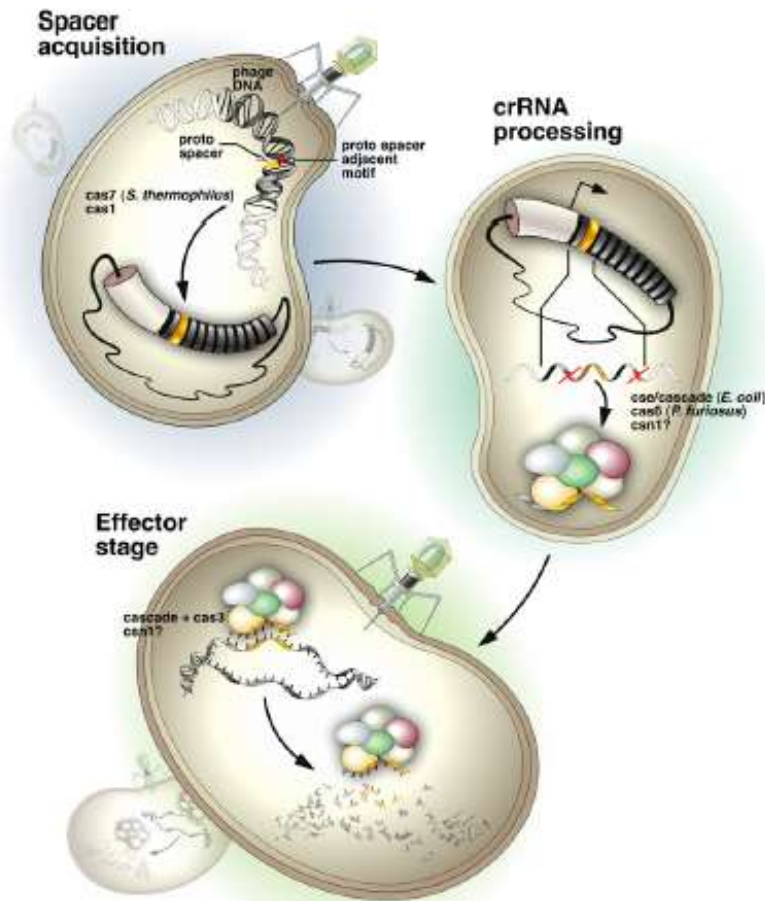
Come fanno le regioni CRISPR ad acquisire le regioni di DNA (SPACERS) che si intervallano tra le sequenze ripetute (REPEATS )?



In seguito all'attacco da parte di un fago, gli acidi nucleici del fago si moltiplicano nella cellula e vengono prodotte nuove particelle fagiche secondo un classico ciclo litico portando a morte le cellule sensibili.

Un piccolo numero di batteri riesce ad acquisire gli "spaziatori fagici" (SPACERS) che permetteranno la sopravvivenza della cellula batterica tramite una degradazione CRISPR-mediata del DNA o mRNA fagico

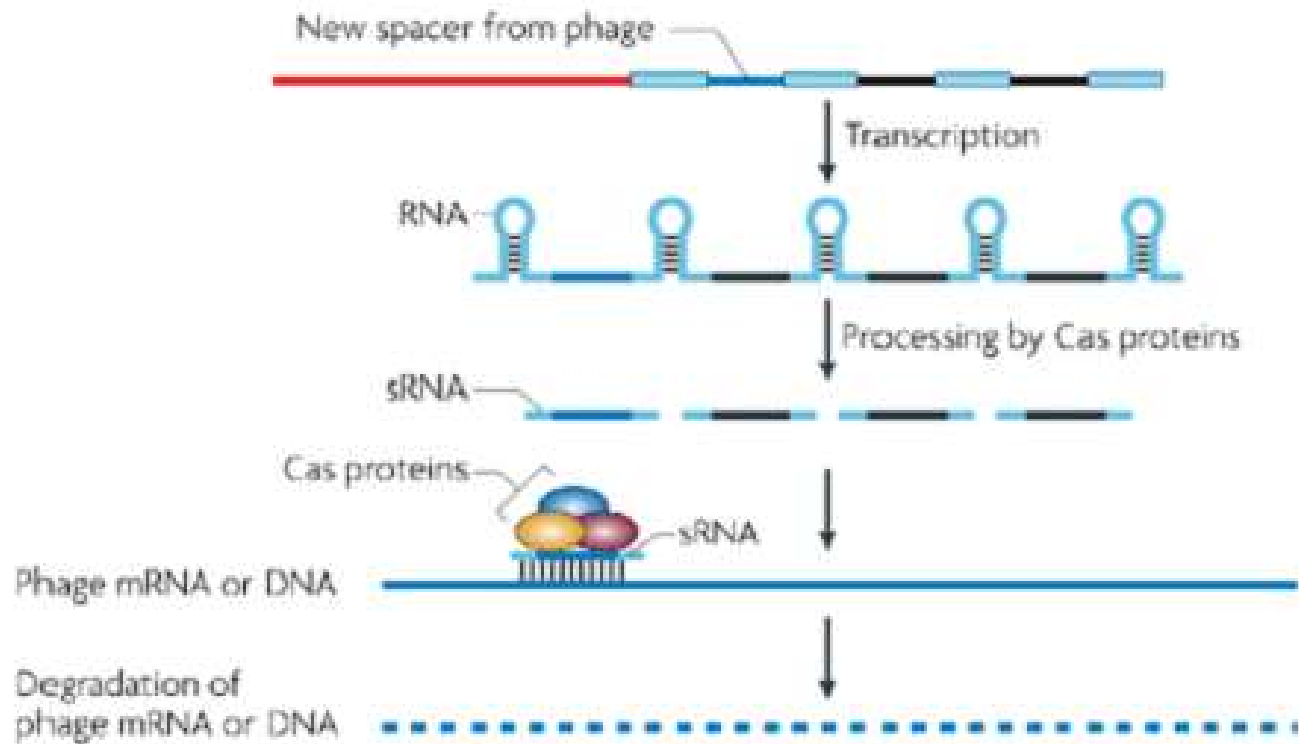
## Le 3 fasi del processo di acquisizione dell'immunità



**Adattamento:** Acquisizione degli SPACERS: durante l'infezione fagica segmenti di acidi nucleici dell'elemento invasore vengono incorporati a valle della sequenza leader del CRISPR.

**Espressione:** Nella fase di processamento il locus è trascritto e processato in crRNA maturo contenente una sequenza REPEAT con un segnale di 8 nt ed un singola sequenza spacer.

**Interferenza:** Durante la fase effettrice i crRNA complessati alle proteine CAS portano alla degradazione il DNA complementare

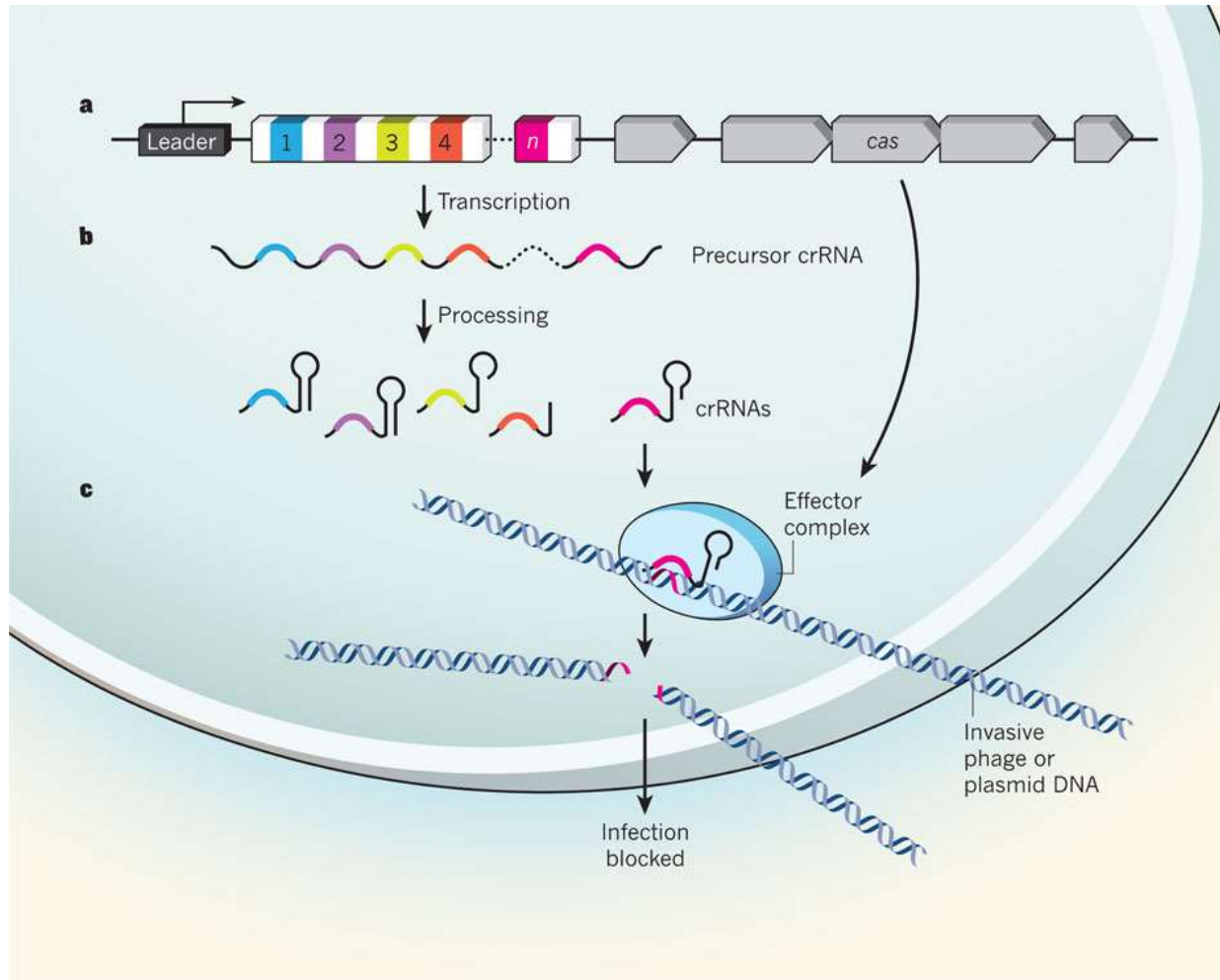


Modello semplificato : la serie di SPACERS-REPEATS è trascritta in un lungo RNA ed i REPEATS assumono una struttura secondaria.

Le proteine Cas riconoscono la struttura secondaria e tagliano l'RNA in modo da produrre dei sRNAs ognuno dei quali contiene una sequenza SPACERS e 2 mezzi REPEATS. Gli sRNA complessandosi alle proteine CAS si appaiano al DNA fagico provocandone la degradazione.

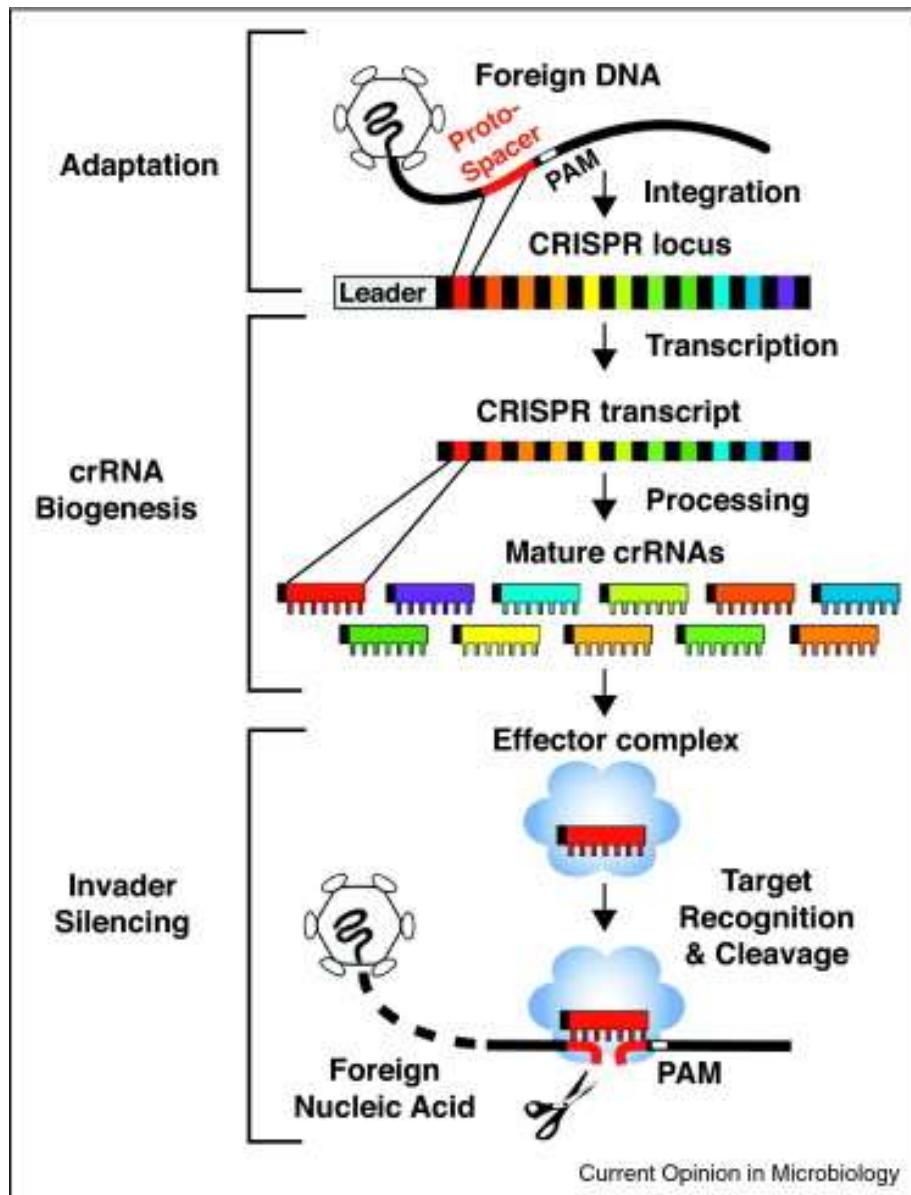
Questo processo è mediato da 1 o più proteine CAS







# Il sistema CRISPR-Cas per la difesa da DNA invasore



Nella fase di adattamento un piccolo frammento di DNA estraneo definito **protospacer** viene acquisito dall'elemento invasore all'interno del locus CRISPR in posizione adiacente al leader.

In seguito a questo processo di acquisizione di sequenze SPACERS dal DNA di elementi invasori, il locus CRISPR risulta costituito da brevi sequenze di sequenze ripetute separate da sequenze diverse (in colore).

Successivamente i trascritti del locus CRISPR sono processati per rilasciare dei piccoli crRNA ognuno in grado di riconoscere una sequenza specifica. I crRNA maturi contengono tipicamente alcuni nucleotidi delle sequenze ripetute che sono il segnale di riconoscimento dei crRNA

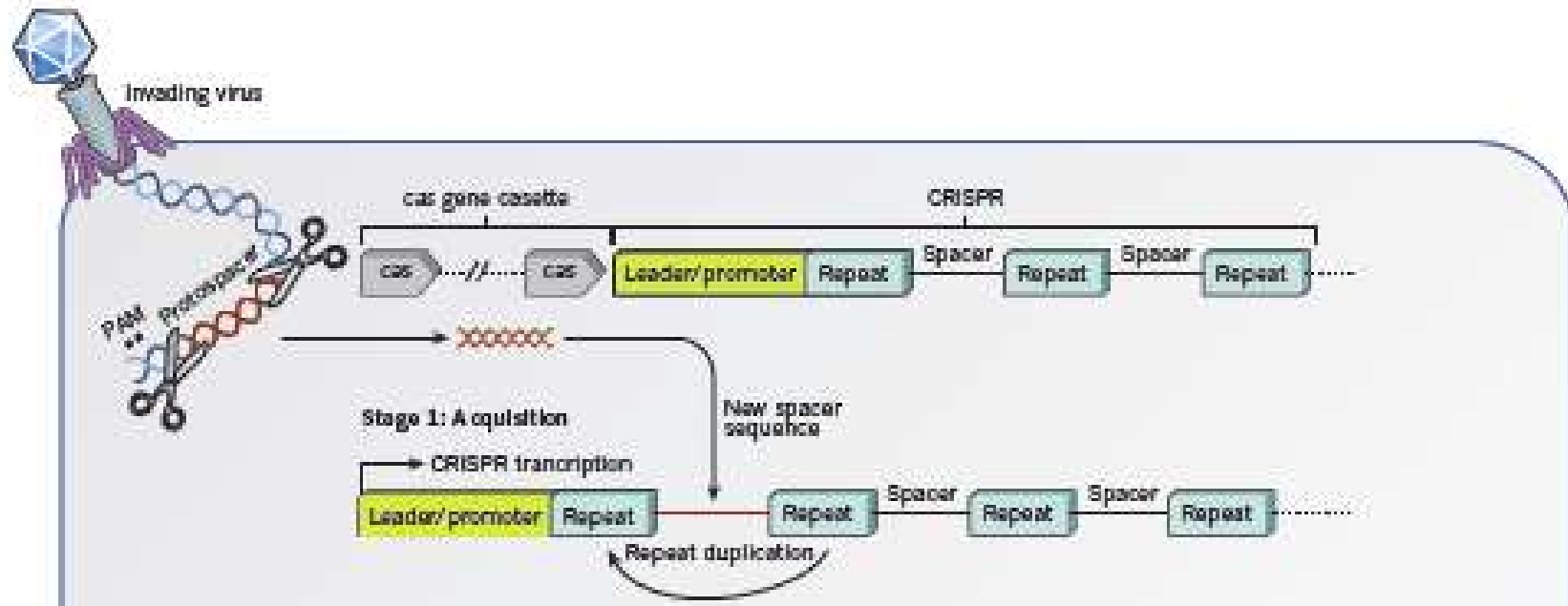
Nella fase di silenziamento il complesso crRNA-CAS riconosce il DNA (oRNA) estraneo tramite appaiamento con i piccoli crRNA.

I sistemi Cmr e Csn colpiscono il taglio dell'RNA target o DNA.

I **PAM Protospacers Adjacent Motif** sono localizzati vicino alle sequenze del DNA "invasore" selezionato per l'integrazione nei CRISPR forniscono ulteriori segnali per il riconoscimento delle sequenze "invasori"

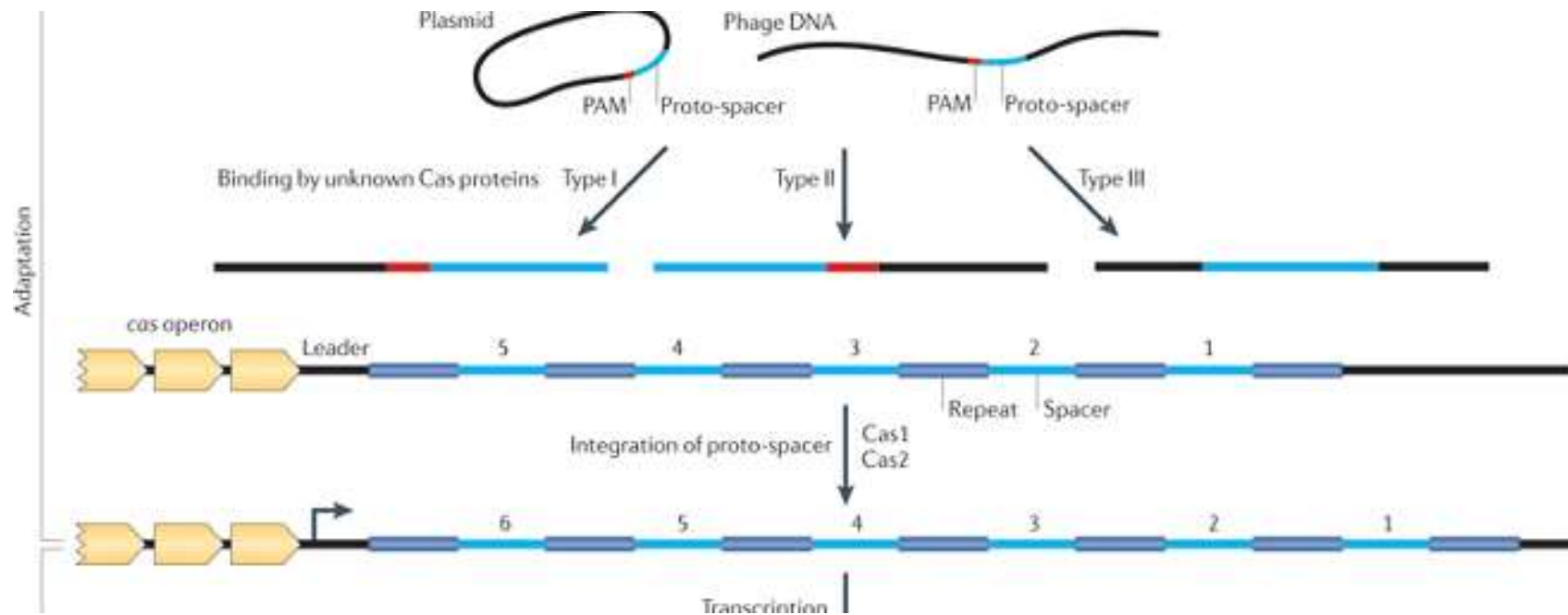
# Diversità nei meccanismi CRISPR.

1. I processi iniziali sono in comune



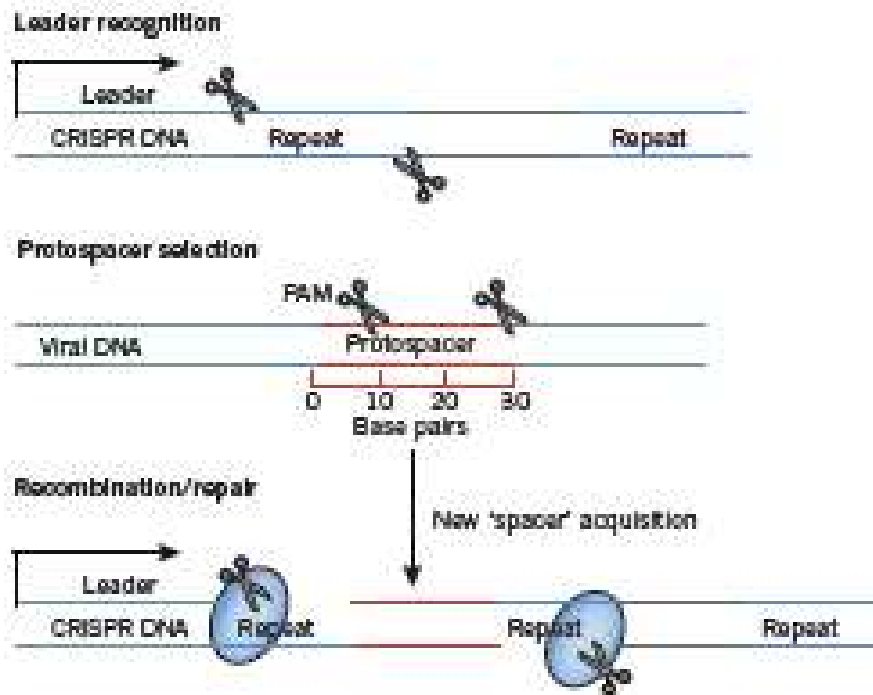
La leader sequence potrebbe contenere un elemento di riconoscimento in grado di reclutare l'apparato di integrazione. Si può anche ipotizzare che l'integrazione avvenga tramite una sola elica di DNA divenuta disponibile in seguito al processo di trascrizione. Il meccanismo di trascrizione/integrazione potrebbe associare la fase di integrazione a quella di trascrizione e garantire che gli spacers dei fagi più recenti siano quelli trascritti per primi

# Fase 1: Adattamento



Nei CRISPR di tipo I e di tipo II il riconoscimento dei PROTOSPACERS negli acidi nucleici "invasori" dipende proprio dai motivi PAM ma come avvenga questo tipo di riconoscimento non è ancora stato chiarito.

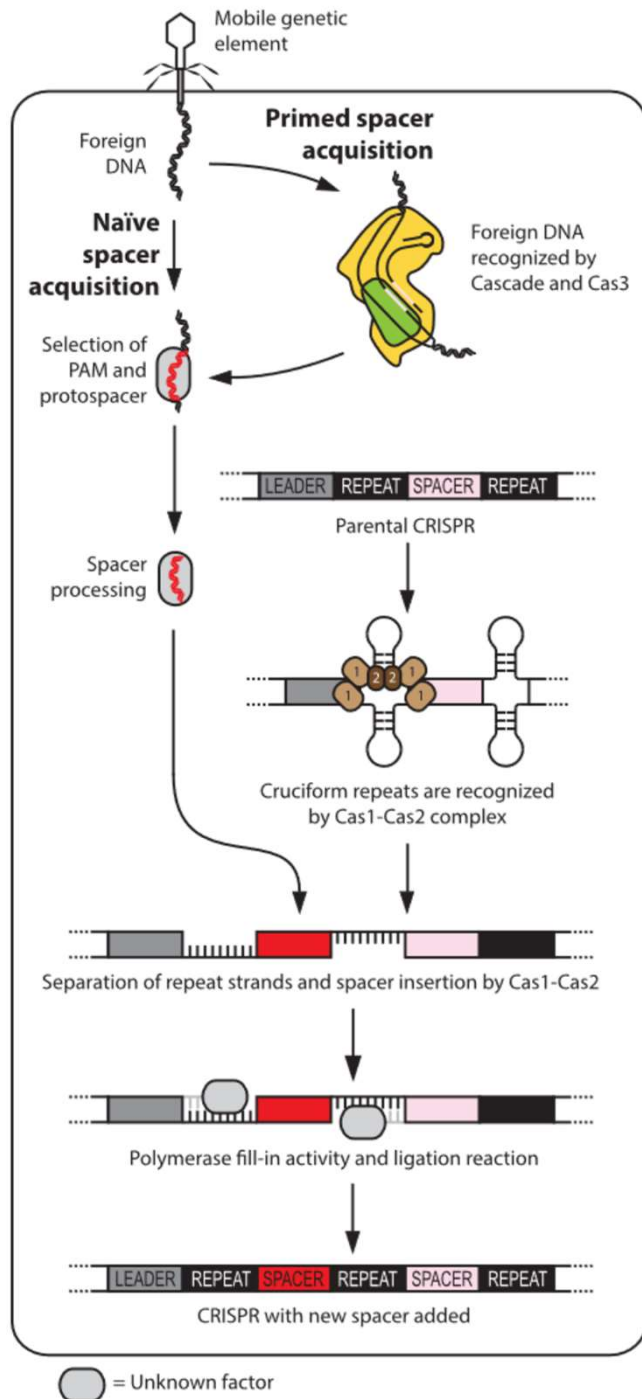
Dopo il riconoscimento iniziale il protospacers viene incorporato grazie alle proteine Cas1 e Cas2 per formare gli spacers.



La fase di adattamento avviene tramite integrazione di frammenti di acido nucleico estraneo vicino alla sequenza Leader. I Protospacers non sono scelti a caso ma vengono selezionati tra le regioni che fiancheggiano le sequenze PAM (Protospacers Adjacent Motifs).

In seguito ad un taglio coordinato del DNA esogeno avviene l'inserimento della sequenza di PROTOSPACER nella locus CRISPR attraverso un meccanismo che duplica la REPEAT SEQUENCE mantenendo l'architettura del locus (Repeat-Spacer-Repeat ).

Il meccanismo non è stato ancora bene definito ma coinvolge la proteina CAS1 e altri fattori implicati nei processi di riparo e di ricombinazione generale



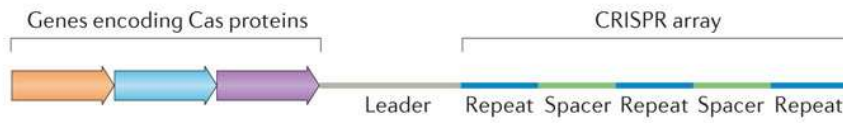
La fase di adattamento conferisce la **MEMORIA GENETICA** ed è un prerequisite per le fasi successive di espressione e di interferenza che neutralizzano gli acidi nucleici re-invasori. La capacità di inserire nuovi spacers è stata riprodotta e testata sperimentalmente in vari sistemi CRISPR - CAS.

Esistono due tipi di meccanismi di acquisizione degli spacer: **NAÏVE** quando un invasore incontra per la prima volta la cellula batterica e **primed** (già istruito) quando già esiste un precedente ricordo del **INVADER** nelle sequenze CRISPR

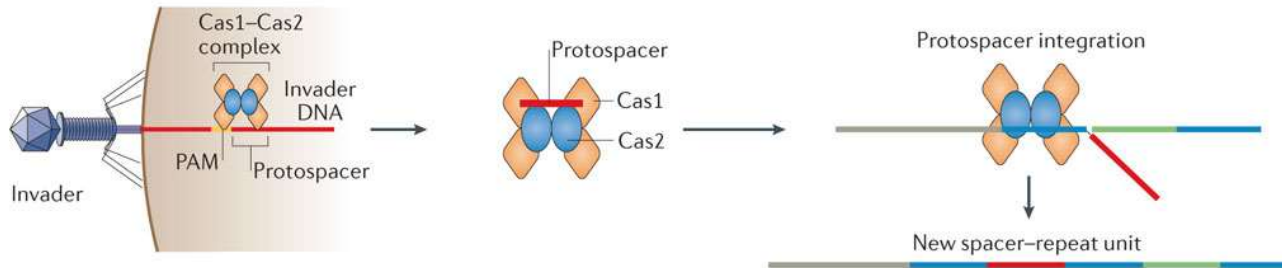
**Fig. 2.** Model of the adaptation in the Type I-E system. There are two types of spacer acquisition, naïve and primed. Both require the presence of a PAM and are dependent on the Cas1–Cas2 complex. The Cas1–Cas2 complex recognizes the CRISPR and likely prepares it for spacer integration. Naïve spacer acquisition occurs when there is no previous information about the target in the CRISPR. Primed spacer acquisition requires a spacer in the CRISPR locus that matches the target DNA and the presence of Cas3 and the Cascade complex. Primed acquisition results in insertion of more spacers from same mobile genetic element. PAM = Protospacer Adjacent Motif.

# Fase di Adattamento

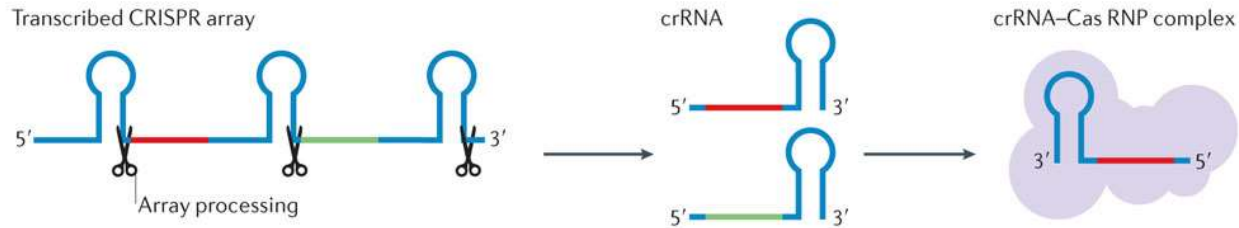
## a Locus organization



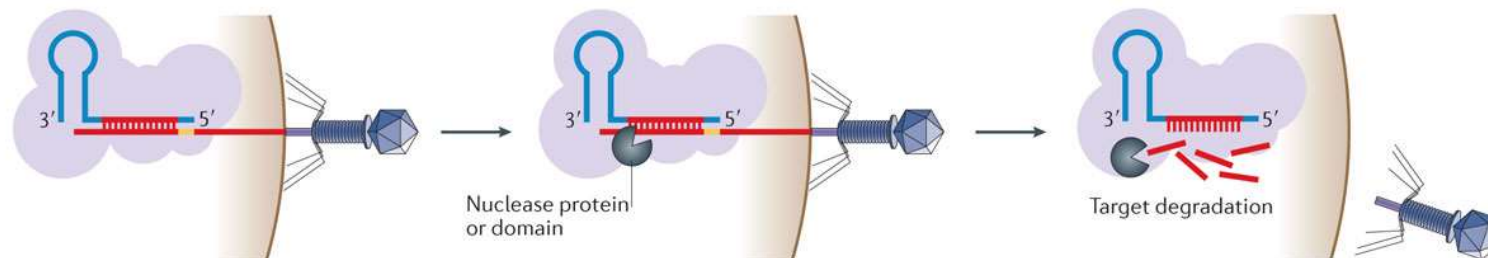
## b Adaptation



## c Expression and maturation



## d Interference





La fase di adattamento CRISPR-Cas è un processo complesso nel quale un protospacer deve essere recuperato dal DNA "invasore" e inserito nella sequenza CRISPR.

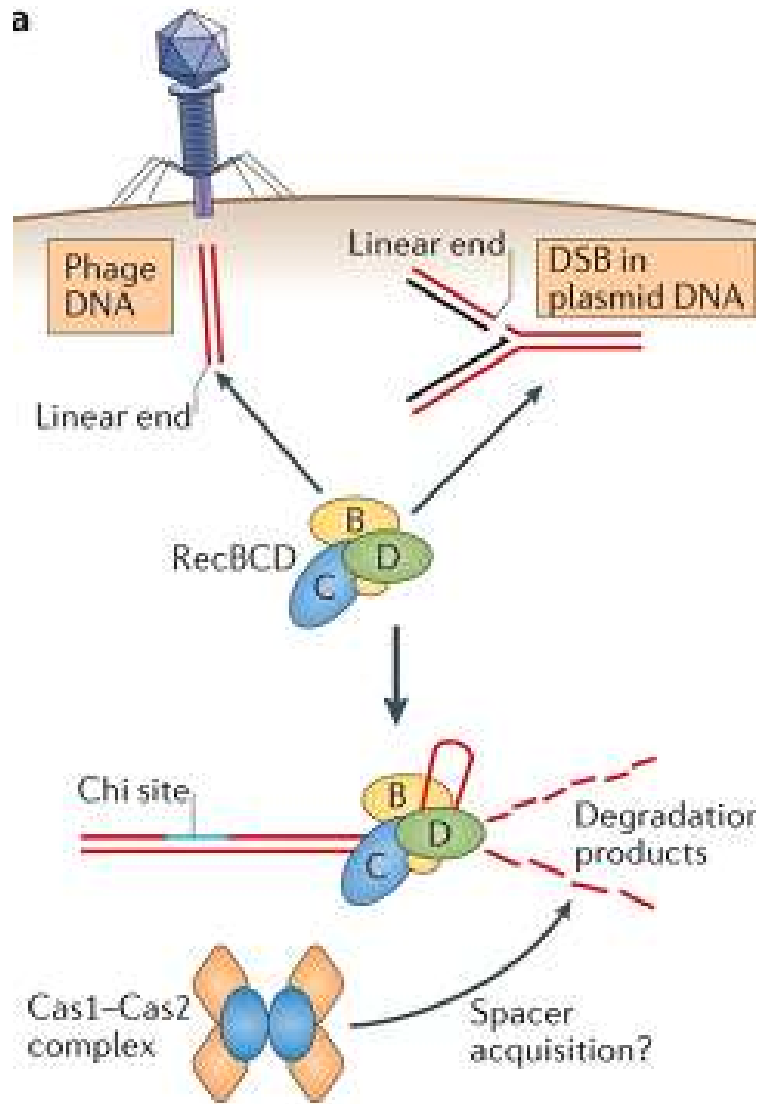
Il processo si divide in più tappe

1. Riconoscimento
2. Taglio
3. Inserimento.

I componenti di questo processo sono le proteine Cas1 (endonucleasi) e Cas2 (attività di taglio RNA e DNA), la sequenza leader e il primo repeat CRISPR.

Le funzioni endonucleasiche di Cas2 non sono direttamente richieste mentre è importante che formi complesso con Cas1

## Coinvolgimento del complesso RecBCD



Da un'analisi genomica delle sequenze spacers risulta che i protospacers sono recuperati dal complesso Cas1/Cas2 dal processo di degradazione del DNA esogeno portato avanti dal complesso RecBCD.

La distinzione tra self e non self quindi avverrebbe grazie alla presenza di siti Chi sul genoma (ogni 5KB nel genoma di *E.coli*).

Nel caso di DNA plasmidico i DSB si genererebbero durante il processo di replicazione.

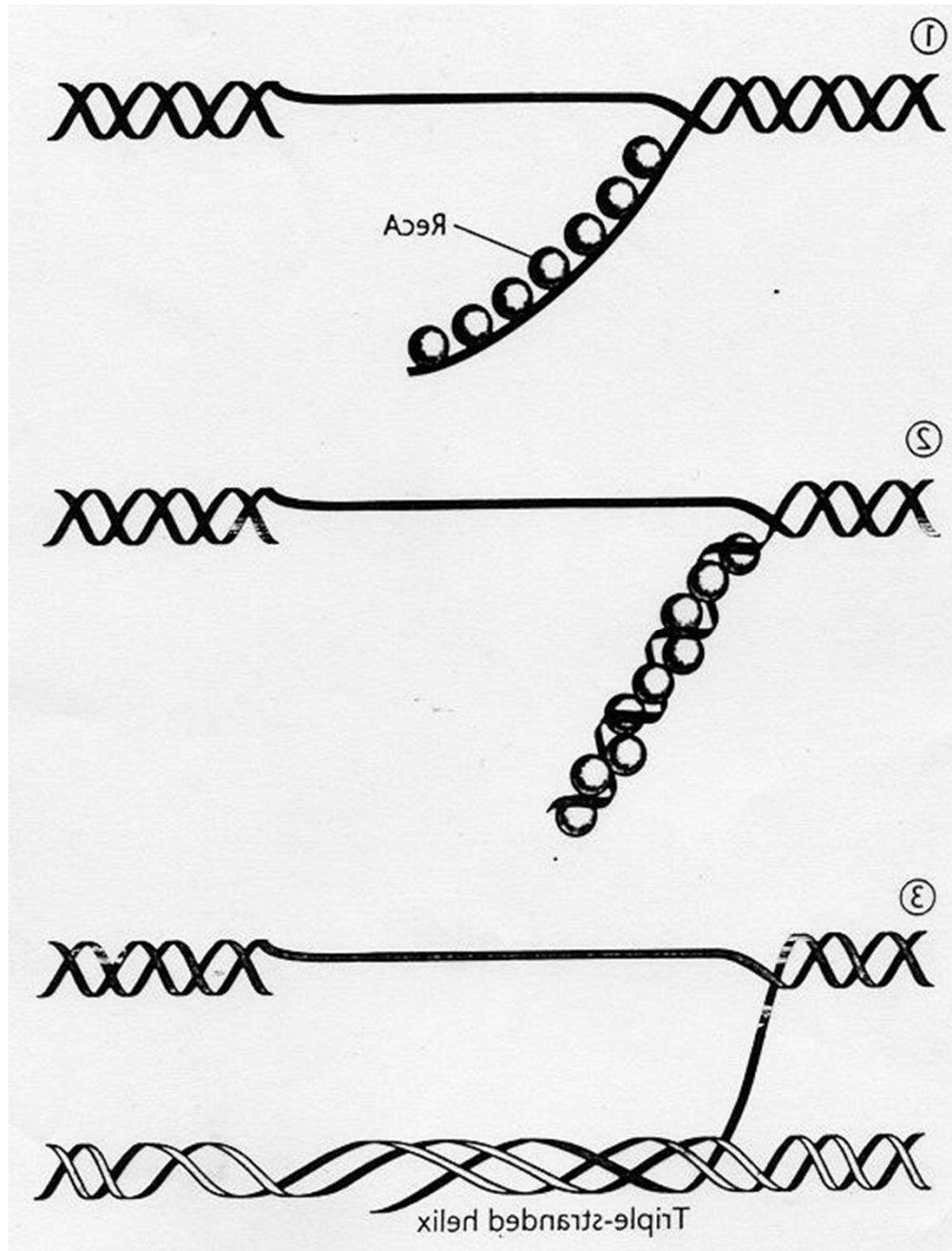
**b**

**Self DNA**



**Foreign DNA**





RecA si lega al DNA a singolo filamento formando un complesso RecA-ssDNA che può invadere una molecola di DNA a doppia elica formando il complesso

ssDNA-RecA-dsDNA

Per questo è necessario il complesso costituito dalle proteine

RecB - RecC - RecD

Il Complesso RecBCD è un complesso multifunzionale

- dotato di attività nucleasica su DNA ss e ds
- dotato di attività elicasica
- dotato di attività ATPasica (idrolizza ATP)

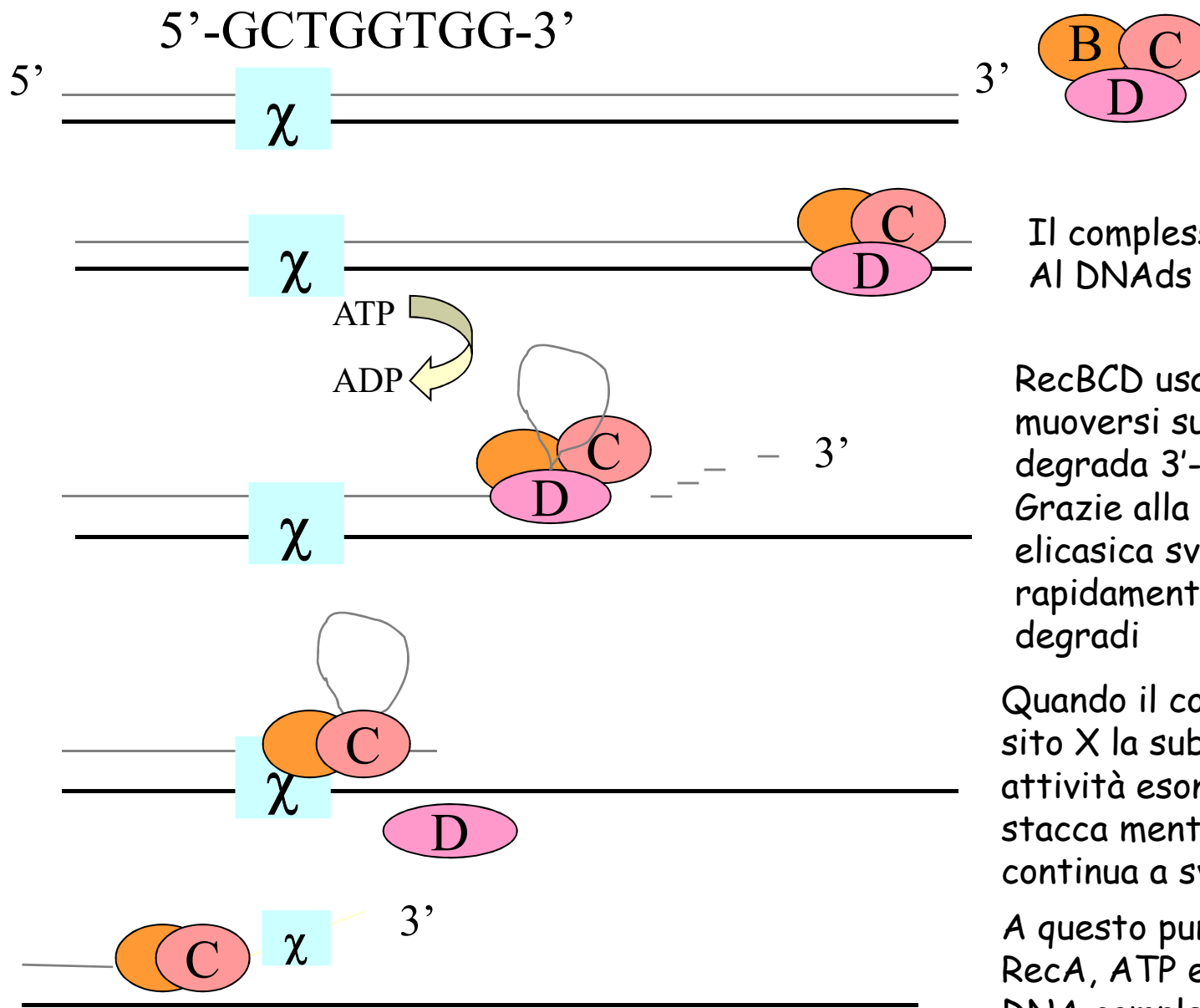
## Le sequenze Chi( $\chi$ )

*E.coli* contiene un numero molto elevato di sequenze chi( $\chi$ )

Sequenza specifica di basi ( 5' GCTGGTGG 3')

Nel cromosoma di *E.coli* è presente

una sequenza Chi circa ogni 5 kb



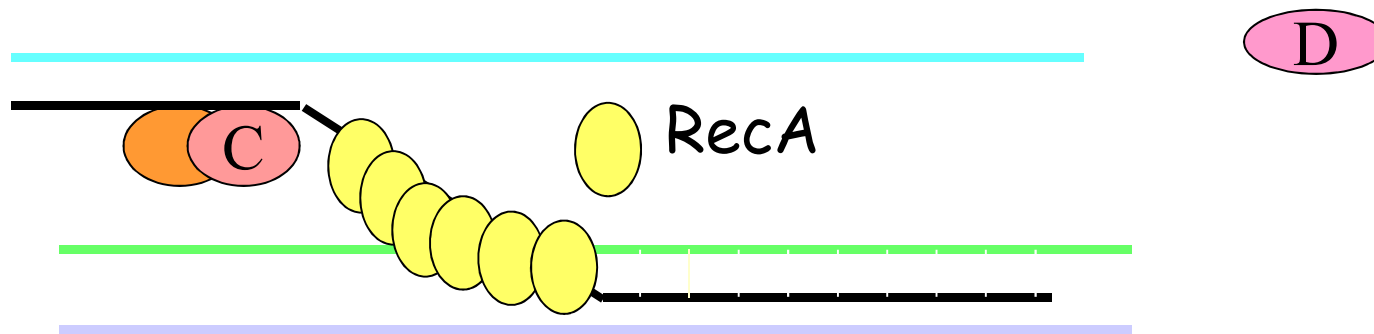
Il complesso RecBCD si lega Al DNAs

RecBCD usa ATP per muoversi sul DNAs e degrada 3'-5' (ExoV) Grazie alla sua attività elicasi svolge il DNA più rapidamente di quanto lo degradi

Quando il complesso arriva al sito X la subunità D dotata di attività esonucleasica si stacca mentre l'elicasi continua a svolgere il DNA

A questo punto se è presente RecA, ATP e una molecola di DNA complementare può avvenire ricombinazione





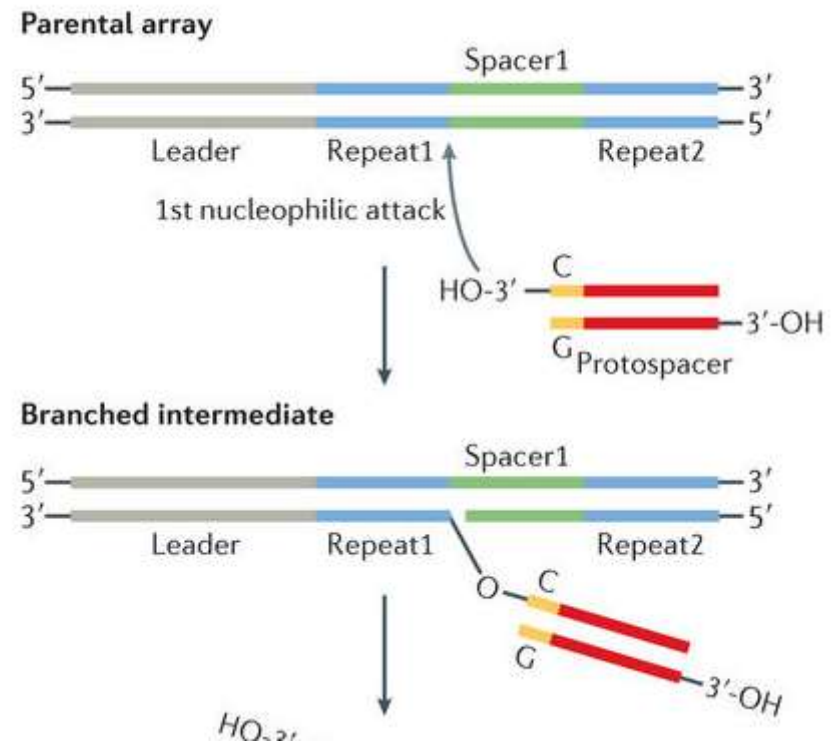
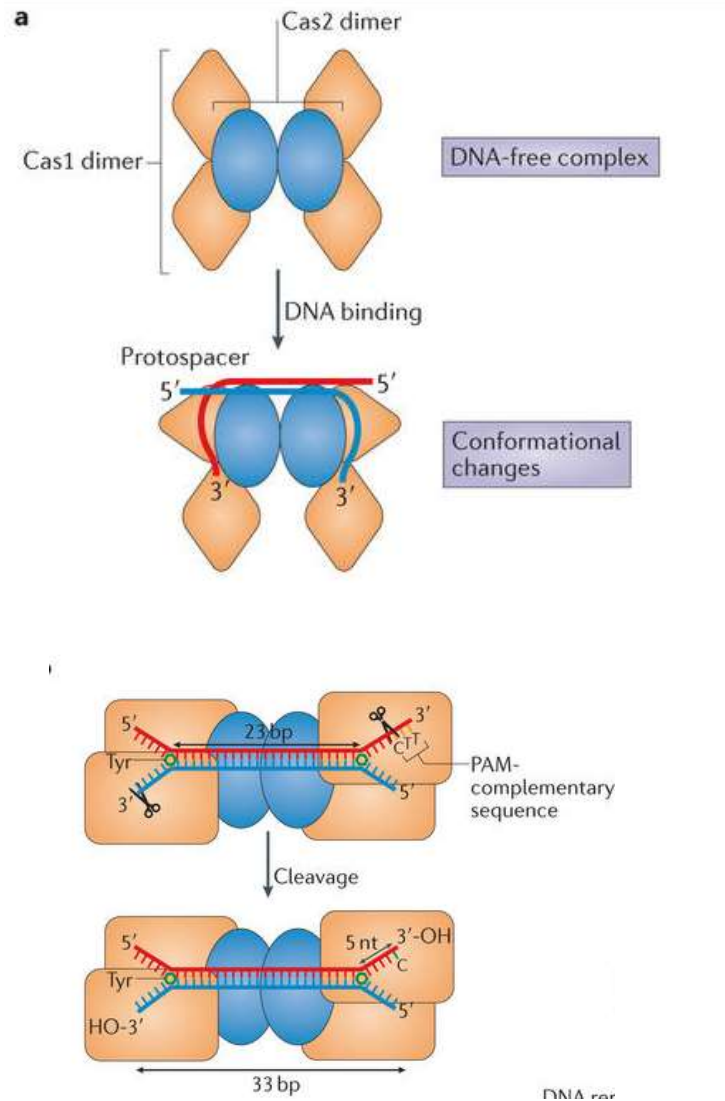
Grazie a questo processamento mediato dal complesso RecBCD si può creare una regione di DNA a singolo filamento.

RecA sarà in grado di riconoscere e legarsi a questo filamento di ssDNA

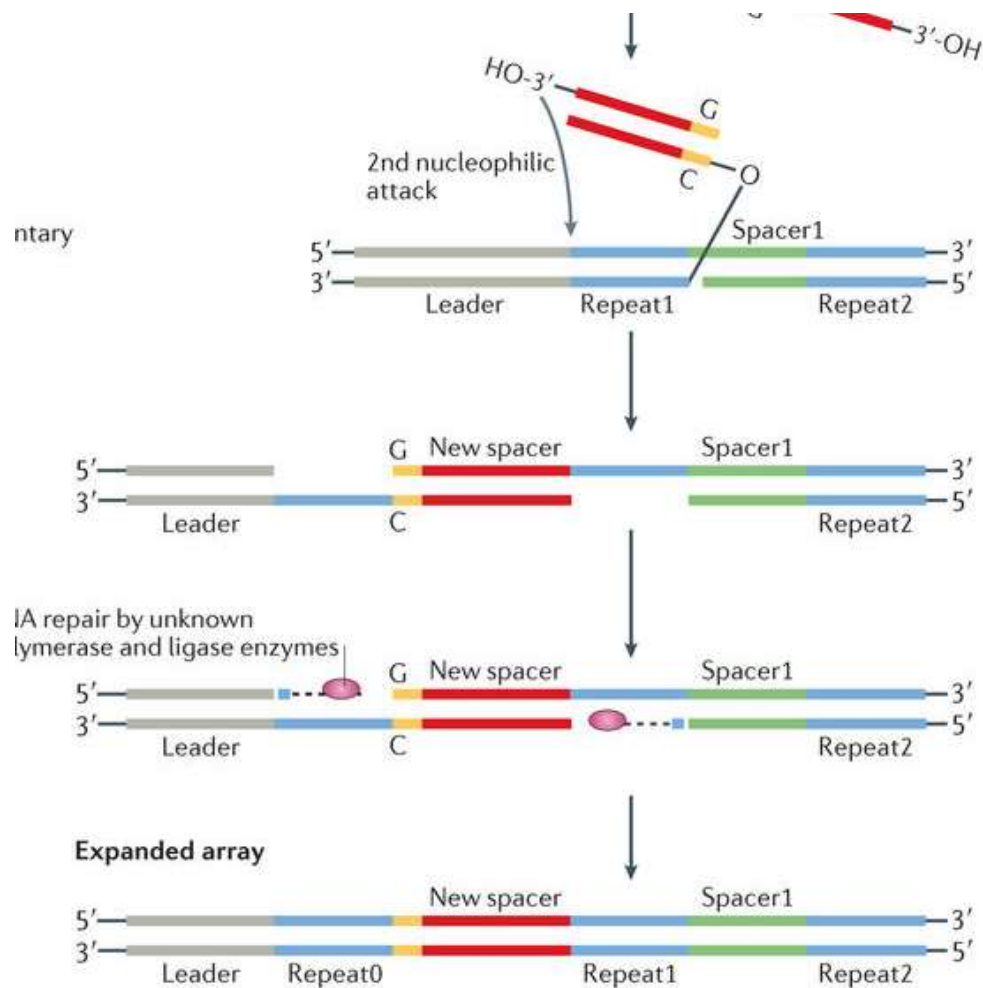
Si potrà avere il processo di ricombinazione per invasione di una regione omologa di dsDNA da parte del complesso RecA-ssDNA

Oltre al sistema RecBCD svolge un ruolo importante nella selezione del DNA fagico da inserire :

- i sistemi di restrizione e modificazione dell'ospite che possono degradare il DNA fagico
- la presenza nella popolazione di particelle fagiche difettive che non sono in grado di dare il via immediatamente ad un ciclo litico e vengono quindi frammentate

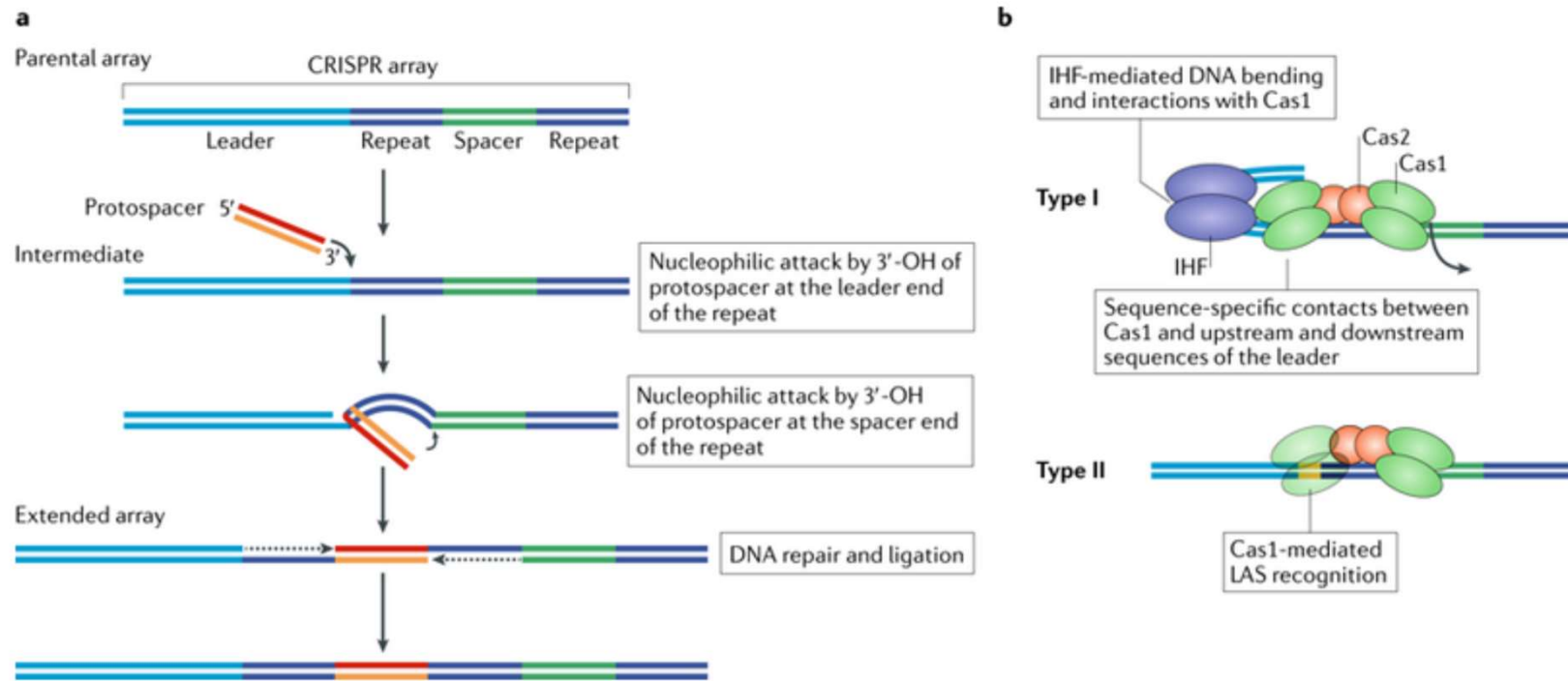


Il complesso Cas1/2 costituito da 2 dimeri di Cas1 e un dimero di Cas2 si comporterebbe come un integrasi



A model for protospacer integration into the CRISPR array.

The protospacer 3'-OH group carries out a nucleophilic attack on the 5' end of the first repeat, thus initiating spacer acquisition by forming a branched intermediate in which a single strand of the protospacer is ligated to a single strand of the CRISPR array. The 3'-OH group on the other protospacer strand generates a second nucleophilic attack on the 5' end of the opposing DNA strand of the repeat, which is juxtaposed to the leader sequence. The product of this reaction is an expanded CRISPR array with a new spacer and a duplicated repeat. The ssDNA gaps that are produced at the repeat sequences are filled and repaired by uncharacterized enzymes.



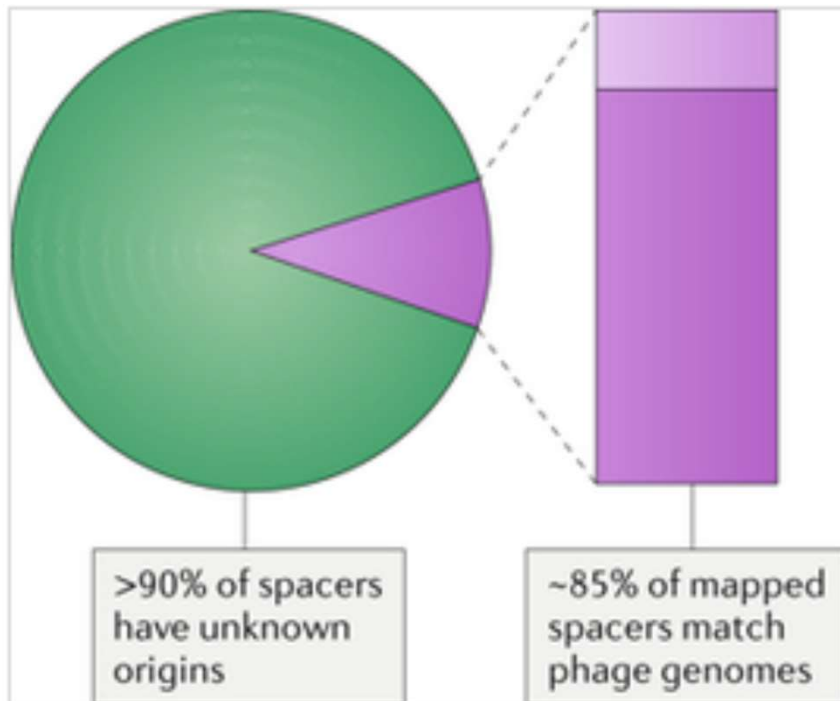
The array is preceded by an AT-rich leader sequence. Integration of new spacers begins with a concerted cleavage–ligation reaction that occurs preferentially at the leader end of the first repeat, whereby the terminal 3'-OH of the protospacer carries out a nucleophilic attack. Next, the repeat DNA is bent, and a second cleavage–ligation reaction takes place at the spacer side of the repeat. The product of this reaction is an intermediate in which the 3' ends of a double-stranded (dsDNA) protospacer are ligated to single-stranded DNA (ssDNA) repeat sequences. The ssDNA repeats are presumably filled by DNA polymerase and ligated to complete the spacer integration process. **b** | Two mechanisms for preferential spacer acquisition at the leader end of the CRISPR array. In the type I system, integration host factor (IHF) binds to a conserved binding site in the leader and induces DNA bending, which enables the Cas1–Cas2 complex to perform the first cleavage–ligation reaction. Cas1–Cas2 makes specific contacts with upstream and downstream sequences in the leader, as well as with IHF.

Le regioni CRISPR evolvono rapidamente e sono spesso ipervariabili anche tra ceppi correlati.

I nuovi spacers sono inseriti sempre all'estremità 5' vicini alla sequenza leader e costituiscono regioni più variabili rispetto ai "vecchi" spacers localizzati al 3'. Si osservano frequentemente delezioni di unità spacer-repeat necessarie per prevenire un ingrandimento eccessivo del locus : non è chiaro se avvengano per ricombinazione omologa o per un processo attivo.

Generalmente i CRISPR loci sono diversi tra le diverse specie batteriche anche se si possono trovare dei repeats simili tra specie diverse ad indicare eventi possibile HGT mediato da megaplasmidi che possono contenere loci CRISPR

## Il mondo oscuro delle sequenze spacers dei CRISPR



Origine degli spacers:  
oltre 90% degli spacers è di  
origine sconosciuta ....

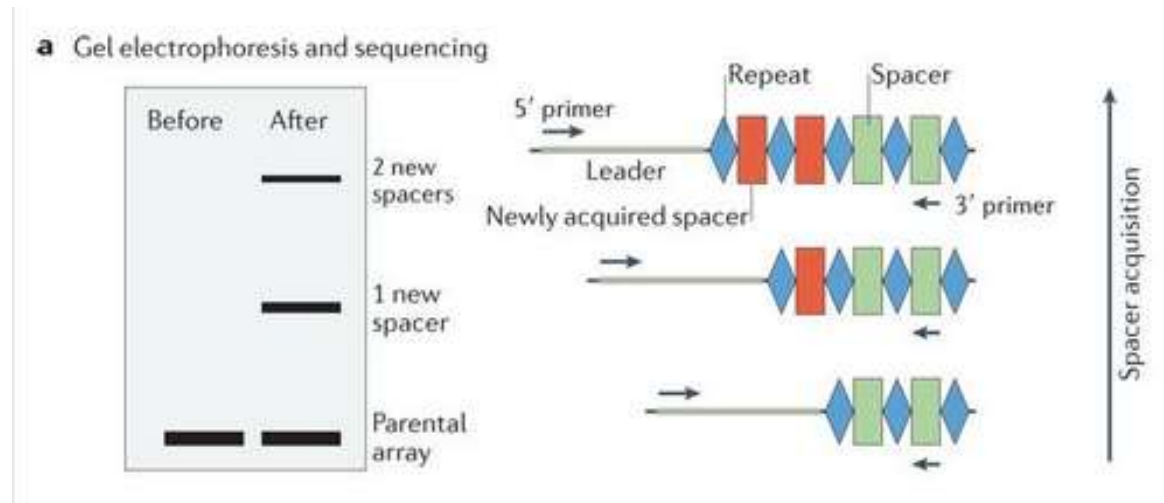
Tra quel 10 % conosciuto  
sono per 85% di origine  
fagica.

Da dove vengono allora?

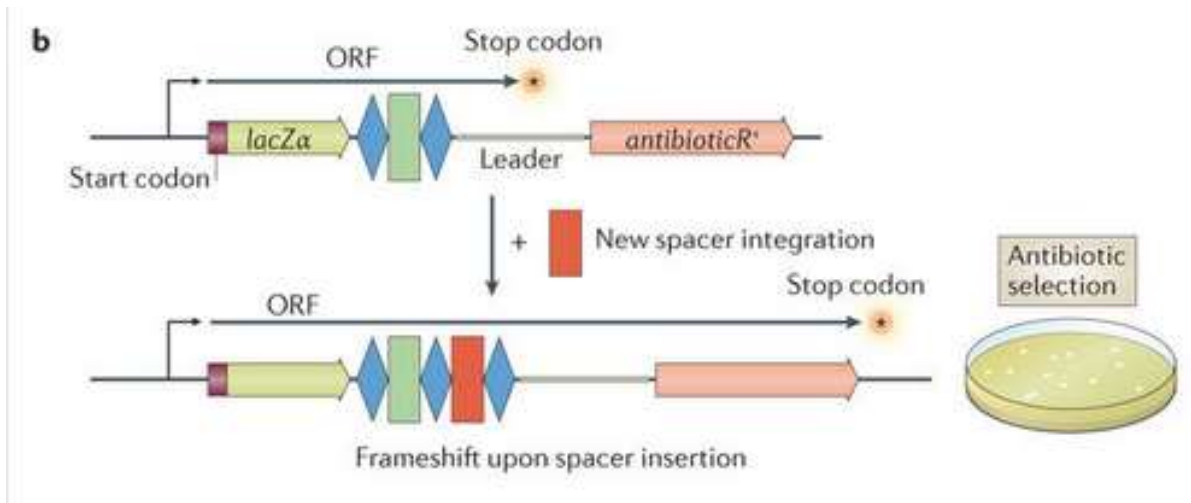
L'ipotesi più semplice è che  
si tratti di fagi o MGE  
ancora sconosciuti



# Come si fa a capire sperimentalmente se vengono inseriti nuovi spacers

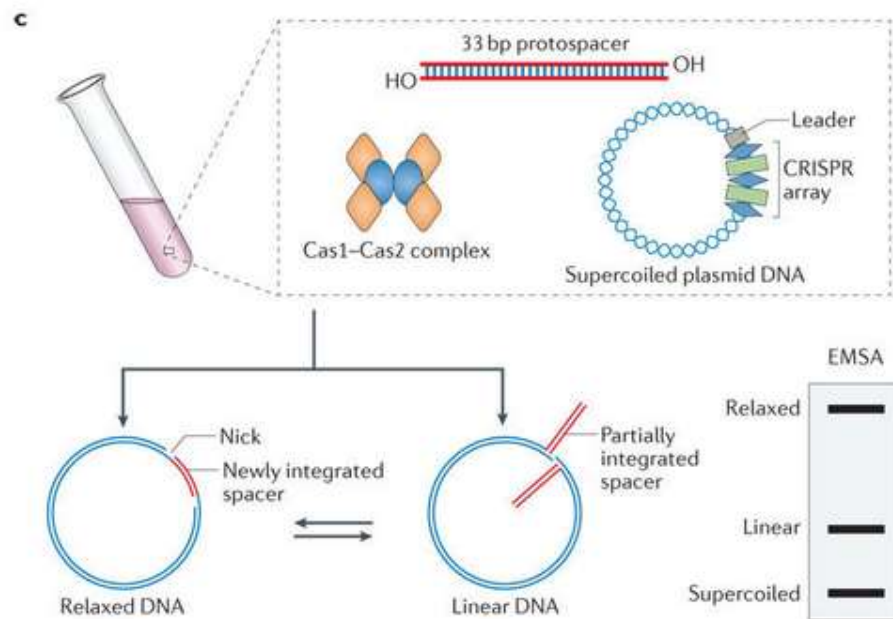


The first approach takes advantage of the fact that new spacers are preferentially added juxtaposed to a regulatory 'leader' sequence that is found directly upstream of the CRISPR array. Each adaptation event entails the expansion of the array by one repeat and one spacer, together sized about 60 nucleotides; hence, PCR amplification with primers that anneal to the leader and a parental spacer can provide data on new spacer integration, either through the use of gel electrophoresis or by direct sequencing<sup>1</sup>. This technique is compatible with various genetic backgrounds that include native or partial repertoires of *cas* genes, as well as manipulated leader and repeat sequences. The biochemical intermediates of spacer integration can also be examined using this approach, by Southern blotting of the PCR-amplified CRISPR array.



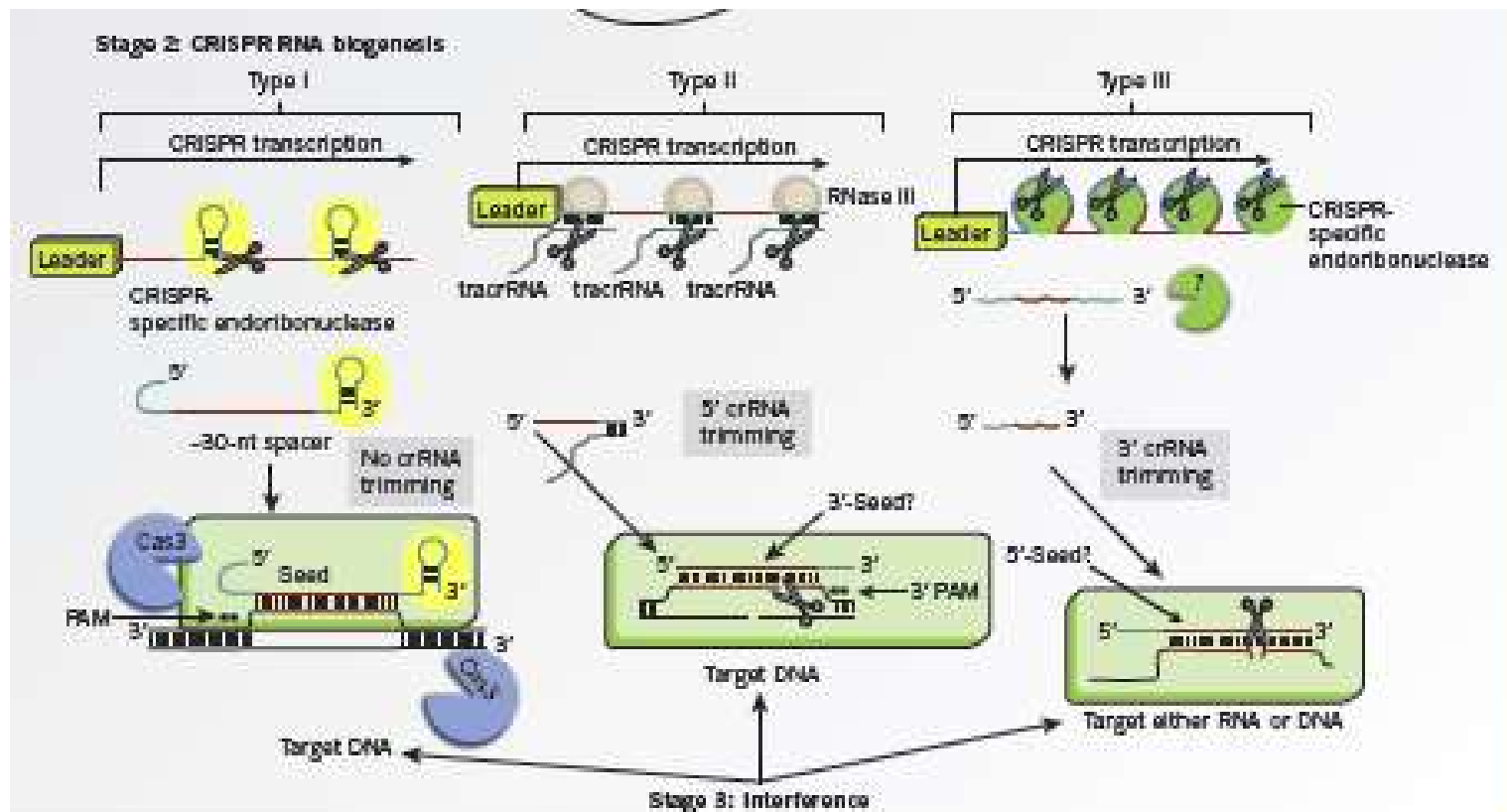
Creazione di una fusione con un codone di stop.

The second approach relies on a viability-based assay for clones that have incorporated new spacers into a CRISPR array that is genetically fused to an out-of-frame antibiotic-resistance gene<sup>26</sup> (*antibioticR<sup>+</sup>*; see the figure, part **b**). When a new spacer is integrated into the CRISPR array, the coding frame is repaired and the antibiotic-resistance gene is expressed, so that the survival of a clone following exposure to antibiotics denotes successful spacer integration. This system therefore provides antibiotic-based selection for spacer acquisition events and enables high-sensitivity detection of newly incorporated spacers without the need for the overexpression of Cas proteins.

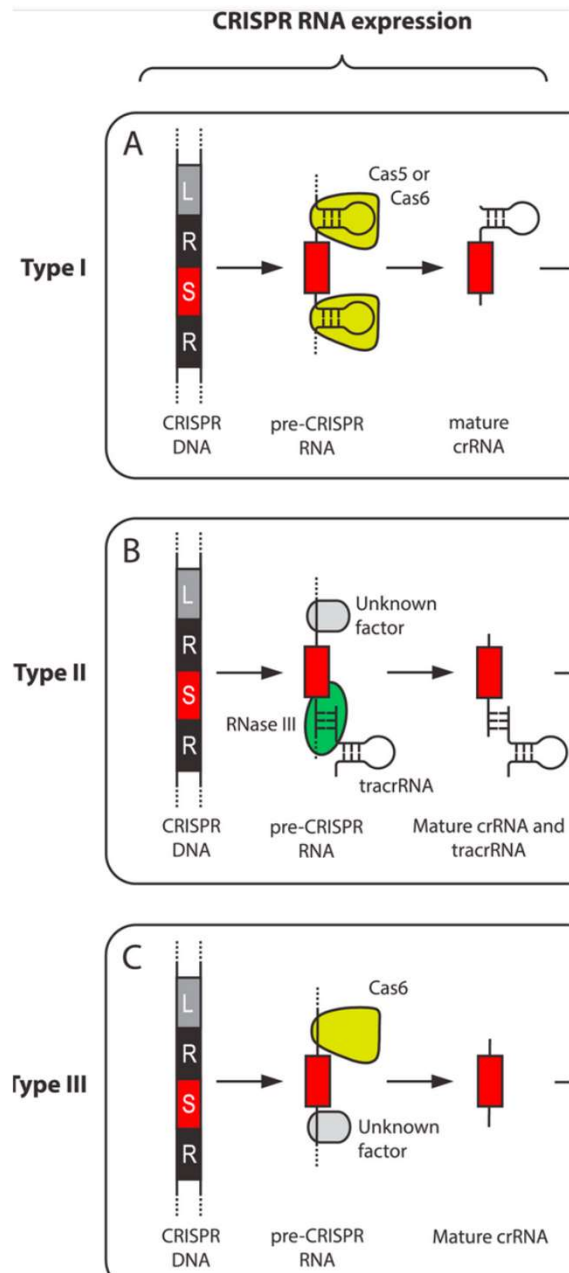


The third approach is to assay spacer acquisition *in vitro*. In this approach, isotopically labelled protospacers are mixed *in vitro* with purified Cas1 and Cas2, which form a protein complex (Cas1–Cas2) that is essential for spacer acquisition, and a CRISPR array-containing plasmid that is an acceptor for spacer integration<sup>29</sup>. The integration of spacers into the CRISPR array results in a relaxation of supercoiled plasmid DNA or the production of linear DNA, both of which can be observed through an electrophoretic mobility shift assay (EMSA) as mobility shifts on the gel (see the figure, part c). This assay demonstrated that the chemical steps of spacer integration are reminiscent of the activity of retroviral integrases. Alternatively, *in vitro* assays have been used to study the reverse reaction of spacer integration: disintegration. These assays expose short labelled branched DNA intermediates of spacer integration to the acquisition machinery, thus producing reaction products that can be observed on a gel as a measure of activity. This assay was used to study the sequence specificity of the integration site in the CRISPR array<sup>33</sup>.

## 2. La diversità nella fase delle strategie per la generazione dei crRNA maturi



# Espressione: punti in comune e diversità tra i 3 sistemi



La trascrizione dei loci CRISPR -CAS per generare un complesso RNA- proteina presenta alcune omologie alcuni punti in comune in tutti e tre i tipi.

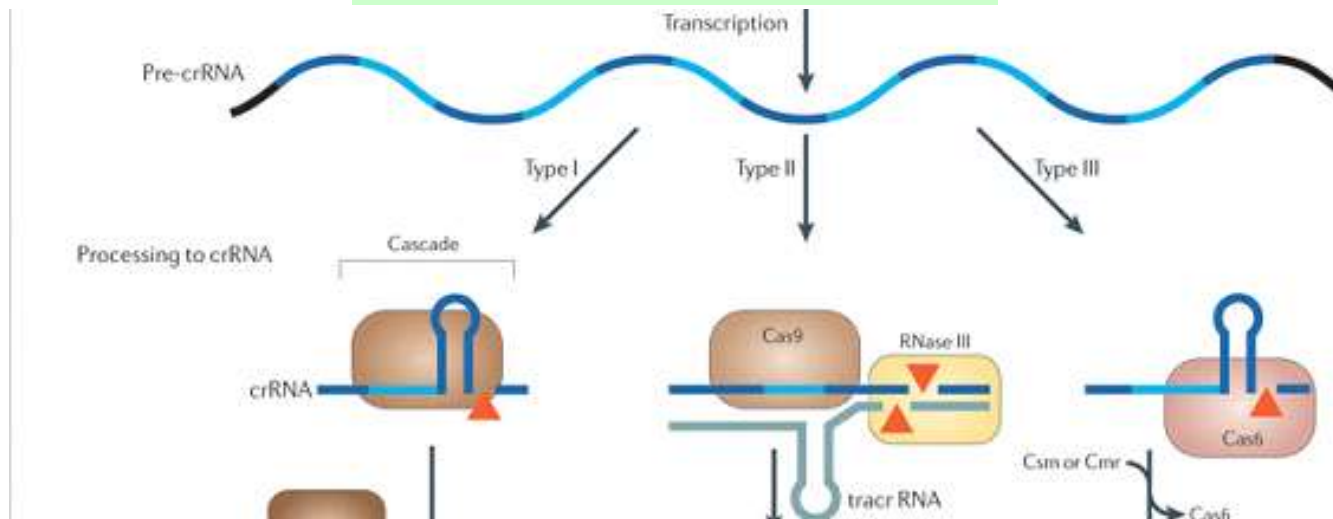
Tutti i sistemi trascrivono il locus CRISPR, processano RNA con le Cas ribonucleasi e formano un complesso CRISPR-ribonucleoproteico crRNP. In alcune specie la trascrizione inizia nella sequenza leader che contiene elementi importanti per la trascrizione.

Come risultato si genera un lungo trascritto precrRNA che contiene una serie di strutture secondarie ad ansa se nel CRISPR sono contenute repeat palindromiche.

Il pre-CR è poi processato in unità più piccole corrispondenti ad un singolo spacer fiancheggiato da repeat parziali.

Nei sistemi di Tipo I e Tipo III ( che ora costituiscono i sistemi CRISPR Cas classe 1) sia lo stadio di maturazione che di espressione che di interferenza sono eseguiti da un complesso multiproteico definito Cascade (**C**RISPR **A**Ssociated **C**omplex for **A**ntiviral **D**Efence) nel Tipo I o Csm complex o il Cmr complex nel Tipo III mentre nei sistemi appartenenti alla classe 2 ( ovvero tipo II, V e VI, questi processi sono portati avanti da un unica polipeptide (Cas) che richiede la transattivazione da parte di uno sRNanc ( tracrRNA)

## Fase 2: Espressione



Durante la fase di espressione il locus CRISPR contenente gli spacers viene espresso dando luogo ad un lungo trascritto definito pre-crRNA (pre-CRISPR RNA).

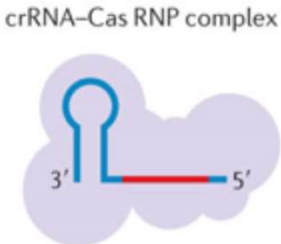
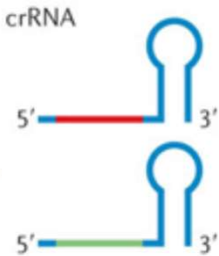
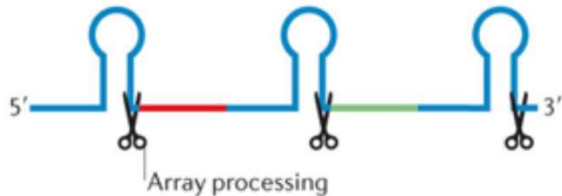
Nel tipo I il complesso di proteine CRISPR associate ( CASCADE) si lega al pre- crRNA che viene poi tagliato dalle subunità Cas6e e Cas6f dando luogo a piccoli crRNA costituiti da una tipica sequenza di 8 nucleotidi all'estremità 5' e dalla restante porzione del repeat che forma poi una struttura ad ansa al 3'

Nel tipo II un *trans encoded RNA* si lega al segmento repeat del pre crRNA e questo appaiamento RNA-RNA viene riconosciuto dalla RNase III dell'ospite in presenza delle proteine Cas.

Nel tipo III vien maturato dal Cas6 e poi trasferito a altri complessi proteici Csm e Cmr

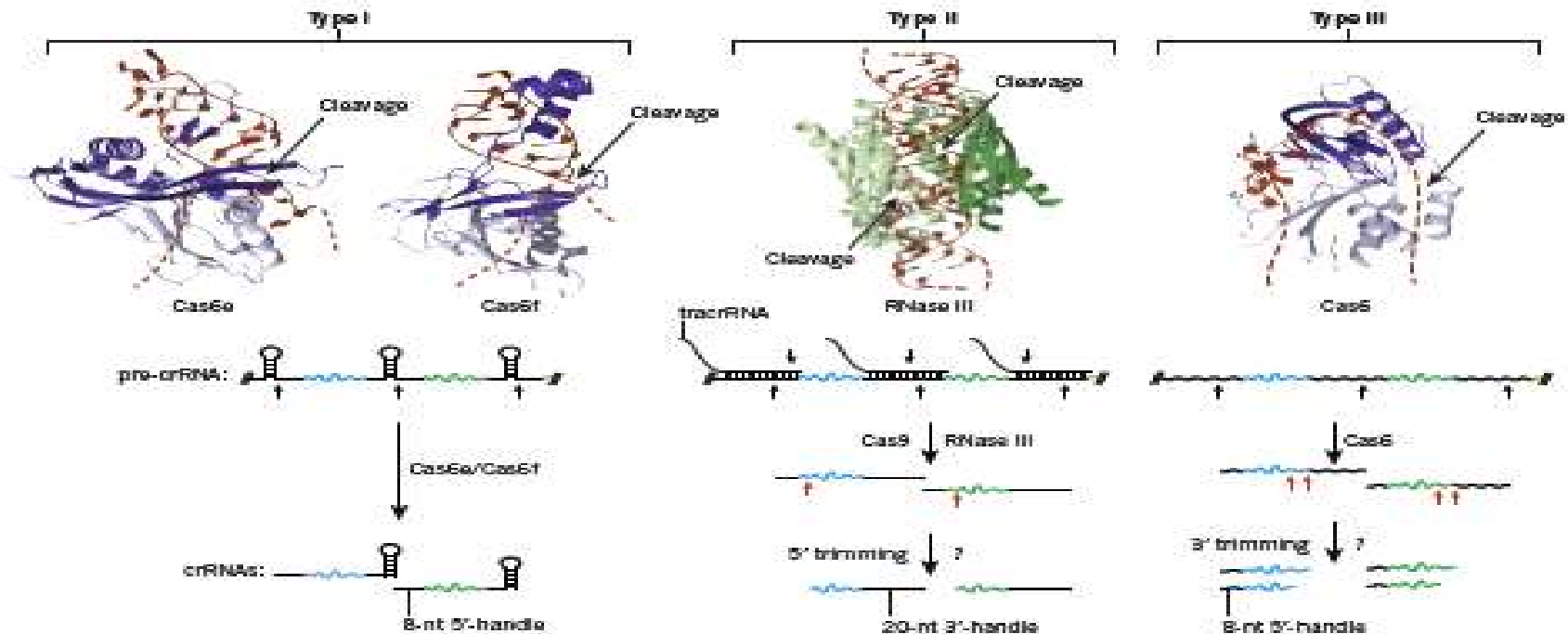
**c Expression and maturation**

Transcribed CRISPR array





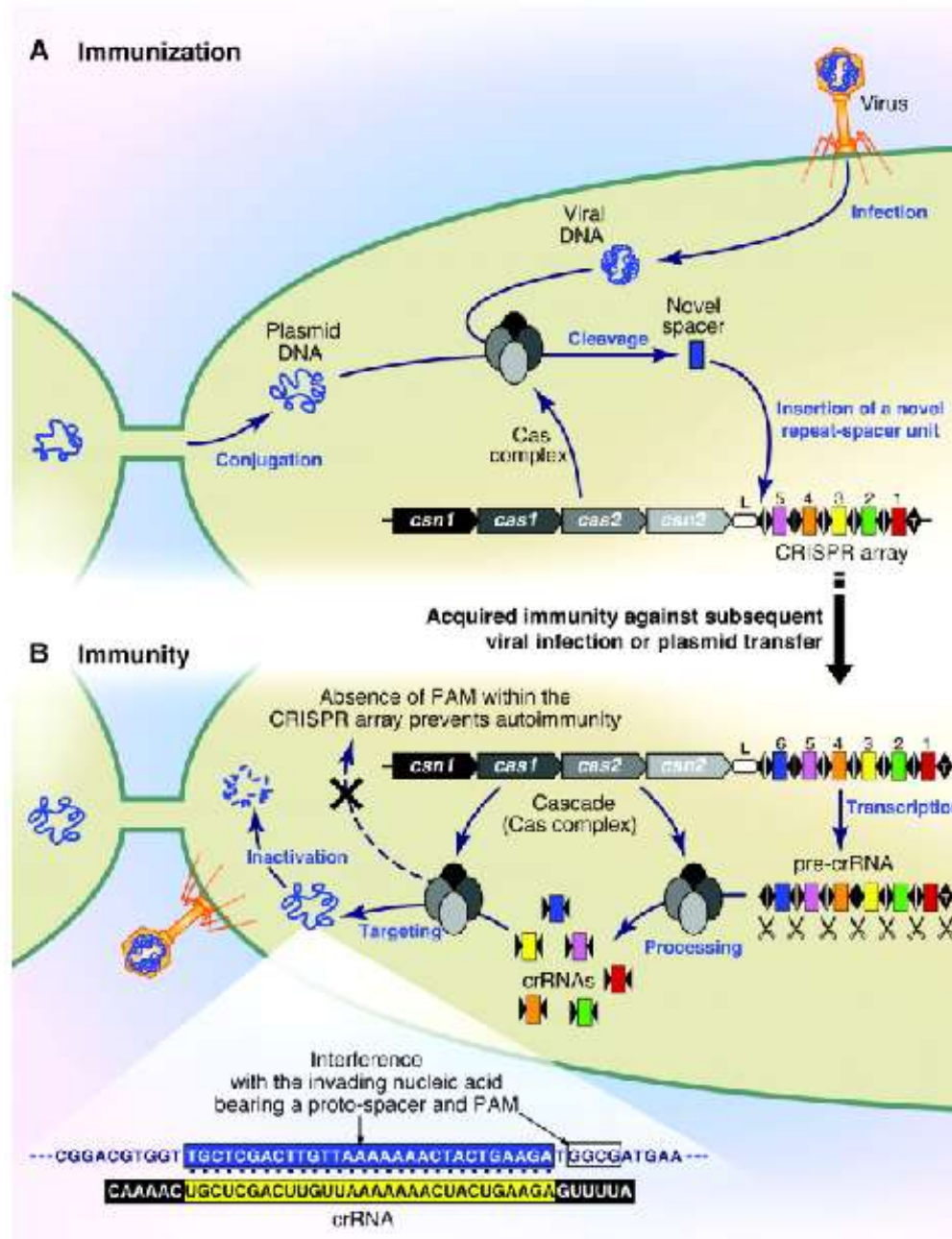
# Il processamento del crRNA.



**Tipo I** Cas6a e Cas6f riconoscono l'ansa maggiore della struttura stem-loop del crRNA tramite interazioni elettrostatiche. Il taglio avviene nella giunzione tra RNA DS-SS, lasciando 8 basi come maniglia al 5'

**Tipo II.** Interviene un tracrRNA (trans acting antisense RNA) codificato sull'altra elica con una complementarietà di sequenza con le sequenze REPEAT del CRISPR RNA per formare un duplex che è riconosciuto e tagliato da RNase III in associazione a Cas9. Questo intermedio è successivamente processato al 5' dando luogo a crRNA di circa 40 nucleotidi con una maniglia di circa 20nt al 3'

**Tipo III.** Troviamo che Cas6 è in grado di tagliare precrRNA 8 nucleotidi a monte della giunzione spacer/repeat. Un'ulteriore maturazione avviene con un processamento al 3'.



## Ruolo delle CAS protein

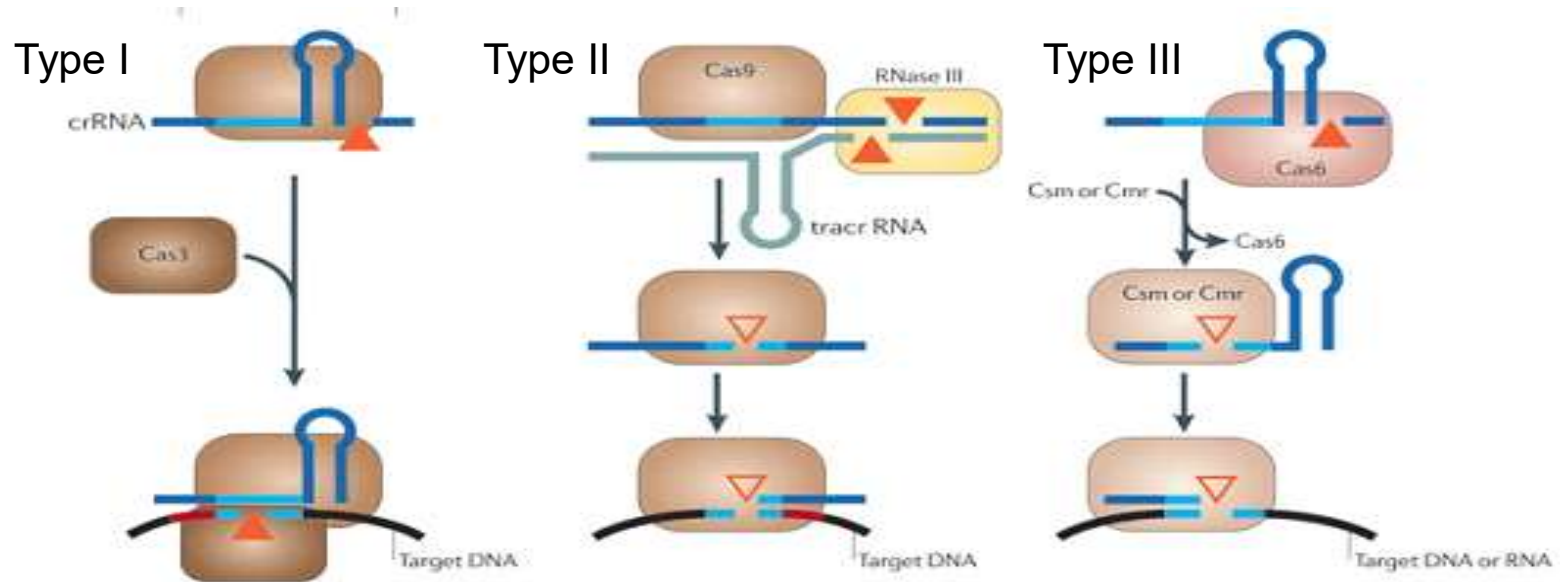
Le proteine CAS sono coinvolte **nell'acquisizione di nuovi spacers** ma contribuiscono a conferire resistenza all'elemento genetico invasivo che può essere un fago o un plasmide

Nella fase di immunità ritroviamo le proteine CAS che **sono essenziali per il processamento del trascritto pre crRNA.**

Infine nella fase di interferenza il **Complesso CAS può indirizzare i crRNA verso il target.**

Nell'appaiamento svolge un ruolo importante la **sequenza PAM**

## Fase 3: Interferenza

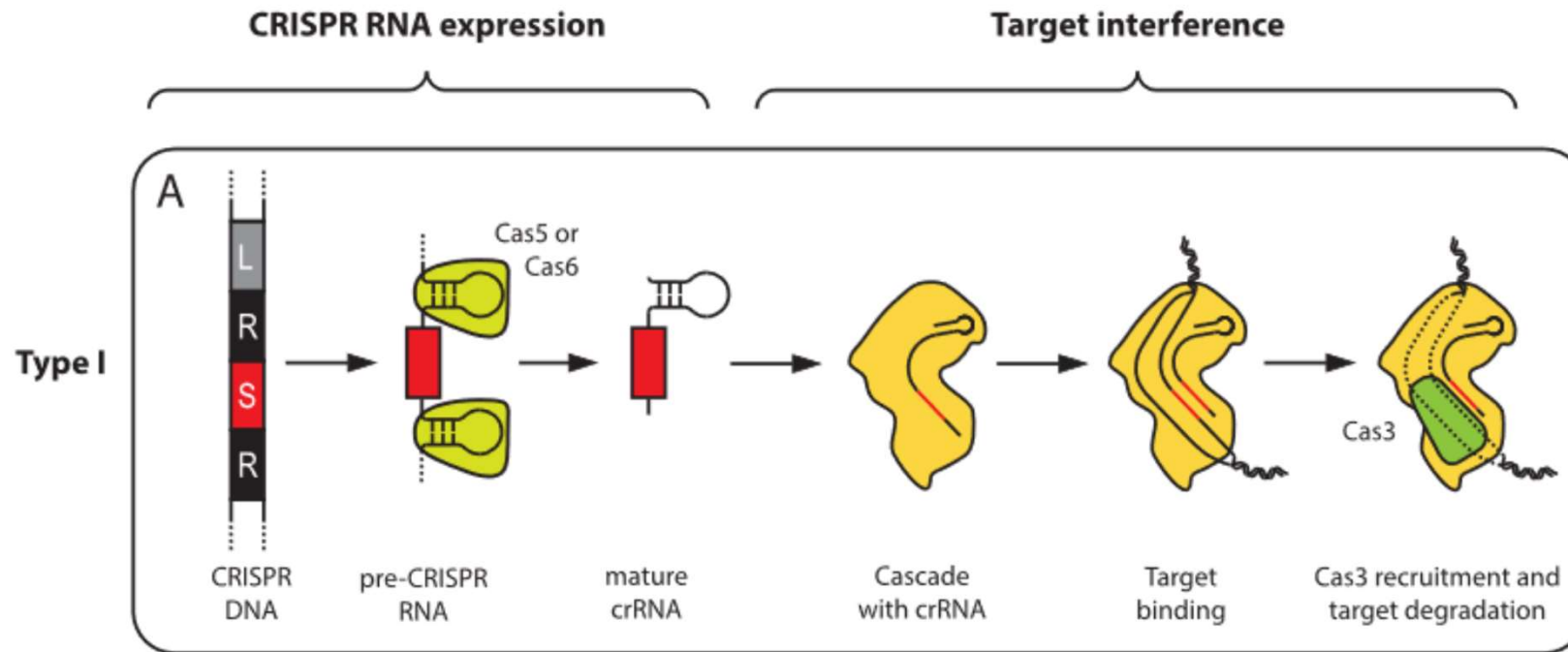


Durante la fase di interferenza l'acido nucleico "invasore" viene tagliato ma con strategie diverse.

Nel Tipo I il crRNA guida il complesso Cascade verso le regioni bersaglio che contengono il DNA complementare che verrà poi tagliato probabilmente dal Cas3. Un ruolo importante per il riconoscimento viene svolto dalle PAM.

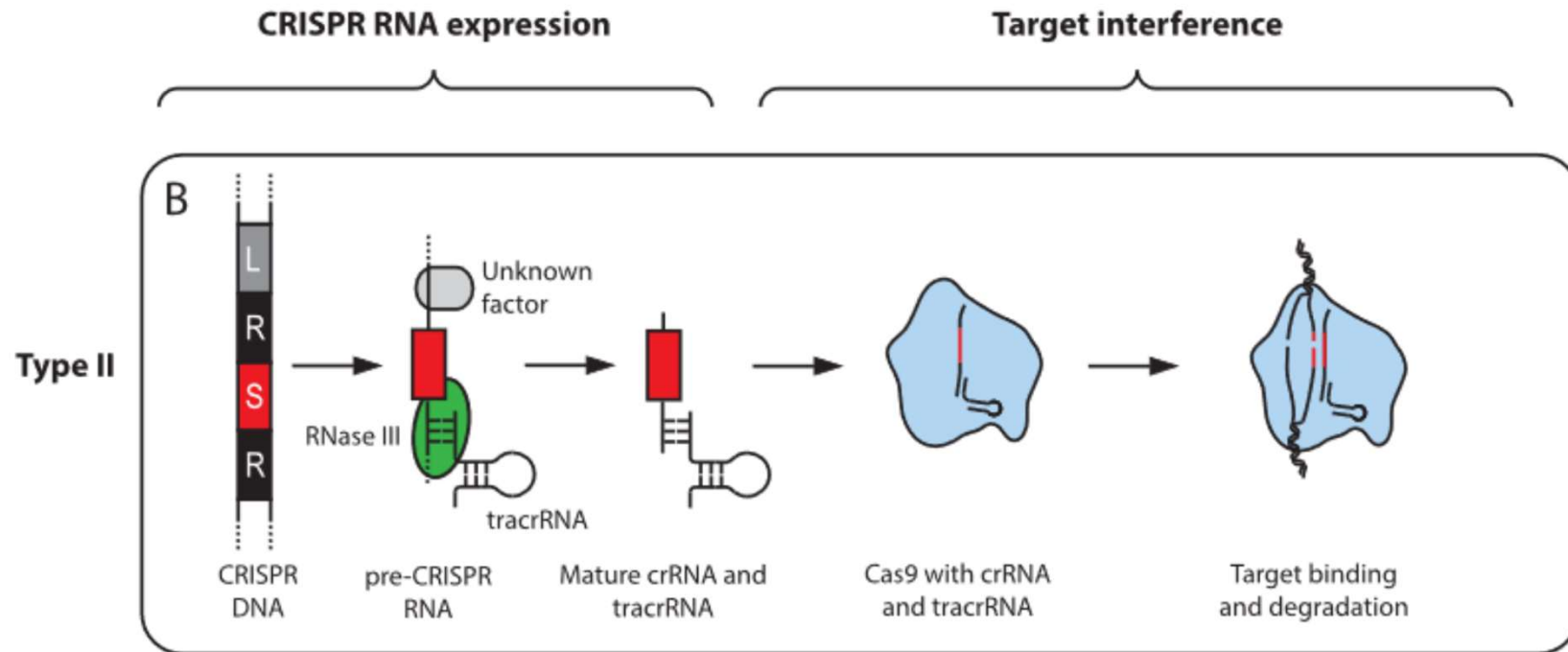
Nel Tipo II Cas9 contenente crRNA punta direttamente il DNA bersaglio con l'aiuto delle PAM.

Nel Tipo III si può avere il riconoscimento sia del DNA che dell'RNA da parte di 2 diversi complessi (Csn o Cmr) ed il successivo taglio da parte di enzimi non ancora identificati

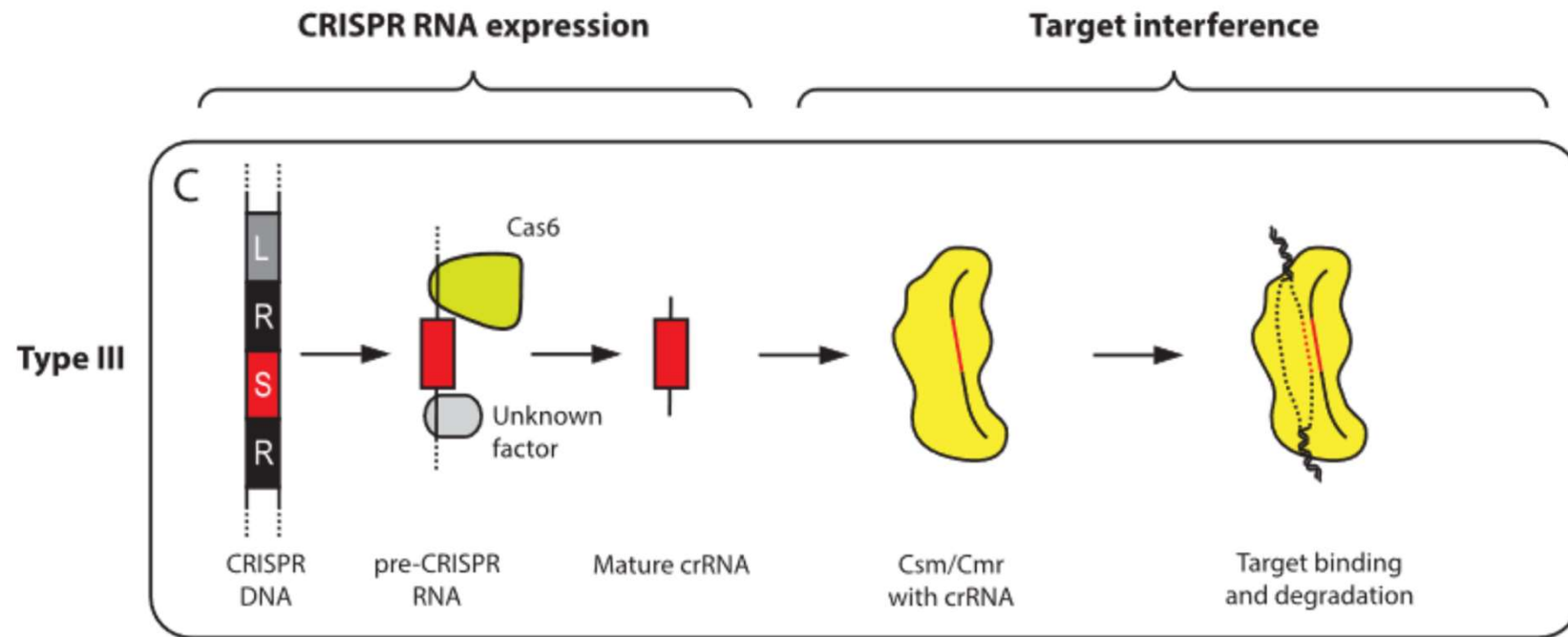


Cas6 e il crRNA sono componenti chiave del complesso Cascade che è un grande complesso multiproteico a forma di dorso di cavallo. Cas5 e 6 legano crRNA.

Per l'interferenza è necessaria Cas3 che viene reclutato in seguito al riconoscimento tra crRNA e target in modo da provocarne la degradazione.

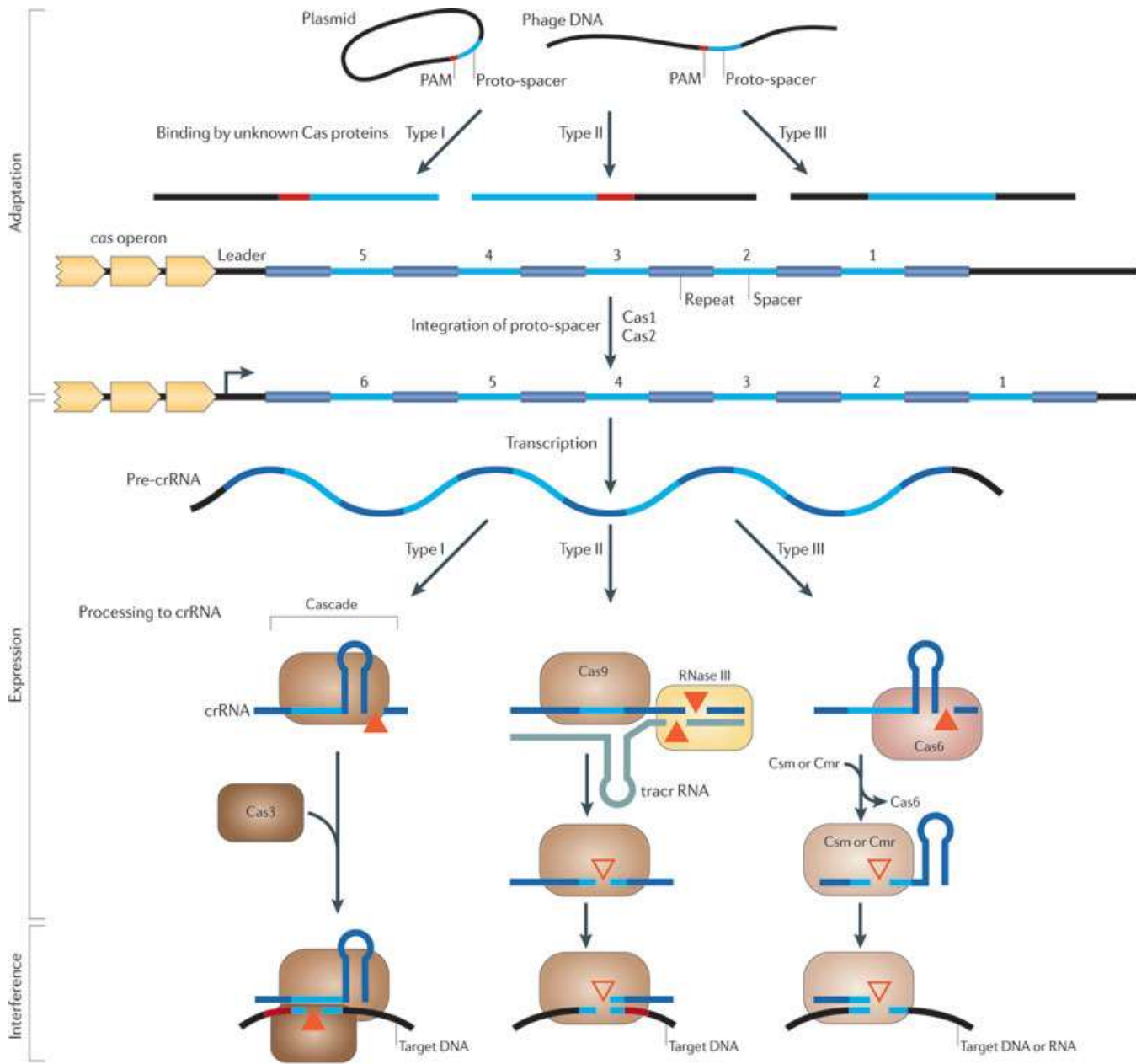


I sistemi di tipo II usano una diversa strategia per la generazione del crRNA in quanto il processamento è dipendente dalla RNasiIII dell'ospite e da un trans encoded sRNA (tracrRNA) che si lega al precrRNA. E' richiesta la proteina Cas9 che porta legato il tracrRNA e crRNA per effettuare il riconoscimento e degradazione

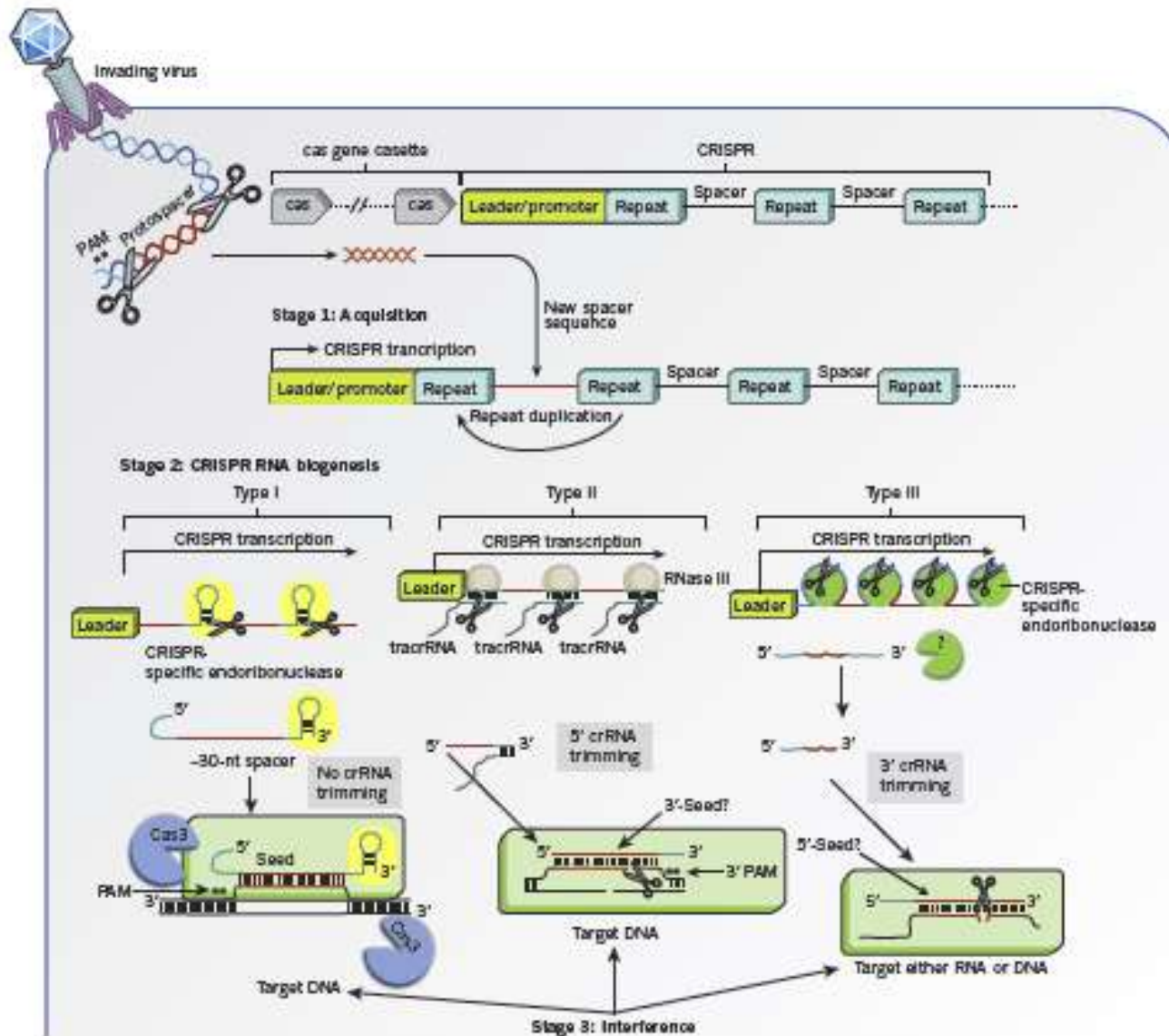


La maturazione del RNA coinvolge Cas6. In seguito tramite un complesso proteico Csm /Csr costituito da diverse proteine si ha il riconoscimento e la degradazione



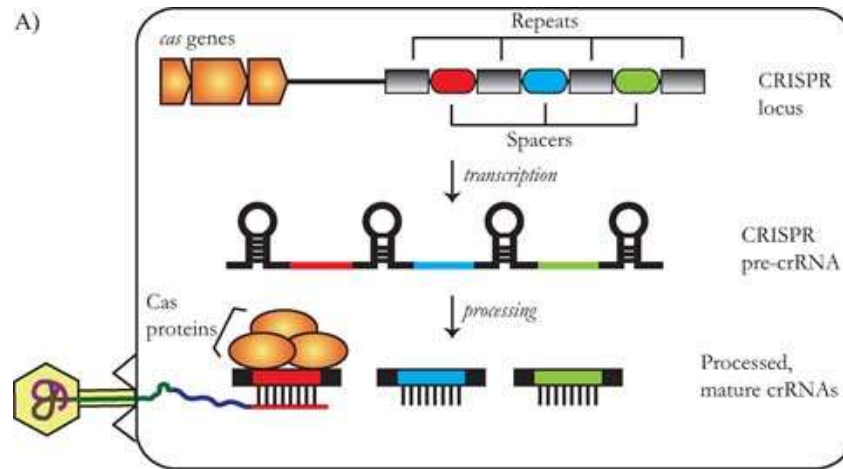


# Visione d'insieme del processo di immunità CRISPR mediato.

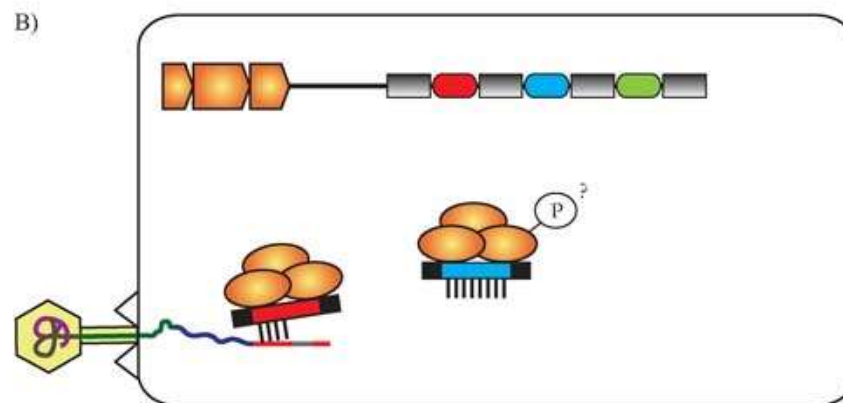




## The phage-host arms race: Shaping the evolution of microbes



Mechanism of action: transcription from the repeat-spacer CRISPR locus generates a long non-coding RNA, with repeats that may sometimes assume a secondary structure. Cleavage of the repeat sequences by the Cas proteins generates crRNAs that target the phage DNA or RNA, and interfere with phage infection.



I fagi possono evadere il interferenza CRISPR con mutazioni o ricombinazioni nella sequenza protospacer. Oppure potrebbero fosforilare le CAS

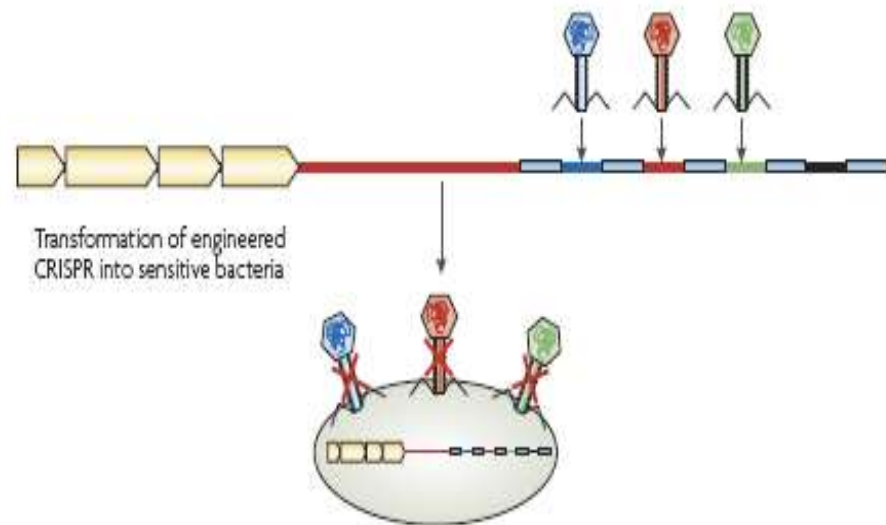


## 2. Costruzione di ceppi fago-resistenti

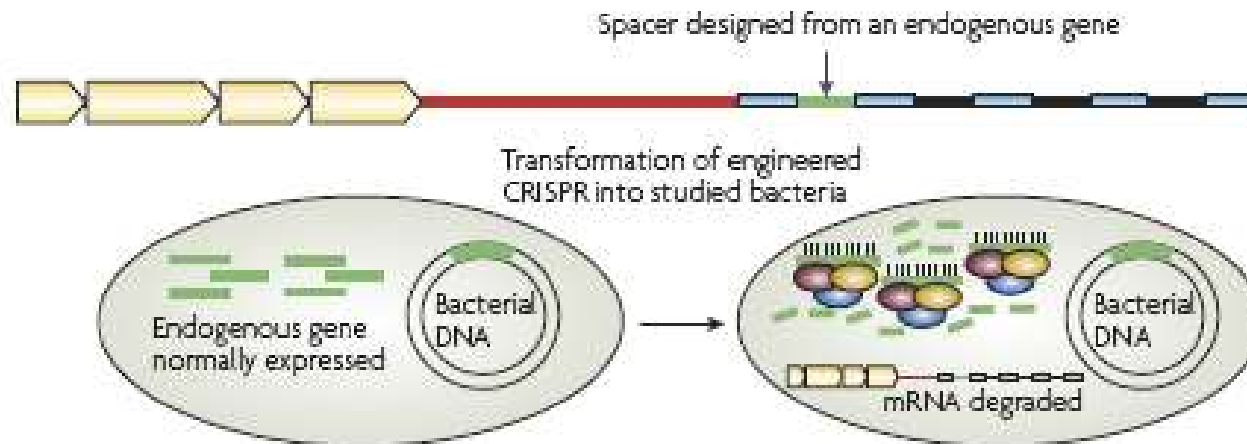
Per molte industrie che utilizzano batteri quali l'industria alimentare e del vino i fagi costituiscono un serio problema che provoca danni ingenti.

Ingegnerizzando i loci CRISPR tramite l'aggiunta in vitro di spacers con sequenze fagiche e poi trasformando questi loci si può indurre resistenza a diversi fagi in batteri sensibili.

Con questa tecnica si può indurre l'immunità nei batteri da utilizzare come starter



### 3. Silenziamento selettivo di geni



Tenendo conto delle analogie tra il sistema CRISPR ed il sistema di RNA interference degli eucarioti si può pensare di introdurre una determinata sequenza all'interno di un locus CRISPR e poi di silenziare il relativo mRNA tramite il sistema CRISPR/CAS.

Si potrebbero costruire dei plasmidi ricombinanti contenenti vari loci CRISPR con spacer idonei per il silenziamento di più geni diversi

## Applicazioni del sistema CRISPR –CAS: 1. Gene therapy

L'applicazione più nota riguarda la dimostrazione che crRNA e tracrRNA possono essere guidati da single guide RNA (sgRNA) verso il target

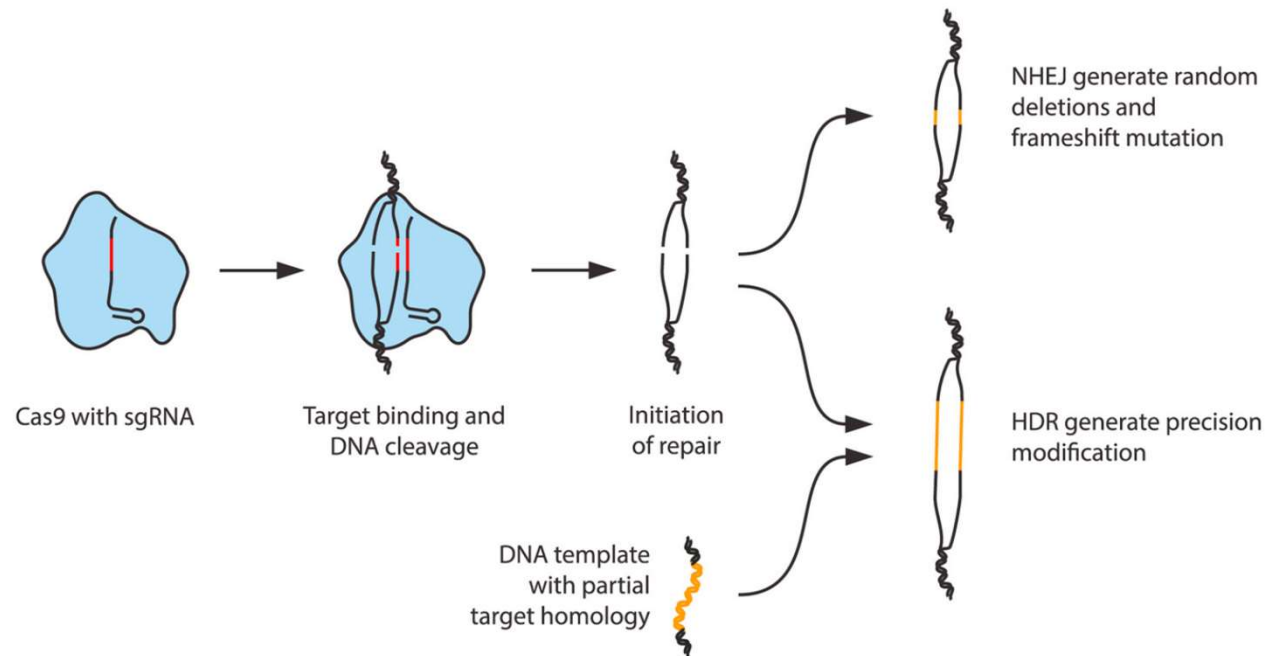


Fig. 5. Examples of basic CRISPR applications. (A) The sgRNA directs Cas9 cleavage of the corresponding target to initiate gene editing. In eukaryotic cells, two main pathways repair DNA damage: Non-Homologous End Joining (NHEJ) and Homology Directed Repair (HDR). NHEJ removes bases, often causing a frameshift and inactivation of the gene. HDR can be used to make specific changes to the target region by providing a designed repair template that becomes inserted in the damaged region. (B)

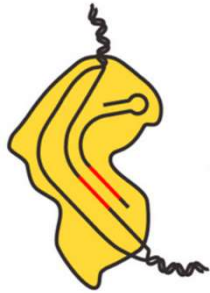
A differenza di Cas3, che progressivamente degrada il target, Cas9 produce un singolo taglio nel DNA a doppia elica. In seguito i sistemi di riparo dell'ospite forniscono due strategie per indurre alterazioni.

In un caso NHEJ ( Non -Homologous End Joining) unisce le estremità ma nel processo può eliminare alcune basi. Nel secondo meccanismo HDR ( Homology Directed Repair (HDR) l'allele danneggiato può essere riparato utilizzando frammento con omologia con il target

## 2. Gene regulation

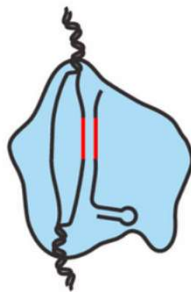
B

Cascade blocks access of RNA polymerase to gene in absence of Cas3



Transcription silencing

Non-cleaving Cas9 blocks access of RNA polymerase to gene



Transcription silencing

Sia il complesso Cascade che una Cas9 nucleasi difettiva possono essere utilizzati per interferire con il legame della RNA polimerasi o con la fase di allungamento.

CRISPR–Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated proteins) systems act in three stages: adaptation, expression and interference. In type I and type II CRISPR–Cas systems, but not in type III systems, the selection of proto-spacers in invading nucleic acid probably depends on a proto-spacer-adjacent motif (PAM)<sup>22,30,31</sup>, but how the PAM or the nucleic acid is recognized is still unclear. After the initial recognition step, Cas1 and Cas2 most probably incorporate the proto-spacers into the CRISPR locus to form spacers. During the expression stage, the CRISPR locus containing the spacers is expressed, producing a long primary CRISPR transcript (the pre-crRNA). The CRISPR-associated complex for antiviral defence (Cascade) complex binds the pre-crRNA, which is then cleaved by the Cas6e or Cas6f subunits (in subtype I-E or I-F, respectively), resulting in crRNAs with a typical 8-nucleotide repeat fragment on the 5' end and the remainder of the repeat fragment, which generally forms a hairpin structure, on the 3' flank. Type II systems use a trans-encoded small RNA (tracrRNA) that pairs with the repeat fragment of the pre-crRNA, followed by cleavage within the repeats by the housekeeping RNase III in the presence of Cas9 (formerly known as Csn1 or Csx12). Subsequent maturation might occur by cleavage at a fixed distance within the spacers<sup>25</sup>, probably catalysed by Cas9. In type III systems, Cas6 is responsible for the processing step, but the crRNAs seem to be transferred to a distinct Cas complex (called Csm in subtype III-A systems and Cmr in subtype III-B systems). In subtype III-B systems, the 3' end of the crRNA is trimmed further<sup>28</sup>. During the interference step, the invading nucleic acid is cleaved. In type I systems, the crRNA guides the Cascade complex to targets that contain the complementary DNA, and the Cas3 subunit is probably responsible for cleaving the invading DNA<sup>21</sup>. The PAM probably also plays an important part in target recognition in type I systems. In type II and type III systems, no Cas3 orthologue is involved ([Table 2](#)). In type II systems, Cas9 loaded with crRNA probably directly targets invading DNA, in a process that requires the PAM<sup>26</sup>. The two subtypes of CRISPR–Cas type III systems target either DNA (subtype III-A systems<sup>31</sup>) or RNA (subtype III-B systems<sup>28</sup>). In type III systems, a chromosomal CRISPR locus and an invading DNA fragment are distinguished by either base pairing to the 5' repeat fragment of the mature crRNA (resulting in no interference) or no base pairing (resulting in interference)<sup>30</sup>. Filled triangles represent experimentally characterized nucleases, and unfilled triangles represent nucleases that have not yet been identified.



Spacer selection appears guided by certain sequence elements in the target. Analysis of target sequences has revealed a short motif next to the target sequence called [protospacer adjacent motif](#) (PAM) that is crucial for discrimination between [self](#) and non-self [\[46\]](#) (see Section [5](#)). While initially thought to be important only for interference, the PAM also has a role in spacer acquisition. This is supported by the fact that most newly acquired spacers have a PAM next to their protospacer [\[35\]](#), [\[37\]](#). In the Type II-A system, Cas9 is responsible for identifying the PAM as mutations that disable Cas9's PAM recognition result in acquisition from protospacers without PAM [\[20\]](#). In Type I-E, PAM recognition during spacer acquisition may be different as they are indicated to be identified by Cas1-Cas2 alone [\[36\]](#). However, Cascade increases the frequency of correct PAMs for inserted spacers [\[](#)

nce located, it is not known [if](#) the protospacer is copied or cut out of the target. The production of spacer material could be linked to other defense systems, such as the [restriction-modification system](#) [\[47\]](#), similar to primed spacer acquisition (see below). Such coupling may facilitate the systems' recognition of “foreign DNA” or provide the system with material suitable for new spacers. Infection by a [phage](#) incapable of reproduction, which *e.g.* only packaged a partial [viral genome](#), could act as a vaccine and facilitate adaptation.

Although spacer acquisition is observed, the mechanism is only partly understood. Conceptually, the process can be divided into two steps: protospacer selection and generation of spacer material followed by integration of the spacer into the CRISPR array and synthesis of a new repeat. Occasional deletion of spacers is required to limit the size of the CRISPR, but there is little knowledge of the mechanism or frequency of such events.

The key factors in spacer integration are Cas1 and Cas2. This function was suggested early as the proteins are ubiquitous but dispensable for interference

[10]

. This was later con

firm

irmed by

overexpression of Cas1 and Cas2 from a Type I-E system in

*E. coli*

,

which resulted in spacer integration even in the absence of all other

Cas proteins

[36]

. Both Cas1 and Cas2 are nucleases

[42

e

44]

and

mutations in the active site of Cas1 abolishes spacer integration in

*E. coli*

[36,37]

. Cas1 and Cas2 from

*E. coli*

form a complex where one

Cas2 dimer binds two Cas1 dimers. F

Il processo di adattamento può essere suddiviso in due steps: selezione dei protospacer e generazione del nuovo spacer seguito dall'inserimento dello spacer nelle sequenze CRISPR con sintesi di un nuovo repeat. Talvolta per un inserire un nuovo spacer è richiesta la delezione di uno spacer precedentemente esistente.

I fattori essenziali per integrazione degli spacer sono Cas1 e Cas2.

Entrambi sono nucleasi e mutazioni in Cas1 aboliscono l'inserimento degli spacers. Cas1 e Cas2 di *E. coli* formano un complesso costituito da dimero di ciascuna proteina