

## Historique de la découverte des systèmes CRISPR-Cas

Christine Pourcel\*

Institut de Biologie Intégrative de la Cellule (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette cedex, France

Reçu le 26 février 2018

**Résumé** – Entre la fin des années 1980 et le début des années 2000, quelques équipes explorant les génomes de bactéries et d'archées mettent en évidence le système immunitaire adaptatif procaryote constitué de la séquence CRISPR et des gènes associés *cas*. Tout d'abord comparé au système d'interférence médié par l'ARN des eucaryotes, ce système se révèle unique et d'une complexité génétique étonnante. Par la suite, l'étude comparative des séquences CRISPR et des gènes *cas*, puis celle de métagénomes microbiens permettent de proposer un schéma évolutif de ces systèmes. Ces résultats rappellent l'importance de l'approche naturaliste, sans *a priori*, pour la compréhension du monde vivant.

**Mots clés** : CRISPR-Cas, immunité adaptative procaryote, génome, évolution, interférence

**Abstract - An history of the CRISPR-Cas systems discovery.** From 1987 and during the following 20 years, a few research teams exploring bacteria and archea genome sequences uncover the prokaryotic adaptative immune system made of the CRISPR sequence and associated *cas* genes. First believed to be similar to the eukaryote RNA interference system, CRISPR-Cas turned out to be unique and of an amazing genetic complexity. The comparative studies of CRISPR arrays and of *cas*, and later of microbiotes metagenomes allowed to propose an evolution scenario for these systems. The results demonstrate the importance of a naturalistic approach, without *a priori*, for the understanding of living organisms.

**Keywords:** prokaryote adaptative immunity, genome, evolution, interference, CRISPR

### La séquence CRISPR, une structure inhabituelle et d'origine inconnue

En 1987, Ishino et coll. observent, après la séquence du gène *iap* dans le génome d'*Escherichia coli*, une structure constituée de la répétition de cinq séquences conservées de 29 nucléotides, séparées par des espaceurs ou *spacers* uniques de 32 nucléotides (Ishino *et al.*, 1987). Une seconde étude par cette même équipe montre que cette séquence n'est pas codante et qu'elle est retrouvée dans d'autres espèces proches. Aucune fonction ne lui est attribuée (Nakata *et al.*, 1989). Des structures similaires sont observées au cours des dix années suivantes par différentes équipes, chez des bactéries et des archées. Différents noms leur sont attribués, reflétant leurs caractéristiques : *Direct Variable Repeat* ou DVR (Groenen *et al.*, 1993), *Tandem Repeat* ou TREP (Mojica *et al.*, 1995), *Short Regularly Spaced Repeats* ou SRSR (Mojica *et al.*, 2000), *Long Cluster of Tandem Repeat* ou LCTR (Nelson *et al.*, 1999),

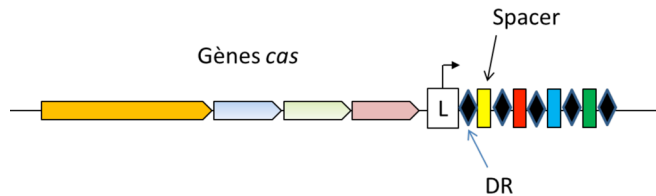
*Spacers Interspaced Direct Repeats* ou SPIDR (Jansen *et al.*, 2002a). L'acronyme CRISPR pour *Cluster of Regularly Interspaced Short Palindromic Repeats* est proposé par Jansen et coll. en 2002 (Jansen *et al.*, 2002b). Plus que les répétitions qui sont des éléments couramment observés, ce sont les séquences uniques qui intriguent. On ne trouve aucune homologie avec les séquences présentes dans les bases de données et elles ne présentent pas de caractéristiques communes. Les répétitions, appelées DR « *direct repeat* », semblent par contre posséder un palindrome qui pourrait leur conférer un rôle biologique (Figure 1). La présence d'un palindrome et la localisation des CRISPR dans certains génomes d'archées conduisent à proposer un rôle dans la réplication (Mojica *et al.*, 2000) ou dans le transfert d'ADN (Nelson *et al.*, 1999), mais ces hypothèses apparaîtront inexactes.

L'étude publiée en 2002 par Tang et coll. démontre que la structure répétée est transcrite abondamment chez *Archaeoglobus fulgidus*. Au cours de l'exploration de petits ARN, souvent impliqués dans la régulation de l'expression de gènes, les auteurs observent des ARN d'une taille correspondant à un DR plus un *spacer*, ainsi que des

\*Auteur correspondant : [christine.pourcel@gmail.com](mailto:christine.pourcel@gmail.com)

<i>Thermofilum pendens</i>	GTATCAACAACGAATGAGTTGA <span style="border: 1px solid black; padding: 0 2px;">AAG</span> 25pb
<i>Verminephrobacter eisenia</i>	TTTCTGAGCTGCCATCCGGCAGTG <span style="border: 1px solid black; padding: 0 2px;">AAC</span> 28pb
<i>Corynebacterium glutamicum</i>	GGAAATACCTCCCGTGTGCGGAGAAGAG 29pb
<i>Desulfotomaculum reducens</i>	GTCAAAACACAAAATAATTCCCTTTGGGAATTGA <span style="border: 1px solid black; padding: 0 2px;">AAC</span> 37pb
<i>Azoarcus sp.</i>	GCATCGCCCCTCGGTGACGGGGGGCGTGGATTGA <span style="border: 1px solid black; padding: 0 2px;">AAC</span> 37pb
<i>Flavobacterium psychrophilum</i>	<span style="border: 1px solid black; padding: 0 2px;">GTT</span> GTAATTATAAGCTAAAATACAATTTTGAAGCAATTAC <span style="border: 1px solid black; padding: 0 2px;">AAC</span> 46pb

**Figure 1.** Diversité des séquences DR. Séquences du DR de six CRISPR appartenant à différentes espèces bactériennes. Le palindrome est surligné et la séquence 3' terminale AAC/G est entourée.



**Figure 2.** Représentation schématique d'un système CRISPR-Cas. Le groupe de gènes *cas* est suivi du *leader* représenté par une boîte marquée L et d'une succession de séquences DR (losanges noirs) séparés par les *spacers* (rectangles colorés).

multiples de ces ARN et hybridant à différentes positions sur trois locus CRISPR (Tang *et al.*, 2002). Ils en concluent qu'un long transcrypt précurseur est mûri par coupures progressives en petits ARN et proposent l'hypothèse selon laquelle ces molécules pourraient jouer un rôle dans le fonctionnement du CRISPR. Cette observation sera faite chez d'autres espèces (Lillestol *et al.*, 2006), confirmant l'importance de l'étape de transcription.

## Les gènes *cas*

En 2002, Jansen *et coll.* montrent l'existence d'un groupe de gènes particuliers toujours situés près d'un CRISPR, et les appellent gènes *cas* pour CRISPR-associated (Jansen *et al.*, 2002a). Ensemble ils forment les systèmes CRISPR-Cas. Parmi les gènes *cas* analysés dans cette étude, quatre sont conservés chez différentes espèces d'archées et de bactéries, et deux présentent des similarités avec des membres des *Cluster of Orthologous Groups* (COG). L'étude montre que dans un génome possédant plusieurs CRISPR avec le même DR, un seul est associé au groupe de gènes *cas*. L'équipe décrit également le *leader*, une séquence flanquant le CRISPR sur un côté et possédant un motif TATA, et qui sera plus tard identifié comme le promoteur pour la transcription du CRISPR (Jansen *et al.*, 2002a; Lillestol *et al.*, 2006) (Figure 2).

Le séquençage de nouveaux génomes procaryotes conduit à découvrir la très grande diversité des gènes *cas*. Un travail de comparaison des protéines CAS présentes dans 200 génomes procaryotes complets permet à Haft *et coll.* (2005) de proposer une première classification constituée de 45 familles et de multiples sous-types.

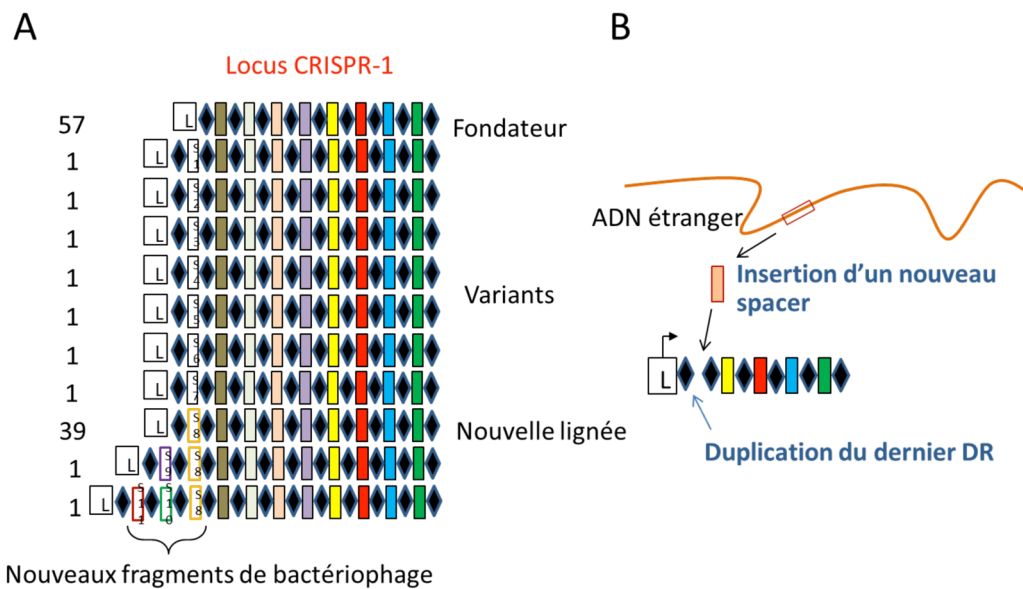
Cette classification continuera à évoluer avec l'analyse de nouveaux génomes séquencés (voir paragraphe « Diversité des systèmes CRISPR-Cas » ci-dessous).

## Origine des *spacers* et première découverte d'un rôle des CRISPR

En 2005 sont publiées trois études montrant de manière indépendante que les *spacers* correspondent majoritairement à des séquences virales ou plasmidiques (Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel *et al.*, 2005). Deux des études *in silico* explorent les séquences génomiques procaryotes disponibles et réussissent à assigner une origine probable à 2% des *spacers* par homologie de séquence (Bolotin *et al.*, 2005; Mojica *et al.*, 2005). Bolotin *et coll.* décrivent de plus la présence d'une courte séquence Pu-Py-A-A proche de la séquence homologue du *spacer*, appelée *protospacer*, et qui s'aligne avec la fin du DR (...ACAAC) chez *Streptococcus thermophilus*. Cette séquence sera à nouveau décrite en 2008 (Deveau *et al.*, 2008), puis appelée *Protospacer Adjacent Motif* ou PAM en 2009 (Mojica *et al.*, 2009). Dans la troisième étude, Pourcel *et coll.* analysent 108 souches clonales de *Yersinia pestis*, obtenues au cours d'une épidémie et montrent que le CRISPR-Cas acquiert de nouveaux *spacers* de manière polarisée (Figure 3). On peut, en comparant leur agencement, retracer une histoire évolutive de l'espèce bactérienne (Pourcel *et al.*, 2005; Cui *et al.*, 2008). Les nouveaux *spacers* ont principalement pour origine un prophage de *Yersinia pestis*.

Dans l'étude de Pourcel *et coll.*, l'observation par analyse bioinformatique des séquences génomiques de souches de *Streptococcus pyogenes* (groupe A) permet de montrer que lorsqu'un *spacer* correspond à un phage, celui-ci n'est pas présent comme prophage dans le génome bactérien (Pourcel *et al.*, 2005). De même, Bolotin *et coll.* font le lien entre le nombre de *spacers* correspondant à un phage dans un CRISPR et le niveau de résistance à ce phage (Bolotin *et al.*, 2005).

Ces observations permettent de proposer pour la première fois l'hypothèse du rôle de CRISPR-Cas comme mécanisme de défense contre les phages, similaire à l'acquisition d'une immunité (Pourcel *et al.*, 2005). Les articles de 2005 avaient ouvert la voie à l'étude fonctionnelle



**Figure 3.** Acquisition de nouveaux *spacers* dans le CRISPR-1 de *Yersinia pestis*. A : organisation du CRISPR dans 11 allèles. Les *spacers* S1 à S11 sont nouvellement acquis pendant l'épidémie. B : mécanisme d'insertion d'un nouveau *spacer* près du *Leader* (L).

et mécanistique des systèmes CRISPR-Cas (Morange, 2015 ; Lander, 2016).

## Analyse du fonctionnement du système CRISPR-Cas

Barrangou et coll. (2007) sont les premiers à confirmer expérimentalement la prédiction faite en 2005 du rôle du système CRISPR-Cas. *S. thermophilus* utilisée comme ferment pour produire des yaourts possède trois CRISPR et deux groupes de gènes *cas*. Certaines parmi les souches résistant à une infection par un bactériophage possèdent dans le CRISPR, nommé CRISPR1, un nouveau *spacer* correspondant à une région du génome viral. En utilisant différents mutants, les chercheurs confirment le rôle du CRISPR dans la résistance et montrent le rôle de deux protéines Cas dans le fonctionnement de l'immunité. Cas7 est impliquée dans l'addition de nouveaux *spacers*, alors que Cas5 (ultérieurement renommée Cas9) est nécessaire à la résistance. Ces travaux enthousiasment la communauté scientifique et de nombreuses équipes entreprennent des recherches sur cette thématique.

Une première hypothèse émise sur le fonctionnement du système d'immunité avait été avancée du fait des similarités observées avec les systèmes d'interférence eucaryote médiés par les siRNA (*small interfering RNA*) et miRNA (microRNA) (Mak, 2005). L'existence de petits ARN dérivés du CRISPR (crRNA) et les similarités fonctionnelles entre certaines protéines Cas et des protéines de type *dicer* et *slicer* avaient conduit Makarova et coll. (2006) à proposer le nom piRNA pour *prokaryotic interference RNA*. Cependant, les études de Van de Oost et coll. sur le fonctionnement du système de Type I de *E. coli* montrent qu'il s'agit d'un mécanisme complètement différent (Brouns et al., 2008). Les chercheurs

caractérisent un complexe de cinq protéines Cas, appelé Cascade, permettant la maturation de l'ARN précurseur CRISPR (pré-crRNA), et également la destruction de la cible en association avec la protéine Cas3. Ils montrent que le crRNA guide le complexe jusqu'à sa cible et que celle-ci n'est pas l'ARN comme précédemment supposé par certains mais l'ADN. Cette observation est également apportée par les résultats de Marraffini & Sontheimer (2008) sur l'inhibition du transfert horizontal d'un plasmide par le système CRISPR-Cas chez *Staphylococcus epidermidis*. Le travail décrit par Garneau et coll. en 2010 démontre que chez *S. thermophilus*, le système CRISPR-Cas (de type II) conduit à une coupure double-brin de l'ADN (Garneau et al., 2010).

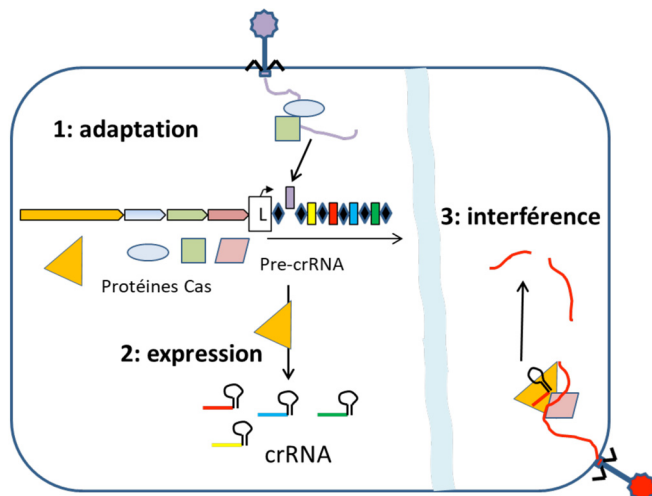
Les trois fonctions principales nécessaires à l'établissement de l'immunité sont ainsi décrites (Figure 4) :

- 1 : l'adaptation ou acquisition qui conduit à l'ajout d'un fragment d'ADN d'un génome invasif dans la structure CRISPR ;
- 2 : l'expression qui se traduit notamment par la maturation du pré-crARN couvrant l'intégralité du CRISPR en petits ARN ou crARN contenant chacun la séquence d'un *spacer* ;
- 3 : l'interférence ou immunité qui correspond au ciblage du génome étranger et à sa coupure par une endonucléase.

Une ou plusieurs protéines Cas sont nécessaires à chacune de ces fonctions et la cible est le plus souvent l'ADN mais peut également être une molécule d'ARN.

## Diversité des systèmes CRISPR-Cas

À ce jour l'analyse de l'ensemble des systèmes CRISPR-Cas des génomes entièrement séquencés permet



**Figure 4.** Fonctionnement du système CRISPR-Cas. Trois fonctions sont communes à tous les systèmes :

- 1 : acquisition d'un nouveau *spacer* ;
- 2 : expression du système, notamment production des petits ARN ;
- 3 : interférence conduisant à la destruction de la cible.

de définir deux classes, six types et 33 sous-types (Koonin *et al.*, 2017). Dans les Types I, III et IV qui constituent la classe 1, l'interférence est réalisée par un complexe de plusieurs protéines Cas (appelé Cascade dans le Type I) alors que pour les Types II, V et VI une grande protéine multifonctionnelle telle que Cas9 est généralement suffisante. Dans les systèmes de type II et V, un petit ARN supplémentaire appelé *tracrRNA* est impliqué dans la maturation des crARN et l'adressage de l'endonuclease jusqu'à sa cible (Deltcheva *et al.*, 2011 ; Jinek *et al.*, 2012). Dans tous les systèmes, l'adaptation est effectuée par les protéines Cas1 et Cas2 avec parfois la protéine Cas4.

La caractérisation du rôle des Cas a permis de proposer leur utilisation comme outil pour la modification des génomes (Jinek *et al.*, 2012). La protéine Cas9 est actuellement la plus utilisée mais d'autres protéines de systèmes de classe 2 montrent des propriétés prometteuses (Barrangou & Doudna, 2016).

## Le mystère de l'adaptation

La nature des protéines Cas entrant en jeu dans la phase d'adaptation a été progressivement décrite mais deux questions se posent encore :

- 1 : comment la cellule contrôle-t-elle le cycle d'un virus ou l'insertion d'un ADN étranger suffisamment longtemps pour permettre la mise en place de l'immunité ?
- 2 : quels sont les signaux permettant au système CRISPR-Cas de distinguer l'ADN étranger de l'ADN cellulaire ?

Des éléments de réponse ont été apportés mais ils ne permettent pas à ce jour de répondre totalement à ces questions.

Au laboratoire, l'adaptation a été observée dans quelques rares espèces bactériennes. *S. thermophilus* semble utiliser CRISPR-Cas comme système de défense de manière prédominante dans plusieurs souches testées (Horvath *et al.*, 2008). Cependant, l'expression du système chez *S. thermophilus* représenterait un coût élevé (*fitness cost*) pour la bactérie (Vale *et al.*, 2015). Chez *E. coli* K12, le système est fonctionnel mais il est réprimé et nécessite la surexpression de Cas1 et Cas2 (Swarts *et al.*, 2012). L'acquisition de nouveaux *spacers* a été observée chez *Sulfolobus solfataricus* uniquement en réponse à l'infection par un mélange de virus directement purifiés de l'environnement et contenant un plasmide conjugatif (Erdmann & Garrett, 2012). Enfin, chez *Pseudomonas aeruginosa* l'adaptation est active contre certains dérivés d'un phage tempéré, DMS3, proche du phage transposable de *E. coli* Mu (Cady *et al.*, 2012). La facilitation est un terme proposé pour expliquer la possibilité de construire une immunité contre un virus avant que celui-ci ne détruise la cellule infectée (Abedon, 2011). Il faut pour ceci invoquer un rôle secondaire pour le système CRISPR-Cas dans la résistance aux virus, la résistance innée étant la première barrière à l'infection virale.

La séquence PAM localisée dans certains types CRISPR-Cas directement à côté du *protospacer* sur le génome d'origine est nécessaire à l'action des protéines Cas dans l'étape d'adaptation (Bolotin *et al.*, 2005 ; Deveau *et al.*, 2008 ; Mojica *et al.*, 2009). C'est également cette séquence qui prévient l'autoimmunité car elle est nécessaire au fonctionnement de l'interférence (Stern *et al.*, 2010). La majorité des *protospacers* identifiés à ce jour correspondent à des virus, plasmides, ou éléments transposables mais il existe un pourcentage non négligeable de *protospacers* correspondant à des séquences chromosomiques (Cui *et al.*, 2008). Le système CRISPR-Cas pourrait donc, chez certains procaryotes, jouer un autre rôle. Chez différentes espèces bactériennes pathogènes, on a pu, par exemple, montrer un effet du système CRISPR-Cas dans le niveau de virulence (Zegans *et al.*, 2009 ; Gunderson & Cianciotto, 2013 ; Louwen *et al.*, 2013 ; Sampson *et al.*, 2013 ; Zhang *et al.*, 2013).

## Origine et dispersion des systèmes CRISPR-Cas

Les structures CRISPR étant différentes des répétitions classiquement observées dans les génomes, des outils informatiques dédiés sont développés dès 2007, permettant d'explorer facilement les génomes. Les premiers programmes sont CRT (Bland *et al.*, 2007), *Piler-CR* (Edgar, 2007) et *CRISPRfinder* (Grissa *et al.*, 2007b). CRISPRdb, une base de données consultable sur internet permet un accès facile à ces séquences (Grissa *et al.*, 2007a). Les analyses montrent que la majorité des archées et 48 % des bactéries possèdent au moins un CRISPR. Plus récemment de nouveaux algorithmes ont été produits permettant l'orientation du CRISPR et la recherche du *leader* (Alkhnbashi *et al.*, 2016 ; Biswas *et al.*, 2016).

L'abondance de données de séquençage massif a suscité la création de nouveaux programmes tels que CRASS (Skenneron *et al.*, 2013) et CRISPR *detection* (Ben-Bassat & Chor, 2016) permettant la recherche des CRISPR directement dans des fichiers de courtes lectures ou *reads*, tels que produits par les principales machines de séquençage. La recherche des protéines Cas et leur classification est facilitée par des outils comme *MacSyfinder* (Abby *et al.*, 2014) et HMMCAS (Chai *et al.*, 2017).

En alignant les protéines Cas1 de 54 génomes procaryotes, Haft *et coll.* (2005) ont remarqué que la phylogénie observée ne correspondait pas à celle des espèces. De plus, des systèmes CRISPR-Cas sont présents sur des plasmides et des îlots génomiques transférables, et des transposons sont fréquemment observés à proximité (Godde & Bickerton, 2006). Le transfert horizontal semble donc jouer un rôle important dans la dispersion des systèmes CRISPR-Cas chez les procaryotes. Leur origine reste inconnue mais le casposon, un transposon auto-synthétisé, pourrait représenter la forme primitive (Krupovic *et al.*, 2014, 2017).

## Analyse de microbiomes complexes

L'équipe de Jill Banfield a été la première à rechercher dans des données de métagénomes, la présence de structures CRISPR et à identifier, grâce à l'analyse des *spacers*, la présence simultanée des bactéries ou archées et de leurs virus (Andersson & Banfield, 2008; Tyson & Banfield, 2008). Ces études ont permis d'observer la dynamique des populations microbiennes dans différents environnements (Emerson *et al.*, 2013). Cette approche a conduit à la découverte de nouveaux CRISPR et de nouveaux gènes Cas ainsi qu'à la reconstruction de la séquence de virus sources de *protospacers* (Rho *et al.*, 2012). Burstein *et coll.* (2017) ont rapporté, par exemple, la présence d'un gène Cas9 chez une archée inconnue précédemment, alors qu'aucun système CRISPR-Cas de type II n'avait été décrit chez une archée.

## Coévolution virus/hôte

Une mutation ou délétion dans le PAM ou dans le *protospacer* permet à un virus d'échapper à l'action d'un système CRISPR-Cas. Ce mécanisme est appelé évasion et il a été décrit dès les premières analyses du fonctionnement du système (Paez-Espino *et al.*, 2015). De manière plus surprenante, plusieurs études ont permis de mettre en évidence des gènes anti-CRISPR chez des bactériophages. Chez *P. aeruginosa*, des phages proches du phage transposable Mu de *E. coli* possèdent de courts gènes codant pour des polypeptides capables de bloquer l'interférence d'un système de type I-E ou I-F (Bondy-Denomy *et al.*, 2013; Pawluk *et al.*, 2014). Plus récemment, ces auteurs ont détecté, par homologie de séquence, l'existence de protéines anti-Cas9 (Pawluk *et al.*, 2016). La grande diversité des systèmes CRISPR-Cas pourrait être une réponse à l'adaptation des virus.

Une catégorie de phages de *Vibrio cholerae* possède un système CRISPR-Cas complet de type I-F. Certains *spacers* correspondent à des fragments d'un îlot chromosomique de 18 kb inhibant la réplication du phage, ce qui laisse supposer que le système CRISPR-Cas contrôle ce mécanisme de défense (Seed *et al.*, 2013). Chez *Clostridium difficile*, des prophages portent un CRISPR possédant des *spacers* contre d'autres phages (Hargreaves *et al.*, 2014).

## Les CRISPR comme outil d'épidémiologie moléculaire

Les éléments CRISPR sont composés d'unités DR+ *spacers* ajoutées séquentiellement après la *leader* (Pourcel *et al.*, 2005). L'insertion à tout autre endroit est très rare et l'on observe donc une évolution de la structure dans le temps avec à une extrémité du CRISPR, les *spacers* les plus anciens, et à l'autre, du côté de la *leader*, les traces des toutes dernières rencontres avec un ADN étranger. Le CRISPR peut perdre des unités DR+ *spacer* par recombinaison interne et l'on observe parfois des duplications. L'analyse de souches d'une espèce possédant un CRISPR permet d'observer un polymorphisme dans le nombre et la nature des *spacers* et d'en déduire des relations phylogénétiques. En 1993, Groenen *et coll.* décrivent une nouvelle méthode permettant le génotypage de *Mycobacterium tuberculosis* et basée sur le polymorphisme d'un locus CRISPR, appelé alors DVR et observé depuis 1991 (Hermans *et al.*, 1991; Groenen *et al.*, 1993). Cette méthode appelée Spoligotypage sera appliquée au typage de dizaines de milliers de souches du complexe *M. tuberculosis* et apportera des éléments essentiels sur l'évolution du complexe (Filliol *et al.*, 2003). Pour *Yersinia pestis* et plus largement *Yersinia pseudotuberculosis*, l'analyse du polymorphisme des CRISPR peut présenter un intérêt en épidémiologie moléculaire (Pourcel *et al.*, 2005; Cui *et al.*, 2008; Riehm *et al.*, 2012; Koskela *et al.*, 2015).

## Conclusion

Les nombreuses études effectuées depuis la découverte d'un CRISPR en 1987 ont permis d'établir que le système CRISPR-Cas est un mécanisme d'immunité adaptative chez les procaryotes. Il semble cependant jouer un rôle essentiel principalement chez les archées et les bactéries hyperthermophiles qui possèdent peu de moyens de défense innés. Leur évolution et leur dispersion, ainsi que les interactions avec les virus, ont donné naissance à une très grande diversité. La découverte de ces systèmes, dont on ignorait totalement l'existence, est le résultat d'une recherche non guidée par l'hypothèse et rendue possible par la curiosité de chercheurs face à un élément biologique nouveau. Ces nouveaux naturalistes utilisent la bioinformatique pour puiser dans les données de génomique microbienne à la découverte d'espèces et de fonctions biologiques insoupçonnées.

## Références

- Abby, S.S., Neron, B., Menager, H., Touchon, M., Rocha, E.P. (2014). MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One*, 9, e110726.
- Abedon, S.T. (2011). Facilitation of CRISPR adaptation. *Bacteriophage*, 1, 179-181.
- Alkhnbashi, O.S., Shah, S.A., Garrett, R.A., Saunders, S.J., Costa, F., Backofen, R. (2016). Characterizing leader sequences of CRISPR loci. *Bioinformatics*, 32, i576i-585.
- Andersson, A.F., Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*, 320, 1047-1050.
- Barrangou, R., Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. *Nat Biotechnol*, 34, 933-941.
- Barrangou, R., Frémaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315, 1709-1712.
- Ben-Bassat, I., Chor, B. (2016) CRISPR detection from short reads using partial overlap graphs. *J Comput Biol*, 23, 461-471.
- Biswas, A., Staats, R.H., Morales, S.E., Fineran, P.C. Brown, C. M. (2016). CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC genomics*, 17, 356.
- Bland, C., Ramsey, T.L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N.C., Hugenholtz, P. (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*, 8, 209.
- Bolotin, A., Quinquis, B., Sorokin, A. Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151, 2551-2561.
- Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, 493, 429-432.
- Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321, 960-964.
- Burstein, D., Harrington, L.B., Strutt, S.C., Probst, A.J., Anantharaman, K., Thomas, B.C., Doudna, J.A. Banfield, J.F. (2017) New CRISPR-Cas systems from uncultivated microbes. *Nature*, 542, 237-241.
- Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., O'Toole, G.A. (2012). The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J Bacteriol*, 194, 5728-5738.
- Chai, G., Yu, M., Jiang, L., Duan, Y. Huang, J. (2017). HMMCAS: a web tool for the identification and domain annotations of Cas proteins. *IEEE/ACM Trans Comput Biol Bioinform*, DOI: [10.1109/TCBB.2017.2665542](https://doi.org/10.1109/TCBB.2017.2665542).
- Cui, Y., Li, Y., Gorge, O., Platonov, M.E., Yan, Y., Guo, Z., Pourcel, C., Dentovskaya, S.V., Balakhonov, S.V., Wang, X., Song, Y., Anisimov, A.P., Vergnaud, G. Yang, R. (2008). Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS One*, 3, e2652.
- Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471, 602-607.
- Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Frémaux, C., Boyaval, P., Romero, D.A., Horvath, P., Moineau, S. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol*, 190, 1390-1400.
- Edgar, R.C. (2007). PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics*, 8, 18.
- Emerson, J.B., Andrade, K., Thomas, B.C., Norman, A., Allen, E.E., Heidelberg, K.B. Banfield, J.F. (2013). Virus-host and CRISPR dynamics in Archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea*, 2013, 370871.
- Erdmann, S. Garrett, R.A. (2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol*, 85, 1044-1056.
- Filliol, I., Driscoll, J.R., van Soelingen, D., Kreiswirth, B.N., Kremer, K., Valetudie, G., Dang, D.A., Barlow, R., Banerjee, D., Bifani, P.J., Brudey, K., Cataldi, A., Cooksey, R.C., Cousins, D.V., Dale, J.W., Dellagostin, O.A., Drobniowski, F., Engelmann, G., Ferdinand, S., Gascoyne-Binzi, D., Gordon, M., Gutierrez, M.C., Haas, W.H., Heersma, H., Kassa-Kelembho, E., Ho, M.L., Makristathis, A., Mammina, C., Martin, G., Mstrom, P., Mokrousov, I., Narbonne, V., Narvskaya, O., Nastasi, A., Niobe-Eyangoh, S.N., Pape, J.W., Rasolofon-Razanamparany, V., Ridell, M., Rossetti, M.L., Stauffer, F., Suffys, P.N., Takiff, H., Texier-Maugein, J., Vincent, V., de Waard, J.H., Sola, C. Rastogi, N. (2003). Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *J Clin Microbiol*, 41, 1963-1970.
- Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Frémaux, C., Horvath, P., Magadan, A.H. Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468, 67-71.
- Godde, J.S. Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol*, 62, 718-729.
- Grissa, I., Vergnaud, G., Pourcel, C. (2007a). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, 8, 172.
- Grissa, I., Vergnaud, G. Pourcel, C. (2007b). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*, 35, W52-57.
- Groenen, P.M., Bunschoten, A.E., van Soelingen, D. van Embden, J.D. (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol*, 10, 1057-1065.
- Gunderson, F.F. Cianciotto, N.P. (2013). The CRISPR-associated gene cas2 of *Legionella pneumophila* is required for intracellular infection of amoebae. *mBio*, 4, e00074-00013.
- Haft, D.H., Selengut, J., Mongodin, E.F. Nelson, K.E. (2005). A guild of 45CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*, 1, e60.
- Hargreaves, K.R., Flores, C.O., Lawley, T.D. Clokie, M.R. (2014). Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium difficile* strains and prophages target multiple phage types within this pathogen. *mBio*, 5, e01045-01013.
- Hermans, P.W., van Soelingen, D., Bik, E.M., de Haas, P.E., Dale, J.W. van Embden, J.D. (1991). Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot

- integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect Immun*, 59, 2695-2705.
- Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Frémaux, C., Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol*, 190, 1401-1412.
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol*, 169, 5429-5433.
- Jansen, R., van Embden, J.D., Gaastra, W., Schouls, L.M. (2002a). Identification of a novel family of sequence repeats among prokaryotes. *Omic*, 6, 23-33.
- Jansen, R., van Embden, J.D., Gaastra, W., Schouls, L.M. (2002b). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, 43, 1565-1575.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821.
- Koonin, E.V., Makarova, K.S. Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol*, 37, 67-78.
- Koskela, K.A., Mattinen, L., Kalin-Mänttari, L., Vergnaud G., Gorgé, O., Nikkari, S., Skurnik, M. (2015). Generation of a CRISPR database for *Yersinia pseudotuberculosis* complex and role of CRISPR-based immunity in conjugation. *Environ Microbiol*, 17, 4306-4432.
- Krupovic, M., Makarova, K.S., Forterre, P., Prangishvili, D., Koonin, E.V. (2014). Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol*, 12, 36.
- Krupovic, M., Beguin, P., Koonin, E.V. (2017). Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr Opin Microbiol*, 38, 36-43.
- Lander, E.S. (2016). The Heroes of CRISPR. *Cell*, 164, 18-28.
- Lillestol, R.K., Redder, P., Garrett, R.A., Brugger, K. (2006). A putative viral defence mechanism in archaeal cells. *Archaea*, 2, 59-72.
- Louwen, R., Horst-Kreft, D., de Boer, A.G., van der Graaf, L., de Knecht, G., Hamersma, M., Heikema, A.P., Timms, A.R., Jacobs, B.C., Wagenaar, J.A., Endtz, H.P., van der Oost, J., Wells, J.M., Nieuwenhuis, E.E., van Vliet, A.H., Willemsen, P.T., van Baarlen, P., van Belkum, A. (2013). A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barre syndrome. *Eur J Clin Microbiol Infect Dis*, 32, 207-226.
- Mak, J. (2005). RNA interference: more than a research tool in the vertebrates' adaptive immunity. *Retrovirology*, 2, 35.
- Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, 1, 7.
- Marraffini, L.A., Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, 322, 1843-1845.
- Mojica, F.J., Ferrer, C., Juez, G. Rodriguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloflex mediterranei* and *Haloflex volcanii* and could be involved in replicon partitioning. *Mol Microbiol*, 17, 85-93.
- Mojica, F.J., Diez-Villasenor, C., Soria, E. Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol*, 36, 244-246.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, 60, 174-182.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155, 733-740.
- Morange, M. (2015). What history tells us XXXVII. CRISPR-Cas: The discovery of an immune system in prokaryotes. *J Biosci*, 40, 221-223.
- Nakata, A., Amemura, M. Makino, K. (1989). Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J Bacteriol*, 171, 3553-3556.
- Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R. J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., Fraser, C.M. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, 399, 323-329.
- Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B. C., Barrangou, R. Banfield, J.F. (2015). CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *MBio*, 6, e00262-15.
- Pawluk, A., Bondy-Denomy, J., Cheung, V.H., Maxwell, K.L., Davidson, A.R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *MBio*, 5, e00896.
- Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E.J., Maxwell, K.L., Davidson, A.R. (2016). Naturally occurring off-switches for CRISPR-Cas9. *Cell*, 167, 1829-1838, e1829.
- Pourcel, C., Salvignol, G., Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151, 653-663.
- Rho, M., Wu, Y.W., Tang, H., Doak, T.G., Ye, Y. (2012). Diverse CRISPRs evolving in human microbiomes. *PLoS genetics*, 8, e1002441.
- Riehm, J.M., Vergnaud, G., Kiefer, D., Damdindorj, T., Dashdavaa, O., Khurelsukh, T., Zöller, L., Wölfel, R., Le Flèche, P., Scholz, H.C. (2012). *Yersinia pestis* lineages in Mongolia. *PLoS One*, 7, e30624.
- Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.L., Weiss, D.S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, 497, 254-257.
- Seed, K.D., Lazinski, D.W., Calderwood, S.B., Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*, 494, 489-491.
- Skenneron, C.T., Imelfort, M., Tyson, G.W. (2013). Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res*, 41, e105.
- Stern, A., Keren, L., Wurtzel, O., Amitai, G., Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet*, 26, 335-340.

- Swarts, D.C., Mosterd, C., van Passel, M.W., Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS One*, 7, e35888.
- Tang, T.H., Bachelier, J.P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J., Huttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci USA*, 99, 7536-7541.
- Tyson, G.W., Banfield, J.F. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol*, 10, 200-207.
- Vale, P.F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S. Gandon, S. (2015). Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. *Proc Biol Sci*, 282, 20151270.
- Zegans, M.E., Wagner, J.C., Cady, K.C., Murphy, D.M., Hammond, J.H., O'Toole, G.A. (2009). Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. *J Bacteriol*, 191, 210-219.
- Zhang, Y., Heidrich, N., Ampattu, B.J., Gunderson, C.W., Seifert, H.S., Schoen, C., Vogel, J., Sontheimer, E.J. (2013). Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol Cell*, 50, 488-503.

**Citation de l'article** : Pourcel, C. (2017). Historique de la découverte des systèmes CRISPR-Cas. *Biologie Aujourd'hui*, **211**, 247-254