

SONDERHEFT
2024

VBio

Verband | Biologie, Biowissenschaften
& Biomedizin in Deutschland

**MATERIAL-
FORSCHUNG**
Gesteinsbesiedelnde
Pilze

PFLANZENGENETIK
Genomsequenzen
sichtbar machen

EXPERIMENT
Pauline und die
Ausreißer

BIOLOGIE

IN UNSERER ZEIT

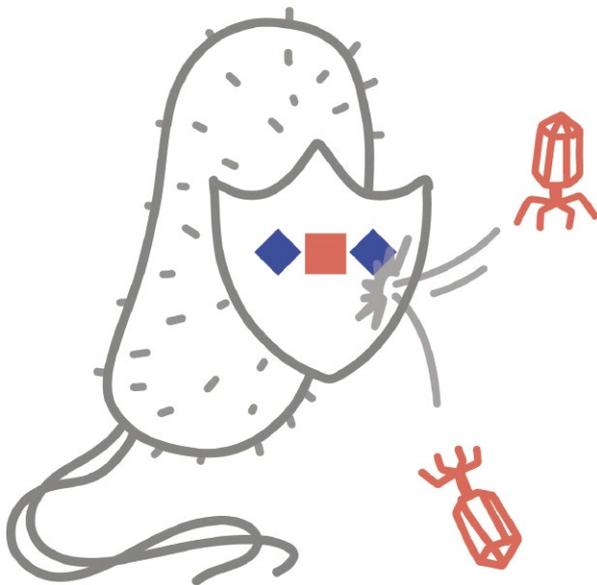
CRISPR-Cas

... mehr als nur
Verteidigung

Das Gedächtnis des bakteriellen Immunsystems

Evolution des CRISPR-Arrays in Bakterien

AXEL FEHRENBACH | FRANZ BAUMDICKER



Copyright: <https://genomeadvisory.com/resources> lizenziert unter <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

CRISPR-Cas wurde ursprünglich in Bakterien gefunden und ist primär als Verteidigungssystem gegen Invasionen von Phagen tätig. Kern dieses Systems ist – neben den cas-Genen – der CRISPR-Array, der dem System als ein vererbbares, immunologisches Gedächtnis dient. Die Evolution von CRISPR-Arrays weist besondere Merkmale auf, die Gegenstand aktueller Forschung sind und den Einsatz neuer bioinformatischer und mathematischer Methoden erfordern.

Die Grundlage für die Entwicklung der revolutionären CRISPR-Cas-Methode für die Genomeditoring wurde ursprünglich in der DNA von Bakterien und Archaeen gefunden. Bereits 1987 fanden Forscher die sich wiederholende *repeat*-Struktur des CRISPR-Arrays in *Escherichia coli* [1]: CRISPR-Arrays bestehen aus sich wiederholenden kurzen (~35 bp), palindromischen *repeats*, die durchbrochen werden durch kleine Stücke variierender Sequenzen (~35 bp), den sogenannten *spacer*. Die Funktion dieser *spacer* war lange unbekannt – bis Forscher 2005 herausfanden, dass viele dieser *spacer* der DNA von Bakteriophagen (Viren) zugeordnet werden können. Bemerkenswerterweise gelingt es Phagen anschließend nicht, Bakterien mit entsprechenden *spacer*-Einheiten zu infizieren [2].

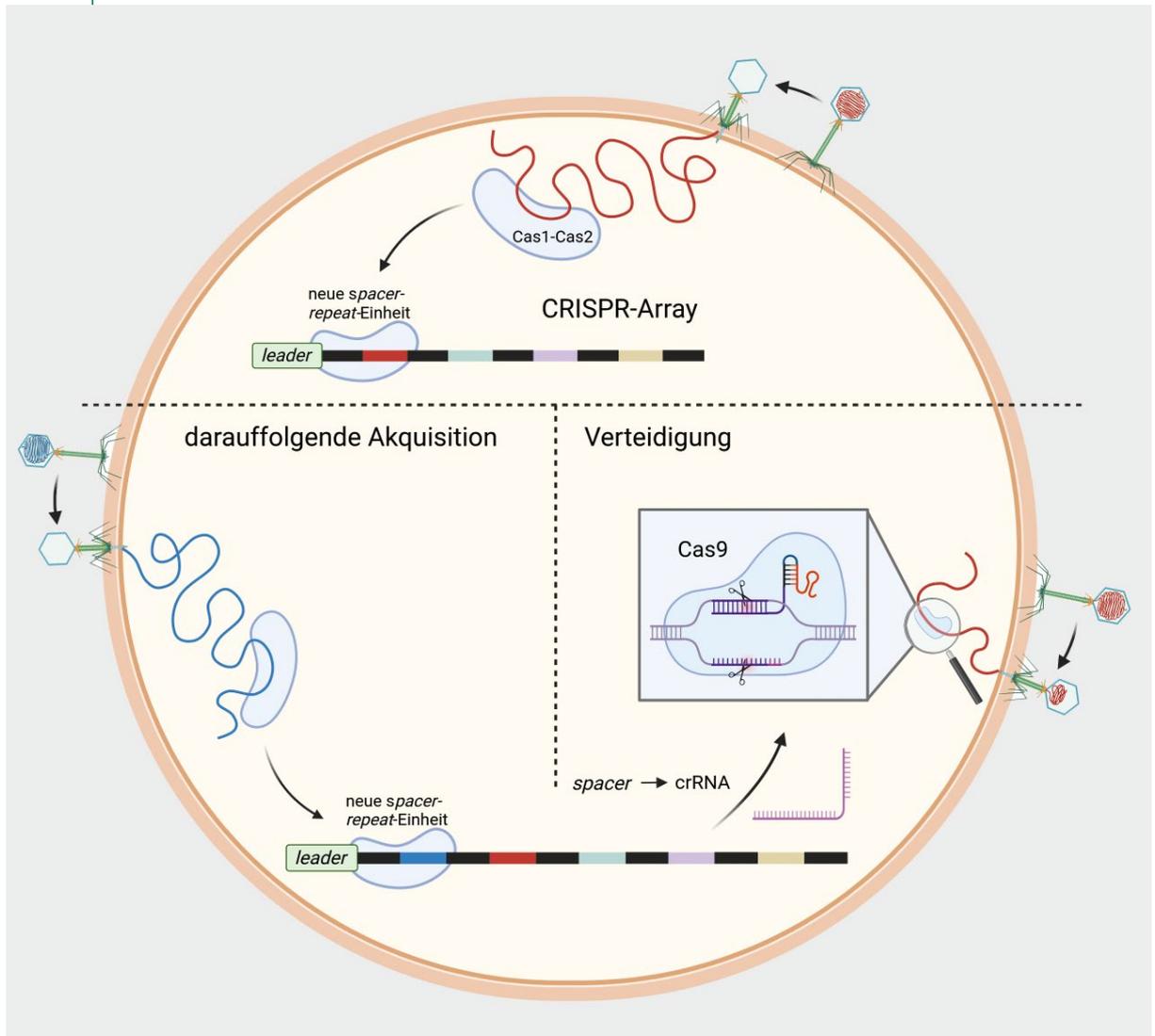
In folgenden Studien wurde klar, dass die Bakterien mithilfe von zugehörigen CRISPR-assoziierten (Cas)-Proteinen (z. B. Cas1 und Cas2) in der Lage sind, kleine Fragmente der DNA von Phagen – sogenannte *protospacer* – zu kopieren und als *spacer* zwischen *repeat*-Einheiten in den CRISPR-Array einzusetzen (Abbildung 1, links) [3]. Bei einer erneuten Infektion können diese *spacer* als sogenannte CRISPR-RNA (crRNA) repliziert werden und in Cas-

Proteinkomplexe (z. B. Cas9) eingesetzt werden. Die crRNAs dienen sozusagen als Wegweiser für die Proteinkomplexe. Diese können dadurch korrespondierende *protospacer* mit großer Präzision erkennen und – je nach CRISPR-Cas-Typ – blockieren oder zerschneiden (Abbildung 1, rechts) [4]. Das bedeutet, dass CRISPR-Cas-Systeme eine Art adaptives, antivirales Immunsystem in Bakterien sind. In diesem Sinne entspricht der CRISPR-Array einem generationenübergreifenden Gedächtnis des CRISPR-Cas-Immunsystems: Denn als Teil der DNA kann der Array relevante *spacer* über lange Zeit erhalten und an die nächsten Generationen weitergeben.

CRISPR-Arrays als chronologisches Gedächtnis

Bei genaueren Untersuchungen des Prozesses zur Aneignung (Akquisition) fremder DNA wurde eine erstaunliche Eigenschaft von CRISPR-Arrays festgestellt: Neue *spacer* werden (fast) immer an einem Ende des Arrays, neben einem DNA-Motiv – dem sogenannten *leader* – eingesetzt [4, 5]. Deshalb ist aus dem CRISPR-Array sogar eine chronologische Historie der eingefügten *spacer* bzw. der Phagen-

ABB. 1 | SPACER-AKQUISITION UND VERTEIDIGUNG GEGEN PHAGEN



Vereinfachte Darstellung eines Bakteriums mit einem CRISPR-Array, der durch Akquisition von spacer-Einheiten eine anschließende Immunreaktion gegen eine Phageninvasion ermöglicht. Alle Abbildungen erstellt mit BioRender.com.

invasionen ablesbar. Dies gibt der Evolution von CRISPR-Arrays eine bemerkenswerte Struktur, die sehr nützlich für die Modellierung und Analyse von CRISPR-Arrays ist.

Verlust von spacer-Einheiten

Die Expression von *spacer*-Einheiten im vorderen Teil des Arrays findet häufiger statt, d. h. die Effizienz des Arrays als Immunantwort ist für neu hinzugefügte *spacer* am stärksten und reduziert sich, je länger der Array wird. Darüber hinaus ist klar, dass die Kosten für den Erhalt und die Replikation des Arrays steigen, je größer der Array ist. Folglich kann ein Array zwar theoretisch beliebig viele neue *spacer* aufnehmen, der regelmäßige Verlust von *spacer*-Einheiten verhindert jedoch, dass das System an Effizienz einbüßt.

Während das Einfügen von *spacer*-Einheiten und die Funktion von (einigen) Cas-Proteinen in diesem Prozess relativ gut verstanden sind, ist der Mechanismus des „Verlusts“ (engl. *deletion*) bisher wenig untersucht. Dies liegt unter anderem daran, dass dieser Prozess im Unterschied zur aktiven Initiation einer *spacer*-Akquisition nur indirekt zu beobachten ist. Dennoch gibt es bereits Theorien da-

IN KÜRZE

- CRISPR-Cas-Systeme dienen Bakterien als **Immunsysteme**.
- Der **CRISPR-Array ist dabei das vererbte Gedächtnis** dieses Immunsystems und erlaubt Rückschlüsse auf die zeitliche Abfolge von Phageninfektionen.
- Dieser Array evolviert über die **Akquisition und den Verlust** von sogenannten *spacer*-Einheiten.
- Durch **mathematische Modelle** ist es uns möglich, die CRISPR-Arrays von **gemeinsamen Vorfahren einer Population** von Bakterien zu rekonstruieren.
- Die Erforschung dieser Ahnenrekonstruktionen erlaubt uns, **neue Erkenntnisse über den Verlustprozess** zu gewinnen und die Umgebung von Bakterien zu erforschen.

rüber, wie Verluste ausgelöst werden. Es wird vermutet, dass die sich wiederholende *repeat*-Struktur des CRISPR-Arrays verantwortlich ist. Während der Replikation des CRISPR-Arrays können *repeats* aufgrund ihrer großen Ähnlichkeit „aneinanderkleben“, wodurch eine Schleife entstehen kann, die *spacer* (und weitere *repeats*) enthält (Abbildung 2) [5]. Wenn nun die DNA kopiert wird, kann der Inhalt dieser Schleifen übersehen werden und in der Folge entsteht ein Strang ohne die enthaltenen *spacer*. Bestimmten Cas-Proteinen konnte bisher noch keine Rolle in diesem Verlustprozess zugeordnet werden.

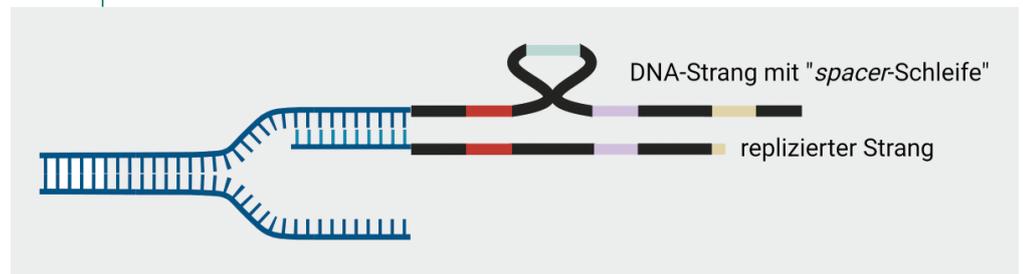
Ahnenrekonstruktion von CRISPR-Arrays

Der besondere Einfügungsprozess und Verlustprozess von *spacer*-Einheiten motivierte uns dazu, mathematische Modelle zu erstellen, um die besonderen Auswirkungen dieser Prozesse zu verstehen und die beobachtete Struktur von *spacer*-Arrays besser interpretieren zu können [6]. Speziell interessiert uns, die historischen Veränderungen nah verwandter CRISPR-Arrays zu rekonstruieren – d. h. entlang des Stammbaums die Akquisitions- und Verlustereignisse zu rekonstruieren, die die derzeitigen CRISPR-Arrays gebildet haben (Abbildung 3). Diese, beziehungsweise ihre Stammbäume, stammen oft aus Laborversuchen, d. h. sie sind in der Regel in einem Zeitraum von Tagen, Wochen und Monaten entstanden. Aber es gibt auch Datensätze, bei denen die CRISPR-Arrays im Abstand von mehreren Jahren sequenziert wurden.

Um die Evolution zu modellieren, nutzen wir stochastische Prozesse, sogenannte „stetige Markovprozesse“. Ein stochastischer Prozess modelliert ein System (in diesem Fall den *spacer*-Array), das sich in verschiedenen Zuständen befinden kann und mit gewissen Wahrscheinlichkeiten zwischen den Zuständen wechselt. Ein Markovprozess ist dabei ein spezieller stochastischer Prozess, bei dem die Wahrscheinlichkeiten für den nächsten Zustand des Systems/Arrays nur von einer begrenzten Vorgeschichte (in unserem Fall dem aktuellen Zustand) abhängt. Dadurch wird für Prognosen über zukünftige Zustände kein Wissen über die gesamte Vorgeschichte des Prozesses benötigt. Bei einem „stetigen“ Markovprozess ändern sich die Zustände nicht zu festen, sondern zu zufälligen Zeitpunkten, d. h. die Übergänge zwischen den Zuständen können sich ständig ereignen. Die Häufigkeit der Übergänge wird dabei durch sogenannte „Raten“ beschrieben.

Zur Vereinfachung beginnen wir mit einem relativ simplen Evolutionsmodell. In unserem Modell für CRISPR-Arrays können *spacer* mit einer gewissen „Einfügensrate“ überall (zwischen jedem bestehenden *spacer*) gewonnen werden und existierende *spacer* können mit einer anderen „Verlustrate“ verloren werden. Allerdings be-

ABB. 2 | SPACER-VERLUSTE WÄHREND DER DNA-REPLIKATION



Schematische Darstellung eines *spacer*-Verlusts durch „Schleifenbildung“ während der Replikation der DNA.

rücksichtigt dieses Modell die Akquisition am *leader* nicht. Entlang des Stammbaums kann mit diesem Modell effizient zurückgerechnet werden, an welchem Ast mit welcher Wahrscheinlichkeit *spacer* gewonnen und verloren werden. Mit einer sogenannten Maximum-Likelihood-Methode können wir so die wahrscheinlichste Konfiguration von Ereignissen rekonstruieren, die zu der beobachteten Population geführt haben.

Solche Methoden – Markovprozesse und Maximum-Likelihood-Methoden – haben eine lange Tradition in der Rekonstruktion von Stammbäumen und werden für DNA-Sequenzen häufig genutzt.

Verlangen wir von dem Modell, dass *spacer* nur am *leader* eingefügt werden können, dann stoßen diese Methoden jedoch an ihre Grenzen, da die Reihenfolge der Ereignisse, die den Array ändern, zeitlich voneinander abhängt. Genauer: Wenn ein *spacer* zu einem Zeitpunkt existiert, können danach nur *spacer* gewonnen werden, die vor dem existierenden *spacer* im Array stehen. Andererseits können alle *spacer* hinter dem existierenden *spacer* nur verloren werden. Wenn man in der Zeit zurückrechnet, um die Ahnen-Arrays zu rekonstruieren, erzeugen diese Abhängigkeiten eine nicht (effizient) auflösbare Komplexität von möglichen Abfolgen von Akquisitions- und Verlustereignissen.

Es gibt jedoch eine Möglichkeit, die bisherige – mit dem einfachen Modell erfolgte Rekonstruktion – zu verbessern, indem wir die rekonstruierten Ereignisse und Wahrscheinlichkeiten so modifizieren, dass die Teile der Akquisitionsreihenfolge, die wir aus den CRISPR-Arrays selbst ablesen können, zu jedem Zeitpunkt berücksichtigt werden. Nun sind wir mit einem Algorithmus [6] ausgerüstet, der uns erlaubt, aus verwandten CRISPR-Arrays, die Array-Ahnen oder – äquivalent dazu – die evolutionären Ereignisse, die zu den beobachteten Arrays geführt haben, zu rekonstruieren.

Mit „Ahnenforschung“ den Verlustprozess verstehen

Die rekonstruierten Ahnen-Arrays ermöglichen uns nun, die Evolution genauer zu verstehen und unsere evolutionären Modelle zu verbessern sowie besondere Muster in den beobachteten Daten zu erkennen [6]. Dabei konzentrieren wir uns im Folgenden auf den Verlustprozess.

Verlustraten und Blockverluste

Eine Facette ist die Möglichkeit, dass mehrere benachbarte *spacer* und *repeats* in Blöcken verlorengehen. Weil beide relativ kurz sind, ist es möglich, dass *repeats*, die weiter als ein *spacer* voneinander entfernt liegen, „zusammenkleben“. Zusammen bilden sie somit eine Schleife, die mehrere *spacer* (und *repeats*) enthält. Um dies zu untersuchen, haben wir einen Markovprozess entwickelt, der genau diesen Umstand modelliert [6]. Sobald ein Verlust stattfindet, wird nicht nur ein *spacer*, sondern ein ganzer Block verloren, dessen Länge nach einer (geometrischen) Wahrscheinlichkeitsverteilung verteilt ist. Dieses Modell hat nun neben den Gewinn- und Verlustraten einen weiteren Parameter: die durchschnittliche Länge von Verlusten. Mithilfe eines statistischen (Likelihood-Quotienten-)Tests können wir nun das Modell mit einzelnen Verlusten mit dem Modell mit Blockverlusten vergleichen. Wir tun dies anhand der Ahnenrekonstruktionen, die wir bestimmt haben. Genauer gesagt: Wir tun dies anhand der rekonstruierten Verlustereignisse, deren Häufigkeit und deren Länge.

In einer Studie über einen großen Datensatz von CRISPR-Arrays waren wir in der Lage, das Modell der einzelnen Verluste in einer signifikanten Anzahl von Fällen zu verwerfen [6]. Daher ist es sehr wahrscheinlich, dass bei den meisten CRISPR-Arrays Blockverluste auftreten. Darüber hinaus ermöglicht uns das Modell, die Verlustrate sowie deren durchschnittliche Länge zu schätzen. Diese Parameter können dann mit der durchschnittlichen Mutationsrate verglichen werden, um vorherzusagen, wie schnell CRISPR-Arrays evolvieren.

Wir fanden heraus, dass die durchschnittliche Verlustrate pro *spacer* ca. 370-mal der Mutationsrate pro Nukleotid entspricht. Das bedeutet, dass CRISPR-Arrays deutlich schneller durch Verluste evolvieren als durch Mutationen. Daraus resultiert auch gleich ein Anwendungsfall für die

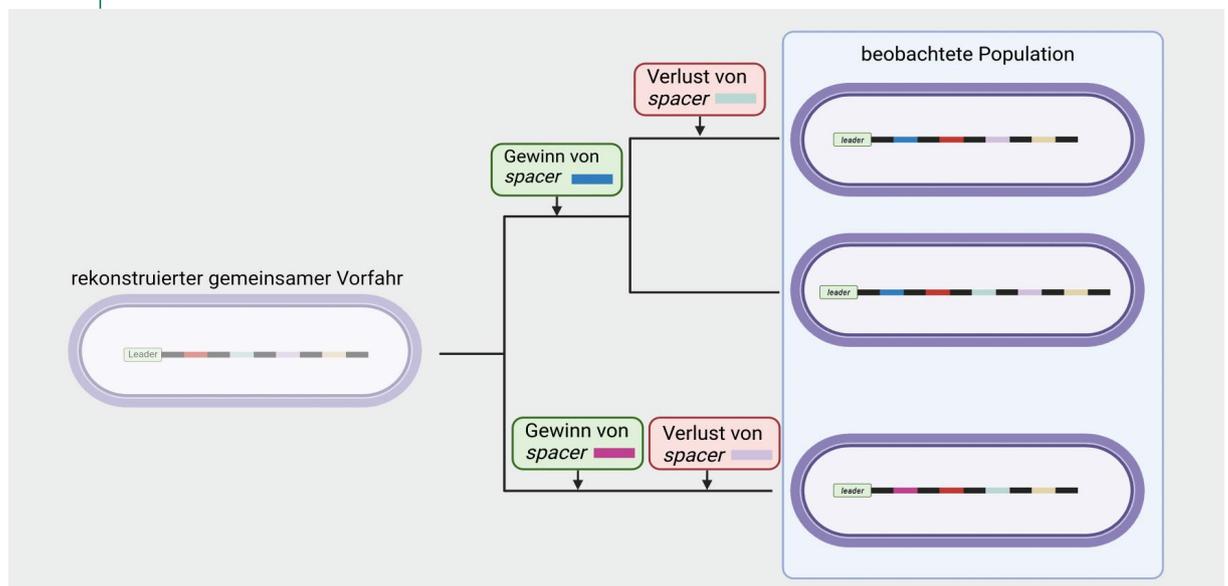
Ahnenrekonstruktion von CRISPR-Arrays, nämlich das Verfolgen der Verwandtschaft von untersuchten Bakterien anhand deren CRISPR-Arrays mit einer höheren Auflösung als anhand von Mutationsmustern, z. B. in einem epidemiologischen Kontext. Darüber hinaus stellten wir fest, dass die Verlustparameter unabhängig sind von den in der Nähe der Arrays gefundenen *cas*-Genen. Das unterstützt die bisherigen experimentellen Beobachtungen, dass *cas*-Genen noch keine Rolle bei der Entstehung von Verlusten zugeordnet werden konnte.

Verlustmechanismus und Selektion

Die Rekonstruktionen erlauben auch, genauer zu untersuchen, wo Verluste im Array auftreten [6]. Es lässt sich beobachten, dass Verluste weniger oft an den Rändern des Arrays auftreten. Dies passt zum vermuteten *repeat*-basierten Verlustmechanismus durch Schleifenbildung. Denn *repeats* am Rande des Arrays können nur mit weniger anderen *repeats* eine Schleife bilden, da auf einer Seite weniger (oder keine) Partner vorhanden sind. Dies ist ein symmetrischer Effekt, der also sowohl für den Anfang des Arrays (beim *leader*) als auch für das Ende des Arrays gilt.

Überraschender ist ein anderer Effekt, den wir beobachten. Es stellt sich heraus, dass keine komplette Symmetrie in diesem Effekt zu beobachten ist. Am Beginn des Arrays werden *spacer* öfter verloren als erwartet oder alternativ: Alte *spacer* am Ende des Arrays werden weniger oft verloren. Das deutet darauf hin, dass junge neu gewonnene *spacer* öfter verloren werden. Vorstellbar wäre hier, dass die ersten *spacer* im Array aufgrund der höheren Expressionsrate solch eine Bedeutung für die akute Verteidigung haben, dass ein Verlust von im Moment unwichtigen *spacer*-Einheiten an dieser Stelle im Array besonders vorteilhaft ist, damit der Array gegen die akut relevanten Bedrohungen effizient und aktiv bleibt.

ABB. 3 | AHNENREKONSTRUKTION VON CRISPR-ARRAYS



Beispiel für die Rekonstruktion von *spacer*-Akquisitions- und Verlustereignissen entlang eines Stammbaumes für eine beobachtete Population von CRISPR-Array tragenden Bakterien.

Anwendungen und Zukunft

Wir haben bereits einen möglichen Nutzungsfall für CRISPR-Array-Evolution beschrieben. Durch die schnelle Evolution der CRISPR-Arrays und die relative Einfachheit, ebendiese Arrays zu sequenzieren, könnten CRISPR-Arrays es ermöglichen, die Evolution von pathogenen Bakterien genauer zu verfolgen. Dies könnte – je nach Aktivität der CRISPR-Arrays – auch in einem epidemiologischen Kontext (über Wochen, Monate und Jahre) stattfinden.

Durch die *spacer*-Phagen-Beziehung besteht außerdem die Möglichkeit, die Umgebung von Bakterien und deren Interaktionen besser zu verstehen. Neben der Untersuchung der Phagenbiodiversität in natürlichen Ökosystemen könnte dies insbesondere im Kontext von Bioreaktoren auch einen wirtschaftlichen Wert haben. Denn Bioreaktoren zur Herstellung von wichtigen Grundstoffen für pharmakologische Produkte mit Bakterien können zum Beispiel durch Virenbefall kollabieren. Anhand der CRISPR-Arrays könnten diese geschlossenen Systeme genauer analysiert werden. Beispielsweise könnte durch wiederholte Probeentnahme und Analyse der CRISPR-Arrays untersucht werden, wann und wie stark Bakteriophagen Druck auf die Bakterienpopulation ausüben. Nach einem Kollaps könnte außerdem durch Analyse der *spacer* der überlebenden Bakterien festgestellt werden, welche Phagen für den Kollaps verantwortlich waren. In den überlebenden Bakterien wird man mit hoher Wahrscheinlichkeit *spacer* gegen genau diese Phagen finden.

Nicht zuletzt ist das Verstehen von CRISPR-Cas als Verteidigungsmechanismus von Bakterien und Archaeen sowie deren Phagen ein faszinierendes Forschungsprojekt, das aufzeigt, wie oft hochkomplexe Immunprozesse schon in den kleinsten Lebewesen stattfinden.

Zusammenfassung

Bakterien beherbergen ein erstaunliches Immunsystem, das mit einem vererbaren Gedächtnis, dem CRISPR-Array ausgerüstet ist. Dieser Array gewinnt neue Sequenzen, sogenannte *spacer*, die mit Sequenzen invasiver Phagen korrespondieren. Deshalb erlaubt der Array Rückschlüsse auf die zeitliche Abfolge von Phageninfektionen. Da der Array als Teil der DNA nicht unbegrenzt wachsen kann, müssen auch wieder *spacer* verlorengehen. Während der Mechanismus des *spacer*-Erwerbs recht gut verstanden ist, weiß man wenig darüber, wie *spacer* eliminiert werden. Die Untersuchung des Arrays erfordert jedoch neue mathematische Modelle, die Hinweise darauf geben, wie diese Vorgänge stattfinden.

Summary

The evolution of the CRISPR array in bacteria

Bacteria harbour a remarkable immune system with an inheritable memory known as the CRISPR array as part of their DNA. This array acquires new segments - so called "spacers" – which correspond to segments of invasive phages. Thus, the array allows to infer a chronological sequence of infec-

tions by bacteriophages. However, being a part of bacterial DNA, the array cannot grow indefinitely and therefore spacers must get lost again. Whereas the mechanism of spacer acquisition is fairly well understood, little is known about how spacers are deleted. However, studying the array requires new mathematical models that provide new clues as to how these processes take place.

Schlagworte:

Evolution, CRISPR-Array, *leader*, *repeats*, *spacer*, Bakterien, Phagen, bakterielles Immunsystem, mathematische Modelle, Stammbaum

Literatur

- [1] Y. Ishino et al. (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(12), 5429–33, <https://doi.org/10.1128/jb.169.12.5429-5433>
- [2] F. J. Mojica et al. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60(2), 174–82, <https://doi.org/10.1007/s00239-004-0046-3>
- [3] J. K. Nuñez et al. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* 21(6), 528–34, <https://doi.org/10.1038/nsmb.2820>
- [4] R. Barrangou et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819), 1709–12, <https://doi.org/10.1126/science.1138140>
- [5] S. C. Garrett (2021). Pruning and Tending Immune Memories: Spacer Dynamics in the CRISPR Array. *Front Microbiol* 12, 664299, <https://doi.org/10.3389/fmicb.2021.664299>
- [6] A. Fehrenbach et al. (2024). SpacerPlacer: Ancestral reconstruction of CRISPR arrays reveals the evolutionary dynamics of spacer deletions. *Nucleic Acids Research*, gkae772, <https://doi.org/10.1093/nar/gkae772>

Verfasst von:



Axel Fehrenbach ist Doktorand im Cluster of Excellence „Controlling Microbes to Fight Infections“ an der Universität Tübingen. Im Zuge seiner Promotion beschäftigt er sich mit der Evolution des CRISPR-Arrays als Teil des bakteriellen Immunsystems. Dafür nutzt er mathematische Modelle sowie computergestützte Simulationen und Analysen von großen Datensätzen.



Dr. Franz Baumdicker ist Gruppenleiter im Cluster of Excellence „Controlling Microbes to Fight Infections“ an der Universität Tübingen. Seine Forschung beschäftigt sich mit der Diversität innerhalb bakterieller Populationen. Um die Dynamiken der bakteriellen Evolution zu entschlüsseln, entwickelt er mathematische Modelle und machine learning-Algorithmen für die Analyse großer genetischer Datensätze.

Korrespondenz

Axel Fehrenbach
Mathematical and Computational Population Genetics
Sand 14
72076 Tübingen
E-Mail: axel.fehrenbach@uni-tuebingen.de
franz.baumdicker@uni-tuebingen.de



Verband | Biologie, Biowissenschaften
& Biomedizin in Deutschland

**GEMEINSAM
FÜR DIE**

BIEWISSENSCHAFTEN

Gute Gründe, dem VBIO beizutreten:

- Werden Sie Teil des größten Netzwerks von Biowissenschaftlern in Deutschland.
- Unterstützen Sie uns, die Interessen der Biowissenschaften zu vertreten.
- Nutzen Sie Vorteile im Beruf.
- Bleiben Sie auf dem Laufenden – mit dem VBIO-Newsletter und dem Verbandsjournal „Biologie in unserer Zeit“.
- Treten Sie ein für die Zukunft der Biologie.



www.vbio.de

Jetzt beitreten!

