

CRISPR RNAs Guide  
of Defense in Prokaryotes

Multiplex Genome Engineering  
Using CRISPR/Cas Systems

Le Cong<sup>1,2\*</sup>, F. Ann Rao<sup>1,2\*</sup>, David Cox<sup>1,2</sup>, Shaocong Liu<sup>1,2</sup>, Robert Bortotto<sup>1</sup>, Nabeel Rizvi,  
Patrick D. Han<sup>1</sup>, Xuebing Wu<sup>1</sup>, Wenyao Jiang<sup>1</sup>, Luciano A. Marandini<sup>1</sup>, Feng Zhang<sup>1</sup>

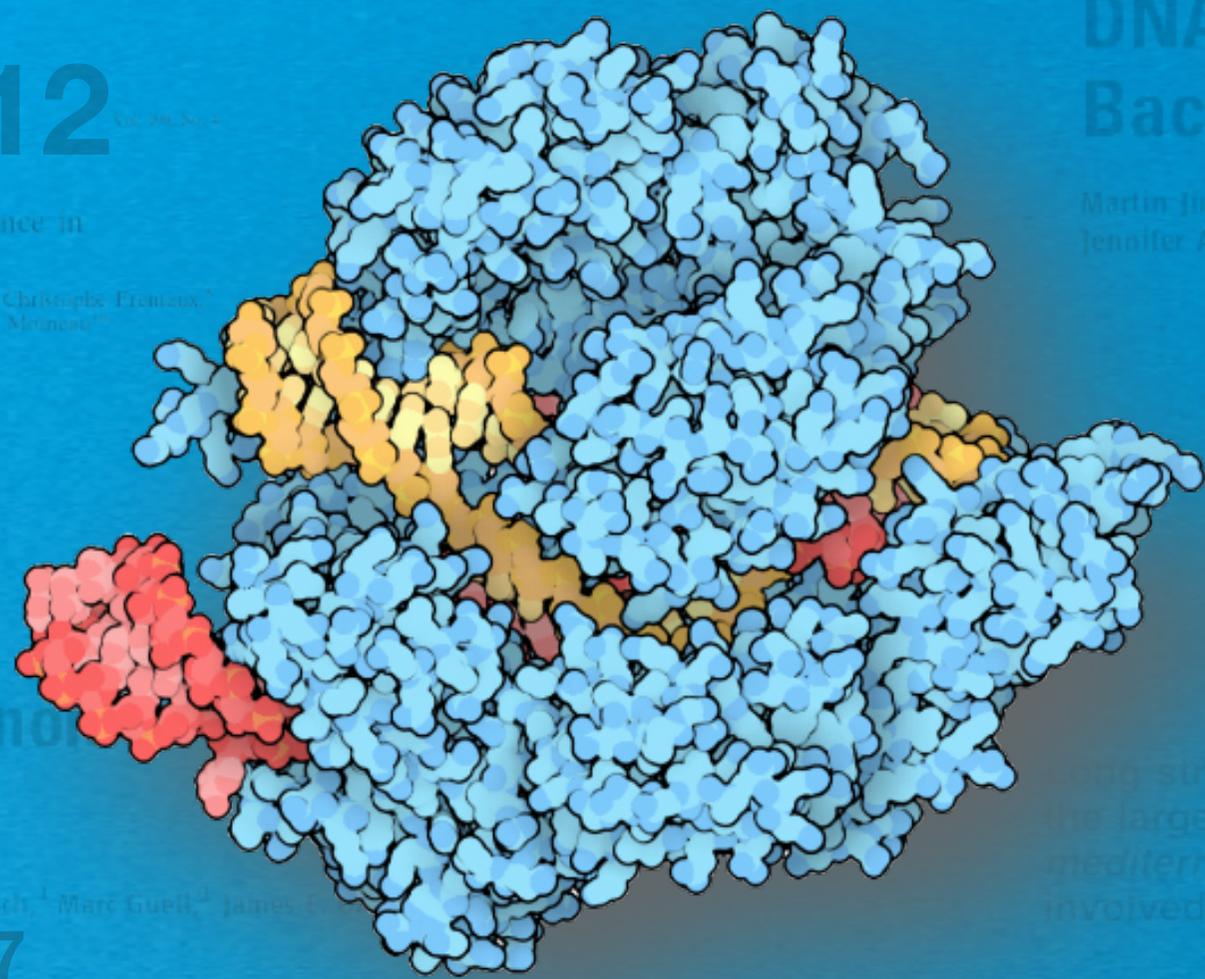
Matthijs Al. Jore,<sup>1,2</sup> Magnus Lundgren,<sup>1</sup> Edze R. Westra,<sup>1</sup>  
Ambrasius P. L. Snijders,<sup>2</sup> Mark J. Dickman,<sup>2</sup> Kira S. Makarova,<sup>1</sup>  
John van der Oost<sup>1</sup>

A Program  
DNA End  
Bacterial

2012

Encoded Resistance in  
*S. aureus*  
Jessica Labonte,<sup>1</sup> Christophe Erenaux,<sup>1</sup>  
Györgyi Horvath,<sup>1</sup> and Sylvain Bréchet<sup>1,2</sup>

Martin Jinek,<sup>1,2\*</sup> Krzysztof  
Jennifer A. Doudna,<sup>1,2,3,4\*</sup>



Genomic  
and  
an Genom  
as9  
svelt,<sup>1</sup> John Aach,<sup>1</sup> Marc Guell,<sup>1</sup> James F.  
2007

2005

stretches of  
the largest replicon  
*mediterranei* and *T.*  
involved in replicon

# #CRISPRhistory

erzählt von Marcus Anhäuser

# Ruud Jansen und die Entdeckung von Cas

25. April 2022

**J**an van Embden ist Mitte der 1990er Jahre überzeugt, dass dieses eigenartige DNA-Muster aus *repeat-* und *spacer*-Sequenzen, das mehr als zwei Jahrzehnte später als CRISPR weltberühmt werden wird, etwas zu bedeuten hat. Seit Beginn der neunziger Jahre hat sich der niederländische Biochemiker und sein Team am staatlichen *Rijksinstituut voor Volksgezondheid en Milieu* (RIVM) in Bilthoven bei Utrecht nun schon damit befasst (siehe das Kapitel „[CRISPR, Tuberkulose und die Suche des Jan van Embden](#)“). Sie wollen das DNA-Muster im Erbgut des Tuberkulose-Erregers *Mycobacterium tuberculosis*, das weltweit fast

niemand kennt, als Werkzeug im Kampf gegen die tödliche Krankheit nutzen.

Eine solche Fragestellung drängt einem neugierigen Wissenschaftler wie van Embden zwangsläufig auch grundsätzlichere Fragen auf: Wieso gibt es solch ein ungewöhnliches DNA-Muster im Erbgut dieser Bakterien überhaupt? Woher kommt es? Wie entstand es? Was hat sich die Natur dabei gedacht?

Auch wenn es sonst an van Embdens Institut um ganz praktische Anwendungen für die öffentliche Gesundheit geht, hat der erfahrene Niederländer schon immer ein großes Interesse an solchen grundlegenden Fragen, insbesondere zur Evolution. Er ist fasziniert von der Einzigartigkeit der neu entdeckten Struktur: „Diese Sequenzen unterschieden sich von all den anderen bekannten sich wiederholenden DNA-Elementen, da stellte sich einfach die Frage nach der biologischen Bedeutung“, schreibt mir van Embden heute, rund 25 Jahre später, als er mit seiner Frau in Coronazeiten in der Normandie im Sommerhaus festsetzt.

In den 1980er Jahren hatten DNA-Stücke, die sich im Erbgut  $x$ -fach wiederholen, noch den Nimbus von nutzlosem genetischem Müll, der sich im Laufe der Jahrmillionen einfach so ansammelt. Erst allmählich begreifen die Biologinnen weltweit, dass solche Muster mit Funktionen verbunden sein können.

Um der Funktion dieses Musters auf die Spur zu kommen, startet van Embden 1995 ein Forschungsprojekt und macht sich auf die Suche nach der Verbreitung des prägnanten *repeat-spacer*-Musters. Wenn es etwas



*Ruud Jansen war um die Jahrtausendwende an der Universität Utrecht angestellt, arbeitete aber für die Arbeitsgruppe von Jan van Embden, der in Bilthoven saß (erstellt mit Datawrapper).*

Bedeutendes ist, so seine These, sollte es sich auch im Erbgut anderer Lebewesen finden lassen.

Van Embden setzt zuerst einen Studenten von der benachbarten Universität Utrecht auf das Thema an. Er soll die öffentlich zugänglichen Gen-Datenbanken nach dem Muster durchsuchen. „Das war keine einfache Aufgabe“, schreibt mir der heute 77-jährige Biologe. Nur allmählich füllen sich damals die Datenbanken wie

die US-amerikanische *GenBank* oder die europäische *EMBL Nucleotide Sequence Databank* mit DNA-Sequenzen. 1995 wird zum ersten Mal überhaupt das Erbgut eines Lebewesens vollständig sequenziert: das Genom des Bakteriums *Haemophilus influenzae*, dessen Name auf den Umstand hinweist, dass man es einst irrtümlich für den Erreger der Grippe gehalten hatte.

Van Embdens Student bricht das Projekt indes nach fast zwei Jahren ab; warum auch immer, der Niederländer erinnert sich heute nicht mehr an den Grund.

Man kann in den Datenbanken nicht einfach einen Filter auswählen, der einem alle „CRISPR“-Sequenzen auswirft. Der spanische CRISPR-Pionier Francisco Mojica durchsucht in diesen Jahren – parallel zur Arbeit der Niederländer – noch händisch die Fachzeitschriften der kleinen Bibliothek seines Instituts an der Universität Alicante nach diesem eigenartigen repeat-Muster (siehe Kapitel [„Francisco Mojicas ‚Salt Lovers‘ und der Beginn der CRISPR-Forschung \(Teil 3\)“](#)). Später schreibt ihm der Student César Díez-Villaseñor ein Programm, das ähnliche Sequenzmuster per Computer findet. Es dauert Jahre bis Mojica eine Liste von 19 Einzellern zusammen hat, bei denen er das prägnante Muster im Erbgut nachweisen kann.

Vielleicht ist diese Aufgabe für van Embdens jungen Mitarbeiter schlicht zu anspruchsvoll, vielleicht hat er andere Gründe. Auf jeden Fall braucht van Embden nach dessen Ausscheiden eine erfahrenere Person für den Job.

An dieser Stelle kommt Ruud Jansen ins Spiel.

Molecular Microbiology (2002) 43(6), 1565–1575

**Identification of genes that are associated with DNA repeats in prokaryotes**Ruud Jansen,<sup>1</sup> Jan D. A. van Embden,<sup>2</sup>Wim Gaastra<sup>1</sup> and Leo M. Schouls<sup>1</sup>  
<sup>1</sup>Department of Infectious Diseases and Immunology,  
Bacteriology Division, Veterinary Faculty,  
Utrecht University, Yalelaan 1, 3584 CL Utrecht,  
The Netherlands;<sup>2</sup>Laboratory of Bacteriology of the Research Laboratory  
of Infectious Diseases, National Institute of Public  
Health and Environmental Protection,  
A. van Leeuwenhoeklaan 1, 3720 BA Bilthoven,  
The Netherlands.**Summary**

Using *in silico* analysis we studied a novel family of repetitive DNA sequences that is present among both domains of the prokaryotes (Archaea and Bacteria), but absent from eukaryotes or viruses. This family is characterized by direct repeats, varying in size from 21 to 37 bp, interspaced by similarly sized non-repetitive sequences. To appreciate their characteristic structure, we will refer to this family as the clustered regularly interspaced short palindromic repeats (CRISPR). In most species with two or more CRISPR loci, these loci were flanked on one side by a common leader sequence of 300–500 bp. The direct repeats and the leader sequences were conserved within a species, but dissimilar between species. The presence of multiple chromosomal CRISPR loci suggests that CRISPRs are mobile elements. Four CRISPR-associated (*cas*) genes were identified in CRISPR-containing prokaryotes that were absent from CRISPR-negative prokaryotes. The *cas* genes were invariably located adjacent to a CRISPR locus, indicating that the *cas* genes and CRISPR loci have a functional relationship. The *cas2* gene showed motifs characteristic for helicases of the superfamily 2, and the *cas1* gene showed motifs of the RecB family of exonucleases, suggesting that these genes are involved in DNA metabolism or gene expression. The spatial coherence of CRISPR and *cas* genes may stimulate new research on the genesis and biological role of these repeats and genes.

Accepted 10 December 2001. For correspondence, E-mail: R.jansen@vet.uu.nl; Tel: (+31) 30 253 4791; Fax: (+31) 30 254 0784.  
© 2002 Blackwell Science Ltd

**Introduction**

Repetitive sequences are common in the genomes of prokaryotic organisms, and their identification is increasingly facilitated by the availability of sequences of complete bacterial genomes. The length, sequence and position of these sequences in the genome are highly variable and often unique for a single strain (Lippski *et al.*, 1996; Bacheller *et al.*, 1997; van Belkum *et al.*, 1998). Two main classes of short sequence repeats (SSRs) can be distinguished, contiguous repeats and interspersed repeats (van Belkum *et al.*, 1998). The number of units of contiguous repeats usually varies from strain to strain, and this genetic heterogeneity in bacterial populations may lead to phenotypic differences due to differential gene transcription or translation. These SSRs are often part of open reading frames or promoter regions, and variation in the number of repeat units may lead to the switch of expression of surface-exposed components (Dybvig, 1993; Belland *et al.*, 1997; van Belkum *et al.*, 1998). Thus, the SSR-mediated variation is thought to provide a bacterial population with the diversity needed to adapt to changing environments.

The group of interspersed SSRs is also commonly present in bacteria. Their unit length is usually less than 200 bp; they are non-coding, interstronic and widely dispersed throughout the genome. Many interspersed SSRs have been disclosed in bacterial species. Examples are the REP and ERIC sequences of the Enterobacteriaceae, the BOX element in *Streptococcus pneumoniae* and the SSRs in *Neisseria gonorrhoeae* and *Haemophilus influenzae* (Correa *et al.*, 1988; Hulton *et al.*, 1991; Martin *et al.*, 1992; Fleischmann *et al.*, 1995). The SSR sequences of different bacterial phyla differ considerably in sequence, and consensus sequences differ from genus to genus. These repeat structures have been implicated in mRNA stabilization, transcription termination and genetic rearrangements. In *H. influenzae*, the SSRs are involved in the DNA exchange by transformation (Fleischmann *et al.*, 1995).

A distinct class of interspersed SSRs was recognized in 1987 in *E. coli* K12 (Ishino *et al.*, 1987; Nakata *et al.*, 1989), and later in other bacterial and archaeal species such as *Halobacterium mediterraneum*, *Streptococcus pyogenes*, *Anabaena* sp. PCC 7120 and *Mycobacterium tuberculosis* (Groenen *et al.*, 1993; Mojca *et al.*, 1995; Massipoli *et al.*, 1996; Hise *et al.*, 1999).



*Ruud Jansen, damals an der Universität Utrecht, hat die ersten vier Cas-Gene entdeckt und die Namen 'CRISPR' und 'Cas' in seinem Artikel 2002 im Fachmagazin 'Molecular Microbiology' eingeführt (links: Titelseite des Artikels). Foto: privat*

## In silico: Die Suche im virtuellen Datenraum des Lebens

Jansen, Jahrgang 1957, kommt von der Küste aus dem kleinen Städtchen Katwijk, das für seine Künstlerinnenkolonie von Impressionisten Anfang des 20. Jahrhunderts und seinen breiten Strand bekannt ist. In den 1980er Jahren studiert Jansen Biologie an der Universität von Utrecht und erstellt dort bis 1994 seine Doktorarbeit über genetische Grundlagen eines Bakte-

riengifts, das eine schwere Lungenkrankheit bei Schweinen auslöst.

Dann wechselt er ans Agrarforschungszentrum in Wageningen, wo er sich am Institut für Pflanzenkrankheiten zwei Jahre lang auf die Suche nach Genen macht, die aktiv sind, wenn ein krankmachender Pilz sich über Nutzpflanzen hermacht. Weil das Projekt nicht weiterfinanziert wird, muss er sich nach neuen Möglichkeiten umschauen. Typisches Leben eines Post-Docs. So führt ihn sein Weg Ende '96 zurück an die Universität Utrecht.

Am dortigen Department für Infektionserkrankungen und Immunologie der veterinärmedizinischen Fakultät übernimmt er gleich drei Projekte. Eines davon ist eine Kooperation mit Jan van Embdens Arbeitsgruppe in Bilthoven: die Suche nach dem „eigenartigen DNA-Muster“ aus *repeat*-Sequenzen. Jansen, fast vierzig, hat definitiv mehr Erfahrung als der Student und er hat längst bewiesen, dass er Projekte zu Ende führen kann.

Van Embdens Auftrag besteht aus zwei Teilen. Zum einen klassische Laborarbeit mit Bakterien. Das kennt Jansen von seinen früheren Jobs. Er will versuchen, etwas über eine mögliche Funktion der *repeat*-Sequenzen herauszufinden. Dazu entfernt er sie aus dem Erbgut der Tuberkulosebakterien, um irgendeine Reaktion zu provozieren. Biologen nennen das Knock-out-Experimente. Das ist ein etwas grobes, aber gängiges Verfahren nach dem Motto: „Willst Du wissen, was es macht, hau' es einfach raus!“ Doch es passiert –

rein gar nichts: keine Verzögerung beim Wachstum, keine Änderung im Aussehen der Zellen oder was auch immer. Die Einzeller zeigen sich völlig unbeeindruckt. Die Funktion des *repeat-spacer*-Musters lässt sich auf diese Weise offenbar nicht herausfinden.

Der zweite Teil des Projekts ist wesentlich erfolgreicher, auch wenn es für Jansen weitgehendes Neuland bedeutet. Mit Fleiß und etwas Glück gelangen ihm entscheidende Entdeckungen.

Es ist die reine Bioinformatik – keine Laborarbeit mit Pipetten, Zentrifugen, Petrischalen oder lebenden Objekten. Diese Aufgabe spielt sich im und am Computer ab. Es geht um Datensätze verschiedenster genetischer Informationen, die der Biologe durchforsten und analysieren muss. Es geht um eine Suche im virtuellen Datenraum des Lebens.

Statt *in vitro* (lat. „im Glas“) oder *in vivo* (lat. „im Lebendigen“) arbeitet der Forscher *in silico*. Dieser an das lateinische „in silicio“ (dt. in Silizium) angelehnte Begriff wurde Ende der 1980er Jahre an der kalifornischen Westküste der USA geprägt. Im Umfeld des Silicon Valley waren Computer in den Forschungsinstituten mehr und mehr als Werkzeug für die Biowissenschaften erkannt worden, um etwa biochemische Vorgänge im Computer zu simulieren. Der Aufstieg der Bioinformatik begann.

Jansen erinnert sich noch heute an seinen Start Ende '96: „Ich hatte wenig Erfahrung in Bioinformatik und wusste nicht viel über DNA-*repeats*. Daher las ich mich ins Thema ein, unter anderem ein Kapitel mit

dem Titel „Repeated Sequences“ von Sophie Bachellier vom Pariser *Institute Pasteur* in einem Band über *E. coli*-Bakterien und Salmonellen.“ Ausgestattet mit dem neuen Wissen, macht er sich ans Werk.

Jansen geht systematisch vor. Zunächst sucht er gezielt nach der bereits bekannten Wiederholungssequenz aus dem Erbgut des Tuberkulose-Erregers. Van Embdens Mitarbeiter Peter Hermans hatte sie einst im Bakterienerbgut von *M. tuberculosis* entdeckt und 1991 erstmals beschrieben (siehe das Kapitel: „[CRISPR, Tuberkulose und die Suche des Jan van Embden](#)“). Weltweit war dies erst der zweite Bericht über CRISPR-Sequenzen gewesen nach der Erstbeschreibung durch die Japaner Yoshizumi Ishino und Atsuo Nakata 1987 bei *Escherichia coli*. (siehe das Kapitel: „CRISPR-Beginn: Herr Ishino und die wunderbare Ahnungslosigkeit“).

Ruud Jansen muss also diese 36 Buchstaben lange *repeat*-Sequenz in den allmählich wachsenden Datenbanken von *Genbank* und *EMBL* aufspüren. Sein Hauptwerkzeug sind nicht Pipetten, Petrischalen oder Mikroskope, sondern ein Suchprogramm namens *BLAST* („Basic Local Alignment Search Tool“). Die *New York Times* wird es einmal „das Google der biologischen Forschung“ nennen.

Jansen tippt dazu die 36 Buchstaben lange Kette der sich wiederholenden *repeat*-Sequenz in ein Word-Dokument:

GTTTCCGTCCCCTCTCGGGGTTTTGGGTCT-  
GACGAC

Jeder der vier Buchstaben steht für eine der vier Basen des genetischen Codes: Guanin, Cytosin, Adenin, Thymin. Dann kopiert er die Buchstabenkette in das BLAST-Suchfeld und klickt auf den ‚Blast‘-Button. Die Suche in den weit entfernten Datenbanken startet.

Sie endet etwas enttäuschend. Wie schon zuvor, Anfang der 1990er Jahre, als Peter Hermans die gleiche Suche durchführt, enthält die Ergebnisliste nur Treffer für zwei Arten: die beiden Tuberkulose-Erreger *Mycobacterium tuberculosis* und *M. bovis* (die Rindervariante des Tuberkulose-Keims), und das, obwohl seit Hermans Versuch zahllose neue Erbgutsequenzen den Weg in die Datenbanken gefunden haben. Ruud Jansens logische Folgerung: Die 36-Basen lange Sequenz scheint tatsächlich nur bei den Tuberkulose-Bakterien vorzukommen.

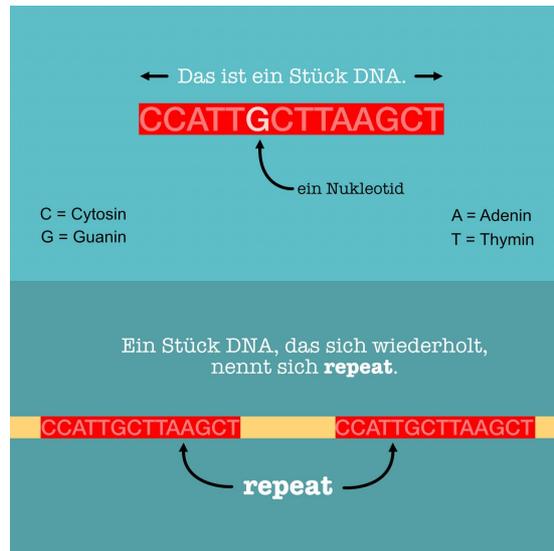
Da der *repeat* der Tuberkulose-Bakterien nichts Neues offenbart, versucht es der Postdoc über die *spacer*, die 43 DNA-Distanzstücke zwischen den *repeats*. Jansen fragt jede einzelne der 35 bis 41 Buchstaben langen, völlig unterschiedlichen Sequenzen mit dem Computer ab.

Doch auch damit hat er keinen Erfolg.

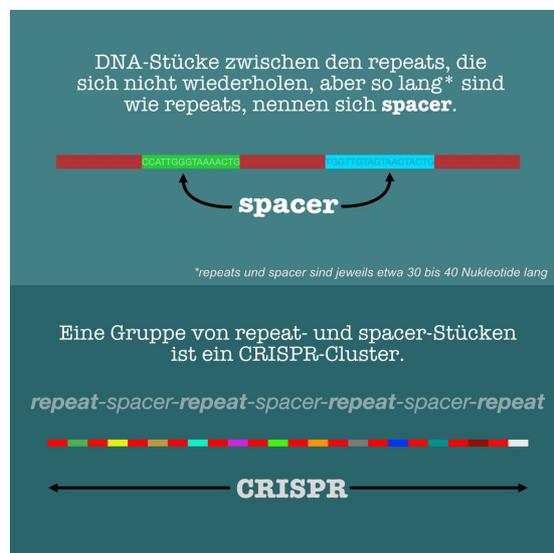
Mit keiner einzigen.

Doch was bedeutet das? Kommen all diese Sequenzen wirklich nur beim Tuberkulose-Erreger vor? Oder hat er einfach nur Pech und es sind immer noch viel zu wenige Erbgutsequenzen in den Datenbanken? Was er nicht wissen kann: Nur wenige Jahre später

werden genau solche *spacer*-Suchen den entscheidenden Hinweis liefern, welche Funktion CRISPR/Cas hat.



*CRISPR-repeats sind DNA-Abschnitte, die sich wiederholen.*



*Spacer sind die gleichlangen DNA-Stücke zwischen repeat-Sequenzen. Jeder spacer hat eine einzigartige DNA-Sequenz.*

Der CRISPR-Cluster liegt auf dem Chromosom eines Bakteriums oder Archaeobakteriums.

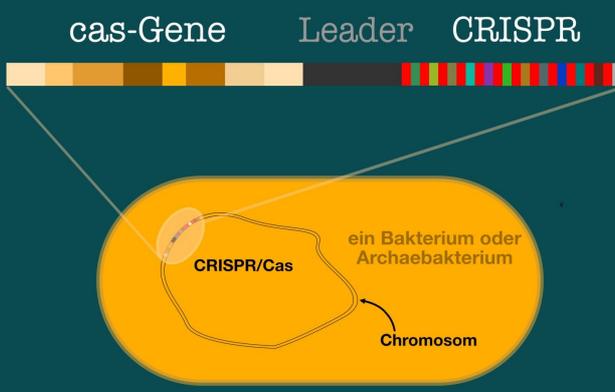
Oft finden sich mehrere solcher CRISPR-Cluster auf einem Chromosom.



The diagram shows a yellow oval representing a bacterium or archaeobacterium. Inside, a circular line represents the chromosome. A small arc of the chromosome is highlighted in red and labeled 'CRISPR'. An arrow points from the text 'ein Bakterium oder Archaeobakterium' to the cell, and another arrow points from 'Chromosom' to the circular line.

1987 wurde das repeat-spacer-Muster erstmals beschrieben.  
2002 erhielt dieses Muster den Namen CRISPR.

*Der CRISPR-Cluster findet sich auf dem Chromosom vieler Bakterien und fast aller Archaeobakterien. Häufig gibt es mehrere Cluster.*



The diagram shows a horizontal bar representing a DNA segment with three regions: 'cas-Gene' (yellow and orange blocks), 'Leader' (black block), and 'CRISPR' (red and blue blocks). Below this, a yellow oval represents a bacterium or archaeobacterium with a circular chromosome. A small arc of the chromosome is highlighted in red and labeled 'CRISPR/Cas'. Arrows connect the 'cas-Gene' and 'CRISPR' regions of the bar to the 'CRISPR/Cas' region on the chromosome. Labels include 'ein Bakterium oder Archaeobakterium' and 'Chromosom'.

2002 wird klar, dass der CRISPR-Cluster, und die cas-Gene zusammengehören.  
Deshalb heißt es CRISPR/Cas.

*Der Komplex aus cas-Genen, leader-Sequenz und CRISPR-Cluster gehören zusammen, Ruud Jansen war der Erste, der dies erkannte und in seinem Artikel 2002 berichtete.*

## Es ist das Muster, nicht die Sequenz, stupid!

Wie weiter? Was kann er tun? Hat er was übersehen? Ruud Jansen vertieft sich noch einmal in Bachelliers Buchkapitel über „Repeated Sequences“.

„Da hatte ich meinen ersten Heureka-Moment“, erzählt er heute.

„Es gibt dort ein Bild der *repeat-spacer*-Sequenzen in der Nähe des *iap*-Gens von *E. coli*-Bakterien.“ Das sind jene Sequenzen, die der Japaner Yoshizumi Ishino 1987 weltweit als Erster beschrieben hatte. Beim Anblick dieses so ähnlichen Musters kommt Jansen die Idee: Es kommt gar nicht auf die einzelne Sequenz, die genaue Buchstabenfolge an, sondern auf die gesamte Struktur: „Das ungewöhnliche Muster bei *E. coli* war genau dasselbe wie die *repeat-spacer*-Struktur unserer *M. tuberculosis*-Bakterien. Jetzt wurde mir klar, dass es mehr solcher *repeat*-Regionen gab, die man aber eben über ihr Muster und nicht über ihre spezifische Sequenz erkennen konnte“, so Jansen.

Um sicher zu gehen, checkt er zunächst weitere *repeat*- und *spacer*-Sequenzen anderer Bakterien und Archaeen, die im Laufe der 1990er Jahre hier und da zufällig beschrieben worden waren. Auch damit hat er – wie vermutet – keinen Erfolg. BLAST spuckt immer nur jene Arten als Ergebnisse aus, aus denen er die Sequenzen kopiert hat. Fazit: Viele Arten besitzen offenbar diese *repeat-spacer*-Struktur, aber jede Art hat ihre ganz eigenen *repeat-spacer*-Sequenzen. Oder anders

ausgedrückt: Jansen muss nach dem allgemeinen Muster suchen, nicht nach den spezifischen Sequenzen, wenn er erfolgreich sein will.

Doch damit hat er ein neues Problem: Das Suchprogramm BLAST kann ihm bei der Suche nach Mustern nicht weiterhelfen. Es gleicht nur die Buchstaben-Reihenfolge mit vorhandenen Sequenzdaten ab. Der Niederländer braucht ein Programm, das nicht nach Buchstabenketten, sondern nach Mustern sucht.

Wie so oft in der Wissenschaft kommt ihm der Zufall zu Hilfe: Just als Jansen sich in van Embdens Projekt einarbeitet und mit dem Bioinformatikteil des Projekts beginnt, veröffentlichen drei US-Forscher 1997 ein Computerprogramm, das wie gerufen kommt für seinen Job: *PatScan*, eine Software mit der man DNA-Sequenzen nicht nach konkreten Sequenzen wie mit BLAST durchsucht, sondern nach *pattern scannt* (daher *Patscan*), also nach Mustern durchsucht, indem man die Parameter dieses Musters unabhängig von der genauen Sequenz definiert.

Jansen findet den Hinweis auf das Programm auf einer Webseite, die – wie in den frühen Internetjahren üblich – ein schmuckloser, dennoch liebevoll zusammengetragener Linkkatalog ist, und sich „*Pedro's Molecular Tools*“ nennt. Dort gibt es eine Liste für „*Molecular Biology Search and Analysis*“, und dort stößt er auf den Link zu *Patscan*, das am *Argonne National Laboratory*, einem der ältesten und größten Forschungsinstitute des Energieministeriums der Vereinigten Staaten entwickelt und bereitgestellt worden war. Die Liste findet sich

heute noch im Netz, auch wenn einige Links längst ins Leere führen.

- [MSA](#) - Multiple Sequence Alignment at [Washington U.](#)
- [MOTIF](#) - Search Patterns in Protein Sequences at [GenomeNet](#), Japan [Motif](#) at [GenomeNet](#), Japan w/ search from [Iowa State](#)
- [Multiple Sequence Alignments](#) - CLUSTAL-W, MAP, and PIMA alignments at [Baylor College of Medicine](#)
- [NDB](#) - The Nucleic Acid Database Project Rutgers, The State University of New Jersey
- [nnPredict](#) - Predict Protein Secondary Structure at [UCSF](#) and [nnPredict](#) at w/ search from [Iowa State](#)
- [NRL\\_3D](#) - Sequence/Structure Database derived from PDB at [Johns Hopkins](#) and [NRL\\_3D](#) at [EMBL-Heidelberg](#)
- [NRSub](#) - A Non-Redundant DNA Sequence Database for *Bacillus subtilis* at [National Institute of Genetics](#), Japan
- [NuclPepSearch](#) - Search SwissProt for a Nucleotide Sequence at [ETH-Zürich](#), Switzerland
- [NYC-MASS](#) - Protein Mass Spectrometry Database at [Rockefeller U.](#)
- [OBSTRUCT](#) - Correlate Sequence Subsets of PDB Protein Structures at [EMBL-Heidelberg](#)
- [OMIM](#) - Online Mendelian Inheritance in Man at [Johns Hopkins](#)
- [Organelle Genome](#) - A Sequence Database at [U. Montreal](#)
- [OWL](#) - [Organism and Strains](#) - Links to Databases Specific to Organism and Cell Collections at [Biotech-Indiana](#)
- [PatScan](#) - Redundant Protein Sequence Database (Updated Daily) at [UCL-London](#), UK
- [PatternFly](#) - Look for Pattern Matches in DNA or Protein Sequences at [Argonne National Laboratory](#)
- [PatternFly](#) - [Patternsearch](#) - Search the SwissProt and Genpept Protein Sequence Databases with a Pattern at [ISREC](#), Switzerland
- [PDB](#) - [The Protein Data Bank](#) is now maintained by the [Research Collaboratory for Structural Bioinformatics \(RCSB\)](#).
- [PDB At A Glance](#) - Hierarchic Access to PDB Files at [NIH](#)
- [PDBFINDER](#) - A Database Containing PDB, DSSP and HSSP Information (see [Documentation](#)) at [EMBL-Heidelberg](#)
- [PDBSTR](#) - Re-organized Protein Data Bank at [GenomeNet](#), Japan
- [PepPepSearch](#) - Search SwissProt for a Peptide Sequence at [ETH-Zürich](#), Switzerland

Die Linkliste auf „Pedro's BioMolecular Research Tools“ weist Ruud Jansen zu PatScan, einem Werkzeug der Bioinformatik, mit dem er Muster in DNA-Sequenzen suchen kann.

Das Programm kommt wie gerufen und Jansens Herz schlägt höher, als er es auf der Tools-Seite entdeckt. Es ist genau das, was er braucht, um den nächsten Schritt anzugehen. Nur: Wir befinden uns Ende der 1990er Jahre, *PatScan* wird genau wie andere spezielle Computeranwendungen nicht über eine schicke und leicht zugängliche grafische Oberfläche bedient – die wird erst zwanzig Jahre später von einem deutsch/

dänischen Team entwickelt – sondern direkt auf der Kommandozeilen-Ebene. Die bekommt heute kaum noch ein Nutzer zu Gesicht, wenn er oder sie nicht gerade programmieren kann oder ein Fan der Filmtrilogie Matrix ist: grüne, manchmal auch bernsteinfarbene Schrift vor schwarzem Hintergrund. „Die wird auch heute noch viel in der Bioinformatik genutzt, aber mein bevorzugter Arbeitsbereich war das nicht“, sagt Jansen.

Er vergräbt sich in die ausführliche Anleitung und macht sich mit der speziellen Befehlssprache vertraut. Dann gibt er verschiedene Parameter für das *repeat-spacer*-Muster an: Er verlangt eine Abfolge von mindestens vier *repeats* und drei dazwischen liegenden individuellen *spacern*. Die Sequenzstücke sollen zwischen 15 und 70 Buchstaben lang sein. Jansen erklärt ein Beispiel: „Einer, der von uns verwendeten Algorithmen war die Suche nach Wiederholungen mit einer Länge von 30 bis 35 Basen, getrennt durch Abstandshalter mit einer Länge von 20 bis 45 Basen.“

Er klickt Enter.

Der Computer arbeitet ...

Und sucht.

Und sucht.

Und sucht.

Treffer!

Endlich.

Das Programm spuckt zahllose Arten mit hunderterten von Sequenzen aus. Doch was zunächst nach einem Erfolg aussieht, entpuppt sich als ein Berg neuer Arbeit, denn so gut sortiert *Patscan* die Muster dann auch wie-

der nicht aus. Die meisten der *repeats* sind große, meist unvollständige, *direkte repeats*, also sich wiederholende Sequenzen, die unmittelbar aufeinanderfolgen, ohne die CRISPR-typischen Distanzstücke dazwischen. Statt *wau-miau-wau-piep-wau-muh* wirft das Programm auch Abfolgen aus, die dem Muster *wau-wau-wau* folgen. Erst ab 2007 wird es Suchprogramme geben, die *repeat* und *spacer* wirklich gut unterscheiden können.

## Treffer, Treffer, Treffer

Ruud Jansen sortiert die Sequenzen mühsam per Hand aus der Ergebnisliste raus. Seine Suche wiederholt er regelmäßig, denn stetig laden Wissenschaftlerinnen auf der ganzen Welt neue DNA-Sequenzen und immer häufiger sogar vollständige Genome in die Datenbanken hoch. Mit jedem Suchlauf gibt es Treffer.

Durch den ersten der Suchläufe wird Ruud Jansen auf einen jungen Kollegen aufmerksam. Das Programm meldet *repeats* der salzliebenden Archaeobakterien *Haloferax mediterranei* und *H. volcanii*. Es sind jene *repeats*, die der Spanier Francisco Mojica an der Universität von Alicante in den Archaeen in den Salzlagunen an der Costa Brava entdeckt hatte (siehe das Kapitel „[Francisco Mojicas 'Salt Lovers' und der Beginn der CRISPR-Forschung \(Teil 1\)](#)“). Mojica erforscht das DNA-Muster offenbar genau so intensiv wie Jansen, nur schon seit viel längerer Zeit.

Schon 1993 hatte der Spanier das Muster erstmals beschrieben. Es war die weltweit dritte Beschreibung

überhaupt und die erste bei Archaeobakterien, nach dem Japaner Ishino und dem Niederländer Hermans. 1995 veröffentlichte er eine Arbeit, in der er als erster auch über die Funktion des „eigenartigen Musters“ spekulierte und kleine Experimente beschrieb, um die Funktion zu verstehen. Jetzt – im Jahr 1997 – schreibt Ruud Jansen dem Spanier eine erste E-Mail, um ihn darauf hinzuweisen, dass sie beide am selben Objekt forschen. Mojica freut sich über Jansens Anfrage und beide bekunden ihr Interesse an einer Kooperation, denn so wie es aussieht, sind sie die einzigen beiden weltweit, die das eigenartige Muster genauer erforschen. Sie werden in den kommenden fünf Jahren immer wieder korrespondieren (siehe das Kapitel: [„RISR, SPIDR, DVR? Wie die Genschere CRISPR zu ihrem Namen kam“](#)).

Ruud Jansens Sammlung mit Einzellern, die dieses eigenartige Muster im Erbgut besitzen, wird derweil mit jedem Suchlauf größer. Neben *Genbank* und *EMBL* durchsucht er auch ein paar Datenbanken kleinerer Genomprojekte. Im Laufe von rund drei Jahren kann er so insgesamt 72 vollständige Genome durchsuchen. Letztlich weist er bei 29 von 61 Bakterienarten und bei zehn von elf Archaeobakterien das prägnante Muster nach. Das Erstaunliche ist: Dieses Zahlenverhältnis hat sich bis heute kaum verändert, auch wenn es inzwischen nicht 72, sondern mehr als 200.000 Genome in den Datenbanken gibt: CRISPR/Cas findet sich immer noch bei fast allen Archaeen und bei weniger als der Hälfte der sequenzierten Bakterien.

Das Muster, das Jansen bei all den Einzellern vorfindet, ist innerhalb bestimmter Grenzen bei allen gleich aufgebaut. Wie bei Ishino, Mojica oder Hermans sind die *repeat*-Stücke zwischen 25 und 37 Buchstaben lang, die individullen *spacer*-Abschnitte in etwa gleich lang. Es finden sich innerhalb der *repeat*-Sequenz häufig palindromische Abschnitte, dank derer sich die lineare Nukleotid-Kette an diesen Stellen parallel aneinanderlegen kann und so eine Struktur bildet, die wie eine Haarnadel aussieht, was bestimmte Funktionen ermöglicht. *Repeats* sind jeweils typisch für eine Art, weshalb man sie nicht bei anderen Arten findet; *spacer* folgen überhaupt keiner Regel, sondern scheinen in jeder Spezies und von Art zu Art völlig wahllos zusammengewürfelt zu sein.

Es gibt Bakterien, die nur einen einzigen *repeat*-Cluster besitzen, häufig sind es aber mehrere, mit bis zu zwanzig Ansammlungen pro Genom. Archaeen haben generell größere Cluster als Bakterien.

Jansen ist zufrieden. Er hat eine ganz besondere Klasse von sich wiederholenden DNA-Sequenzen entdeckt. Sein Chef Jan van Embden hatte gehnt, dass es sich lohnen würde, dieses Muster zu erforschen.

Und, so findet Jansen, da es sich um eine ganz eigene Klasse von *repeat*-Sequenzen handelt, scheint es nur angemessen, dafür auch einen eigenen Namen zu vergeben, der anders als bisher, die Bedeutung der Distanzstücke, der *spacer*, berücksichtigt, die ja auch zum Muster gehören. Statt DR (für *direct repeats*), tauft der Niederländer sie: SPIDR, das für „**SP**acer **I**ntersper-

sed **Direct Repeats**“ steht. Es bedeutet in etwa: mit Distanzstücken durchsetzte direkte Wiederholungen.

Als Ruud Jansen den Namen 1998 oder '99 kreiert (so genau erinnert er sich heute nicht mehr), weiß er noch nicht, dass Francisco Mojica fast zur selben Zeit eine ähnliche Herleitung wählt, und das „eigenartige Muster“ SRSR tauft: **Short Regularly Spaced Repeats**, auf Deutsch: kurze, regelmäßig unterbrochene Wiederholungssequenzen. Obwohl beide Forscher in Kontakt stehen und sie anfangs auch über eine Kooperation nachdenken, ist ihr Austausch nicht so eng, dass sie sich über jeden Punkt verständigen: „Es war eine gesunde Konkurrenz zwischen uns beiden“, sagt Jansen heute.

Nur wenig später sollte er zu spüren bekommen, was das heißt.

## **Mojica kommt Jansen zuvor**

Seine aufregenden Ergebnisse, die er seit 1997 durch akribische Arbeit am Computer gesammelt hatte, präsentiert der niederländische Forscher im April 2000 in Paris auf der internationalen Konferenz Genomes 2000 der *American Society for Microbiology* am berühmten *Institute Pasteur*. Doch das eigenartige Muster interessiert kaum jemanden der Konferenzteilnehmerinnen. Doch es kommt noch härter.

Zur Überraschung von Jansen und seinen niederländischen Kollegen veröffentlicht fast zeitgleich Francisco Mojica die Ergebnisse seiner eigenen mühevollen

Arbeit der letzten sieben Jahre in der April-Ausgabe des renommierten Fachmagazins *Molecular Microbiology*. Im Prinzip berichtet er das gleiche wie das, was Jansen auf der Konferenz vorstellt (siehe das Kapitel [„Francisco Mojicas 'Salt Lovers' und der Beginn der CRISPR-Forschung \(Teil 3\)“](#)).

Da ein Konferenzbeitrag lange nicht die Bedeutung einer Magazinveröffentlichung hat, wollte Jansen seine Ergebnisse natürlich auch bei einem Fachmagazin einreichen, doch Mojica ist ihm zuvorgekommen. „Als ich seinen Artikel las, war ich gar nicht amüsiert, denn er hatte uns davon gar nichts gesagt. Von diesem Zeitpunkt an war klar, dass wir keine Kooperationspartner, sondern Konkurrenten waren“, blickt er heute zurück.



*Ruud Jansen und Francisco Mojica waren die einzigen, die in der Anfangszeit CRISPR/Cas so intensiv erforschten. Anfangs wollten sie kooperieren, dann wurden sie zu Konkurrenten.*

Jansen schafft es zwar noch, dass ein Magazin seine Arbeit annimmt, aber es wird letztlich das noch völlig unbekannte [neue Journal Omics](#), dessen allererste Ausgabe 2001 erscheinen soll, dann aber erst im Januar 2002 herauskommt.

Mojica war schneller. Und das wurmt Jansen.

„Wir hatten bis dahin nur diese stetig wachsende Liste von Mikroben, die das SPIDR-Muster besaßen. Ich brauchte mehr als eine Liste, die Mojica ja schon publiziert hatte“, erzählt Jansen heute. „Also nahm ich mir nochmal Sofia Bachelliers Kapitel über „Repeated Sequences“ vor.“

Und da findet der Niederländer tatsächlich die Information, die ihn rettet, die seine Arbeit über Mojicas Untersuchung aus dem Jahr 2000 hinausbringen wird. Die ihm nicht nur einen Platz im gleichen, renommierten Journal sichern wird, sondern auch einen Platz in der Ahnengalerie von CRISPR/Cas.

In Bachelliers Kapitel fällt ihm auf, dass *repeat*-Sequenzen immer wieder mit „springenden Genen“ in Verbindung gebracht werden. Das sind Gene, die die Fähigkeit haben, selbstständig in einem Genom umherzuwandern. „Springende Gene“ oder Transposons hatte 1948 die Botanikerin Barbara McClintock in Mais entdeckt und 1983 den Nobelpreis dafür bekommen. In der Nähe vieler solcher Gene mit Sprungfähigkeit finden sich immer ein paar *repeats*, denn sie sind Teil der Maschinerie. „Vielleicht sind die SPIDR-*repeats* mit Transposon-Genen verbunden?“, fragt sich Jansen.

Er sucht nach genähnlichen Sequenzabschnitten in der Nähe der SPIDR-Clusters, wie etwa für eine Transposase, ein Enzym, das das springende Gen heraus-schneidet. „Aber ich fand nichts transposase-artiges“, sagt Jansen. Stattdessen beginnt er mit BLAST nach ähnlichen Gensequenzen in den Datenbanken zu suchen, findet aber auch nichts. Er versucht es mit Proteinsequenzen, die man aus den vermuteten Gensequenzen ableiten kann.

Wieder nichts.

Doch dann fällt ihm etwas auf.

Die Protein-Sequenz bei einem der Bakterien, die er ins Suchfeld von BLAST eingeben will, kommt ihm bekannt vor.

Genau! Sie ähnelt sehr stark der Proteinsequenz eines Einzellers, den er nur wenige Minuten zuvor untersucht hat.

Das ist es! „Da wird mir klar, dass beide Einzeller das gleiche Gen in der Nähe des SPIDR-Clusters besitzen“, sagt Jansen. Jetzt hat er einen Ansatz, um weiter zu suchen.

Und tatsächlich: Das Herumstöbern lohnt sich.

## **CRISPR und Cas: „You complete me.“**

Ruud Jansen macht seine wichtigste Entdeckung. Erst diese komplettiert das Gesamtbild in der CRISPR-Historie. Erst jetzt geben sich CRISPR und Cas vollständig als untrennbares Gespann zu erkennen, so wie Oli-

ver Hardy erst durch Stan Laurel komplett ist, Romeo durch Julia oder Asterix durch Obelix.

Ruud Jansen entdeckt das Cas in CRISPR und Cas.

In nächster Nähe zum *repeat-spacer*-Cluster und einer 300 bis 500 Basen langen *leader*-Sequenz, eine Art Vorspann, die ebenso typisch ist, findet der Forscher echte Gene, also DNA-Stücke, die ein Rezept, eine Anleitung für den Bau von Proteinen liefern, die eine ganz bestimmte Aufgabe erfüllen. Proteine sind die Arbeitstiere innerhalb jeder Zelle und in jedem Lebewesen. Sie werden auch biologische Roboter genannt, weil sie unermüdlich zahllose, aber spezifische Aufgaben erledigen – ob als Verdauungsenzym, Sehfärbstoff, Muskelaufbaustoff, Signalrezeptor, Hormon oder was auch immer. An jeder Funktion im Körper sind Proteine beteiligt. Und zu dieser Gruppe von komplexen Makromolekülen gehört eben auch das Protein Cas9, die Gen-Schere, für die zwei Forscherinnen 18 Jahre später im Jahr 2020 den Nobelpreis bekommen werden.

Es sind insgesamt vier Gene, die Jansen mehr oder weniger konstant bei den verschiedenen Arten wiederfindet. Sie sitzen eng an eng nur wenige hundert Basenpaare entfernt vom SPIDR-Abschnitt und der *leader*-Sequenz. Das weist auf eine mögliche Verbindung hin. Noch stärker wird diese Verbindung durch Jansens Entdeckung, dass die vier Gene ausgerechnet bei den Bakterien und Archaeen fehlen, denen auch das SPIDR-Muster fehlt.

**Januar 2002**

**OMICS**  
Band 6, Nr. 1, S. 23-33

**Identification of a Novel Family of Sequence Repeats among Prokaryotes**

**Ruud Jansen  
Jan van Embden  
Wim Gaastra  
Leo Schouls**

**März/April 2002**

**Molecular Microbiology**  
Band 43, Nr. 6, S. 1565-1575

**Identification of genes that are associated with DNA repeats in prokaryotes**

**Ruud Jansen  
Jan D. A. van Embden  
Wim Gaastra  
Leo Schouls**

2002 erscheinen gleich zwei Artikel von Ruud Jansen, die beide zum Teil dasselbe berichten. Der zweite Artikel in Molecular Microbiology präsentiert zum ersten Mal den Namen CRISPR/Cas und Jansen berichtet zusätzlich zur CRISPR-Liste aus dem ersten Artikel, dass er vier cas-Gene entdeckt hat.

Die noch namenlosen vier Gene, die leader-Sequenz und das SPIDR-Muster gehören offensichtlich zusammen.

Ruud Jansen ist der Erste, der diese enge Verbindung von Genen und dem „eigenartigen Muster“ aus *repeats* und *spacern* auffällt.

Dem Postdoc ist bewusst, dass dies eine wirklich aufregende Entdeckung ist und seine Arbeit auf ein neues Level hebt. Er nimmt einen neuen Artikel in Angriff, den er bei *Molecular Microbiology* einreichen kann, dem Magazin, in dem auch Mojicas Artikel erschienen war. Jansen versucht sogar, seinen ersten, von *Omic*s bereits angenommenen Artikel zurück-zuziehen, doch der Chefredakteur will den Artikel unbedingt in seinem neuen Journal bringen. So kommt es zu dem Kuriosum, dass es heute zwei *Jansen et al.*-Artikel aus dem Jahr 2002 gibt, die zum Teil identische Ergebnisse präsentieren (und beide mit „*Identification* (...)“ beginnen).

## Das Kind braucht einen neuen Namen

Für seine neue Publikation trifft der Niederländer eine wegweisende Entscheidung: Es soll ein Ende haben mit der ganzen Namensverwirrung. Neben seinem SPIDR-Locus gibt es so viele andere Bezeichnungen für das *repeat-spacer*-Muster, dass sie keine Hilfe sind, sondern nur Verwirrung stiften. Er kontaktiert wieder einmal Francisco Mojica und schlägt ihm vor, einen neuen, prägnanten Namen zu suchen, der alle anderen, auch sein SPIDR und Mojicas SRSR, ersetzen soll.

Es geht ein, zwei Wochen hin und her, dann einigen sie sich auf: CRISPR, *Clustered Regularly Interspaced Short Palindromic Repeats*, zu Deutsch etwa „gruppierte, regelmäßig unterbrochene, kurze, palindromische Wiederholungssequenzen“ (siehe das Kapitel: „[RISR, SPI-DR, DVR? Wie die Genschere CRISPR zu ihrem Namen kam](#)“). Jansen wird seine neu entdeckten Gene daran angelehnt Cas-Gene nennen: *CRISPR associated genes*; Gene, die mit CRISPR verbunden sind. Weil es vier Cas-Gene sind, die er entdeckt hat, nennt er sie *cas1*, *cas2*, *cas3* und *cas4* (Gen-Namen erkennt man am Kleinbuchstaben am Anfang, Protein-Namen am Großbuchstaben). Cas1 findet er bei allen Einzeller-Arten, die auch CRISPR haben, cas2 bis cas4 finden sich nicht bei jeder Art, aber bei den meisten. Ein wenig kann Jansen sogar über die Funktion der Gene spekulieren, denn die Gensequenzen geben Hinweise auf mögliche Aminosäuresequenzen. Und damit lassen sich Aufgaben der Proteine erahnen, zumindest für die Proteine Cas3 und Cas4.

Cas3 könnte so etwas wie eine Helikase sein, das sind Proteine, die helfen, die verdrehte DNA-Helix zu entwinden, während Cas4 zu den Exonukleasen zählt und womöglich am Auf- und Abbau von DNA-Sequenzen beteiligt ist oder bei dem Vorgang, der Genexpression genannt wird, bei dem aus einzelnen Genen, Proteine werden.

Diese Beschreibungen passen tatsächlich zu den heute bekannten Aufgaben im CRISPR/Cas-Komplex. Doch sie sind noch so allgemein, dass man nicht wirk-

lich behaupten kann, Jansen hätte schon damals irgendeine Ahnung davon gehabt, welche Aufgabe das ganze „CRISPR/Cas-Ding“ tatsächlich hat. Damit ist er indes nicht allein. Auch eineinhalb Jahrzehnte nach der Erstbeschreibung von CRISPR durch Ishino und Nakata, weiß noch niemand welche Aufgabe das eigenartige Muster aus *repeats* und *spacers* in den Einzellern erfüllt. Daran ändern auch die *cas*-Gene nichts.

Festzuhalten aber bleibt: Im Jahr 2002, fünfzehn Jahre nachdem CRISPR entdeckt wurde, betritt endlich auch Cas die Bühne. Und man könnte sich – etwas kitschig natürlich – das personifizierte CRISPR vorstellen, das seinem Cas-Gegenüber entgegentritt und haucht: „You complete me.“ Auch wenn dies erst den Anfang einer sehr langen Bühnenromanze markiert, das noch viele weitere Höhe- und Tiefpunkte erleben wird.

Ruud Jansen jedenfalls fasst all seine Ergebnisse und den neuen Namen in einem Artikel zusammen und reicht ihn mit Jan van Embden und zwei weiteren Kollegen bei *Molecular Microbiology* ein. Der Spanier wird im Artikel der Niederländer natürlich nicht als Autor gelistet, aber sie danken ihm für die „anregenden Diskussionen“. Jansen und Mojica haben sich in dieser Zeit nie persönlich getroffen oder gesprochen und sind sich bis heute nie persönlich begegnet.

Der Artikel wird am 10. Dezember 2001 von der Redaktion akzeptiert und im Band 43 in der zweiten Märzangabe (Ausgabe 6) 2002 auf den Seiten 1565 bis 1575 mit dem Titel „[Identification of genes that are associated with DNA repeats in prokaryotes](#)“ veröffent-

licht. Online erscheint die Ausgabe am 25. April 2002. Die gedruckte Ausgabe war zwar schon früher herausgekommen (anders als es heute üblich ist), aber der Verlag kann heute – auf meine Nachfrage hin – nicht mehr sagen, an welchem Tag dies genau war.

**Was genau ist eigentlich CRISPR/Cas?**

Der Begriff CRISPR/Cas wird sehr breit verwendet. Die einen bezeichnen damit die berühmte Gen-Schere, die Jennifer Doudna und Emmanuelle Charpentier 2012 in Science vorgestellt und für die sie 2020 den Nobelpreis für Chemie bekommen haben.

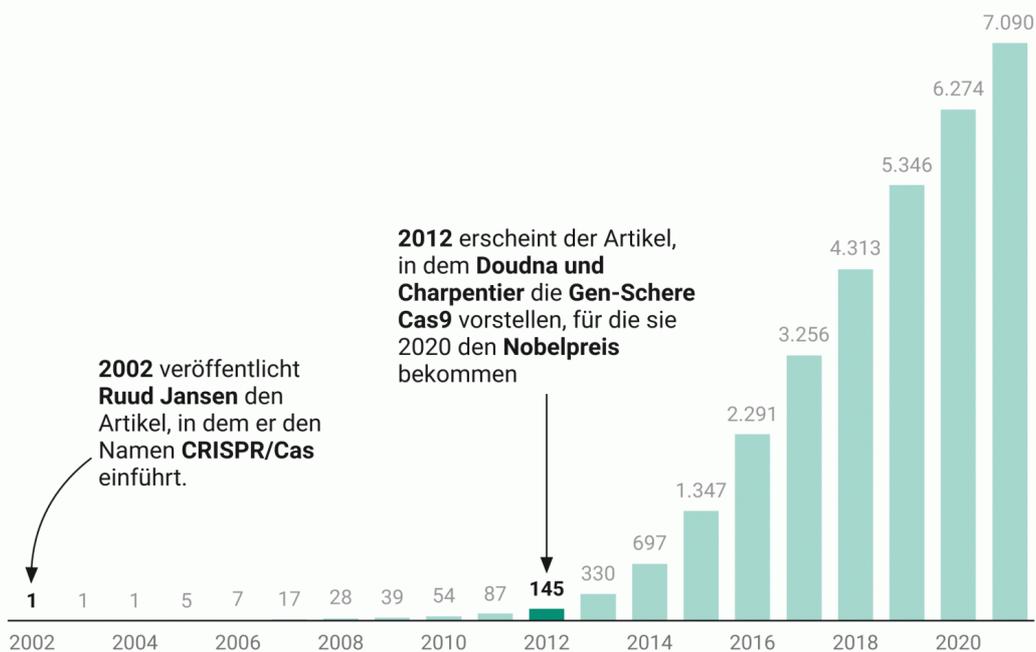
Seit 2002 beschreibt der Begriff aber vor allem ein Immunsystem von Bakterien und Archaeen gegen Viren und andere eindringende genetische Elemente.

Die Gen-Schere ist von diesem Immunsystem abgeleitet und müsste genau genommen einfach nur Cas9 heißen (oder zumindest CRISPR/Cas9), weil das Enzym, das die Viren-DNA zerschneidet, das Protein Cas9 ist, an das eine spezifische RNA-Sequenz angehängt wird. Die Sequenz weist dem Protein den Weg an eine ganz bestimmte Stelle einer DNA-Sequenz, an der die „Schere“ dann die Sequenz zerschneidet. Diese RNA-Adresse kann sehr vielfältig von Forscherinnen ausgewählt werden. Mittlerweile werden auch andere Proteine wie Cas12 oder Cas3 als Gen-Scheren eingesetzt.

Beim Immunsystem der Einzeller stammt diese Ziel-Sequenz aus dem CRISPR-Cluster, zusammengesetzt aus einem der spacer und einem Teil der benachbarten repeat-Sequenz. Der spacer wurde bei einem früheren Virenangriff aus einer Viren-Sequenz herausgeschnitten und im CRISPR-Cluster archiviert für künftige Angriffe genau dieser Virenart.

Ruud Jansens Artikel ist einer der Meilensteine in der Geschichte von CRISPR und Cas. Die Arbeit wurde seitdem mehr als eintausend Mal in anderen Fachartikeln zitiert, so häufig wie Ishinos Artikel der CRISPR-Erstbeschreibung fünfzehn Jahre zuvor und mehr als doppelt so oft wie Mojicas Arbeit aus dem Jahr 2000.

**Seit dem ersten Fachartikel mit dem Namen "CRISPR/Cas" sind mehr als 30.000 weitere Artikel erschienen.**



Grafik: Marcus Anhäuser • Quelle: PubMed • Erstellt mit Datawrapper

Laut der Analyseplattform *Dimensions* wird Jansens Artikel 58-mal häufiger zitiert als eine ganz gewöhnliche Arbeit in diesem Feld. Nicht nur weil Jansen die ersten cas-Gene im Zusammenhang mit CRISPR entdeckt hatte, sondern weil es der erste Artikel weltweit ist, in dem der Name CRISPR/Cas vor-

kommt. Zwanzig Jahre später sind es allein für das Suchwort CRISPR fast 30.000 Artikel und jedes Jahr kommen mehr dazu.

## Der Ausstieg

Doch das war's dann auch. Anders als Kollege Mojica spielt Ruud Jansen in der CRISPR-Historie keine weitere Rolle mehr. Er hat – wie er sich später erinnert – regelrecht die Nase voll und steigt aus. So spannend er das Thema findet, hatte er schon die letzten Jahre seiner CRISPR-Forschung keine Finanzierung mehr bekommen und sich mit dem Geld für die anderen Projekte über Wasser gehalten. Wie vielen anderen Wissenschaftlerinnen verleiden auch dem inzwischen 45-Jährigen das ständige Betteln um Geld, die Befristung der Stellen, der pausenlose Kampf um externe Finanzierung und die anhaltende Unsicherheit die Freude an der akademischen Grundlagenforschung.

Er sieht in diesem Bereich keine Zukunft mehr: „Man darf nicht vergessen: CRISPR war zu dieser Zeit wirklich etwas Seltsames, etwas am Rande der Wissenschaft“, sagt Jansen heute. Das macht ihm den Abschied leichter. Niemand ahnt damals, dass CRISPR/Cas eines Tages eine wissenschaftliche Revolution auslösen wird. „Ich dachte, CRISPR und die Cas-Gene verschwänden in der schieren Menge anderer Entdeckungen, die durch all das Sequenzieren ganzer Genome, allen voran das „Human Genome“-Projekt, hervorgebracht würden. Doch zu meiner Überraschung

ist das nicht passiert“, sagt er heute. Kurz nach Erscheinen seines CRISPR-Artikels kehrt Jansen der Universität den Rücken und tritt eine besser bezahlte, unbefristete Position an einem unabhängigen, privaten Institut für medizinische Diagnostik an, wo er bis 2021 arbeiten wird. Inzwischen ist er freier Berater in diesem Bereich.

Er vergisst CRISPR und Cas für ein paar Jahre. Erst um das Jahr 2007 herum kommt er darauf zurück. Er will wissen, ob irgendjemand seine Arbeit zitiert hat und durchsucht die biomedizinische Literaturdatenbank *PubMed* nach dem Stichwort CRISPR/Cas. Und tatsächlich: Er findet einige wenige Artikel. Einige Forscherinnen, darunter sein alter Bekannter Francisco Mojica, scheinen doch endlich einer möglichen Funktion der Struktur auf die Spur zu kommen. Und es ist die Suche nach den *spacer*-Abschnitten mit Hilfe des BLAST-Programms, die die entscheidenden Hinweise liefert, genau der Ansatz, den Jansen und vor ihm seine niederländischen Kollegen in der van Embden-Gruppe versucht hatten – damals leider ohne Erfolg.

Er war wohl einfach einen Moment zu früh in der Geschichte, als noch zu wenige Sequenzen in den Datenbanken dieser Welt lagern, um einen Treffer zu landen, um so der wahren Funktion von CRISPR auf die Spur zu kommen.

Ihm sei damals, 2007, als er die ersten Hinweise auf eine Funktion von CRISPR/Cas las, sogar ein Experiment eingefallen, wie man diese Funktion bei den Tuberkulosebakterien überprüfen könnte, schreibt er

mir. Wenige Wochen später sei dann genau ein solches Paper in *Science* erschienen. Es habe hieb- und stichfest die Funktion und Herkunft von CRISPR/Cas bei Bakterien bewiesen.

„Schade, dass ich da nicht mehr in diesem spannenden Bereich der Grundlagenforschung tätig war“, so Jansen. „Trotzdem ist es schön zu wissen, dass ich den Namen eingeführt und die Cas-Gene entdeckt habe.“ Und mit einem hat Ruud Jansen definitiv Recht behalten: Wer heute CRISPR in die Suche von *Pubmed* eingibt, landet ganz sicher nur bei Forschung, die sich um das einst so ominöse *repeat-spacer*-Muster dreht. CRISPR ist ein einzigartiges Akronym geblieben (siehe das Kapitel: [„RISR, SPIDR, DVR? Wie die Genschere CRISPR zu ihrem Namen kam.“](#)).

Jansens Projektleiter Jan van Embden hatte jedenfalls das richtige Näschen gehabt. Es hatte sich gelohnt, dieses ungewöhnliche DNA-Muster genauer zu erforschen. Doch auch mit ihm geht die CRISPR-Geschichte nicht weiter. Zwei Jahre nach Jansens Abtritt verabschiedet sich Jan van Embden in die Rente (siehe das Kapitel: [„CRISPR, Tuberkulose und die Suche des Jan van Embden“](#)). Damit endet die niederländische CRISPR/Cas-Linie, die parallel zur spanischen Linie Francisco Mojicas so erfolgreich dieses völlig neue Forschungsgebiet erschlossen hat.

Die spanische Linie hingegen läuft bis heute weiter. Nur wenig später – im Jahr 2005 – wird Mojica eine weitere entscheidende Entdeckung berichten, die wesentlich dazu beiträgt, endlich das Rätsel zu lösen,

welche Aufgabe CRISPR und Cas in Bakterien und Archaeen erfüllen.

Aber auch dieses Mal ist er nicht der Einzige.

Dies ist die PDF-Version von Kapitel 6 der #CRISPRhistory-Serie. Die Webversion findet sich unter dieser URL:

<https://www.riffreporter.de/de/wissen/crispr-cas-entdeckung-name-biotechnologie-gentechnik-nobelpreis>

Dies ist ein Beitrag der Serie [#CRISPRCas9 – Die Biografie der revolutionären Gen-Schere CRISPR/Cas9](#) von Marcus Anhäuser. Verpasse keine Folge und [abonniere den Newsletter](#) oder verfolge [Twitter](#), [Facebook](#) und [Instagram](#).

Fragen, Lob oder Kritik? Einfach eine E-Mail an mich, Marcus Anhäuser: [m.anaeuser@me.com](mailto:m.anaeuser@me.com)

#CRISPRhistory  
Die Biografie der revolutionären  
Gen-Schere CRISPR/Cas9.  
Von Marcus Anhäuser  
[https://www.riffreporter.de/de/magazine/  
crispr-biotechnologie](https://www.riffreporter.de/de/magazine/crispr-biotechnologie)