CRISPR

2017 July 22 (Sat.)
R. Newlon M1
Literature Seminar

First mention, 1987

Journal of Bacteriology, Dec. 1987, p. 5429-5433
Copyright © 1987, American Society for Microbiology

Nucleotide Sequence of the *iap* Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in *Escherichia coli*, and Identification of the Gene Product

YOSHIZUMI ISHINO, HIDEO SHINAGAWA, KOZO MAKINO, MITSUKO AMEMURA, AND ATSUO NAKATA*
Department of Experimental Chemotherapy, The Research Institute for Microbial Diseases, Osaka University, Yamadaoka, Suita, Osaka 565, Japan

Ishino, Y, *et al.* J Bacteriol. 1987 Dec; 169(12): 5429–5433.

codon, a nucleotide sequence with a transcript that may form a stable stem-and-loop structure was found. Transcription may end at this region.

An unusual structure was found in the 3'-end flanking region of *iap* (Fig. 5). Five highly homologous sequences of 29 nucleotides were arranged as direct repeats with 32 nucleotides as spacing. The first sequence was included in the putative transcriptional termination site and had less homology than the others. Well-conserved nucleotide sequences containing a dyad symmetry, named REP sequences, have been found in *E. coli* and *Salmonella typhimurium* (28) and may act to stabilize mRNA (18). A dyad symmetry with 14 nucleotide pairs was also found in the middle of these sequences (underlining, Fig. 5), but no homology was found between these sequences and the REP sequence. So far, no sequence homologous to these has been found elsewhere in procaryotes, and the biological significance of these sequences is not known.

- Unknown, unusual structure found
- Repeats, interspaced by 32 apparently random nucleotides
- Only 5 repeats identified at first
- Significance unknown

5432 ISHINO ET AL. J. BACTERIOL.

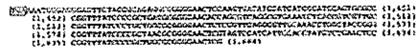


FIG. 5. Comparison of direct-repeat sequences consisting of 61 base pairs in the 3'-end flanking region of *iap*. The 29 highly conserved nucleotides, which contain a dyad symmetry of 14 base pairs (underlined), are shown at the bottom. Homologous nucleotides found in at least two DNA segments are shown in boldface type. The second translational termination codon is boxed. The nucleotide numbers are in parentheses.

Ishino, Y, *et al.* J Bacteriol. 1987 Dec; 169(12): 5429–5433.

Members of the same group, 1988/1989. Found 21 repeats, split in two clusters.

Journal of Bacteriology, June 1989, p. 3553-3556
Copyright © 1989, American Society for Microbiology

Unusual Nucleotide Arrangement with Repeated Sequences in the *Escherichia coli* K-12 Chromosome

ATSUO NAKATA,* MITSUKO AMEMURA, AND KOZO MAKINO
Department of Experimental Chemotherapy, Research Institute for Microbial Diseases, Osaka University, Yamadaoka, Suita, Osaka, Japan 565

Nakata, A, *et al.* J Bacteriol. 1989 Jun; 171(6): 3553-6.

What does CRISPR mean?

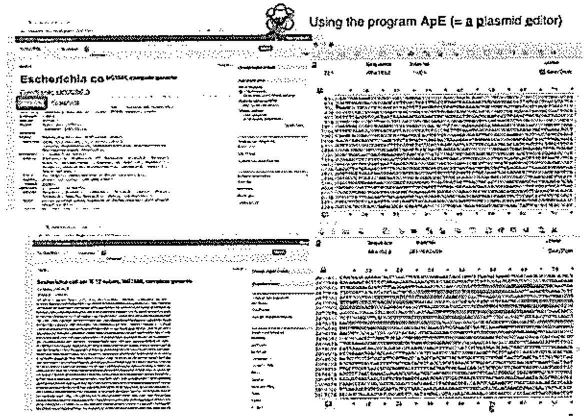
Clustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (name selected in 2002)

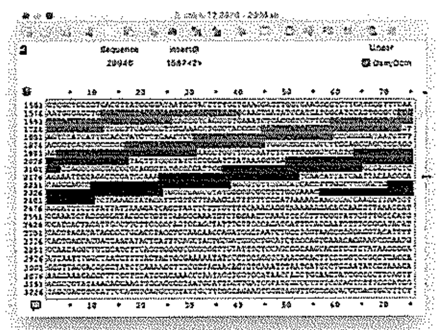
Previously called:

- DVR (direct variable repeats)
- TREP (tandem repeats)
- LTRR (long tandemly repeated repetitive sequences)
- SRSR (short regularly spaced repeats)
- LCTR (large clusters of tandem repeats)
- SPIDR (spacer interspersed direct repeats)

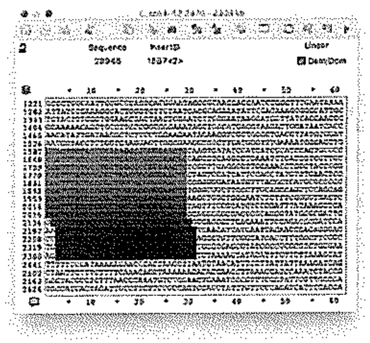
Barrangou, R. and Horvath P. *Nat. Microbiol.* 2, 17092 (2017)

Using the program ApE (= a plasmid editor)

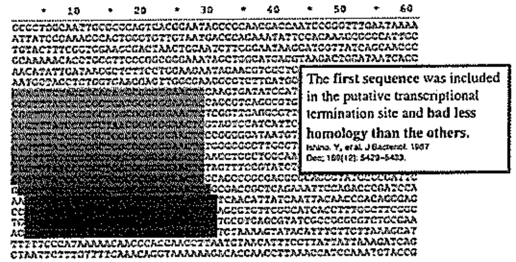




Region of interest shown clipped from complete genome; numbers not accurate.



Region of interest shown clipped from complete genome; numbers not accurate.



20 of the 21 repeats differ from this sequence by only two bases or less. (The first repeat being less homologous)

5' CGGTTTATCCCCGCTGGCGCGGGGAACWC 3'

The repeats contain a
7-base inverted repeat

AT
GC
GC
GC
GC
CG
GC

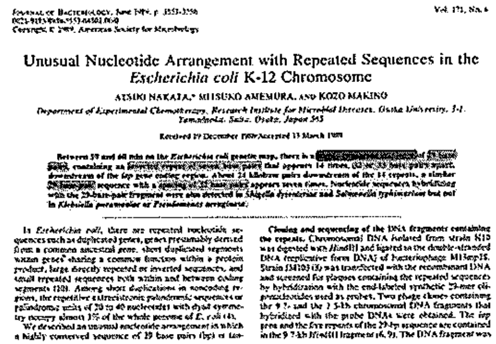
5' CGGTTTATCCCCGCTGGCGCGGGGAACWC 3'
3' GCCAAATAGGGGCGACCGCGCCCCTTGSG 5'

A palindrome has no intervening nucleotides, but an inverted repeat can be called "partially palindromic."

5' TTATCCCGGGAAC 3'
3' AATAGGGCCCTTG 5'

Example palindrome

palindrome = ...

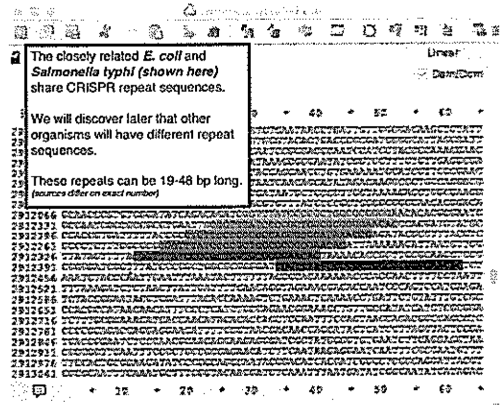

Journal of Bacteriology, June 1989, p. 3553-3556
Vol. 171, No. 6

Unusual Nucleotide Arrangement with Repeated Sequences in the *Escherichia coli* K-12 Chromosome

ATSUO NAKATA,* MITSUKO AMEMURA, AND KOZO MAKINO

Department of Experimental Chemotherapy, Research Institute for Microbial Diseases, Osaka University, 3-1, Yamadaoka, Suita, Osaka, Japan 565

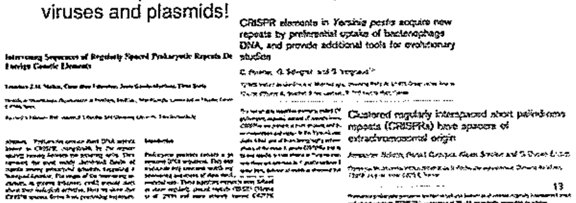
Nakata, A. *et al.* J Bacteriol. 1989 Jun;171(6):3553-6.

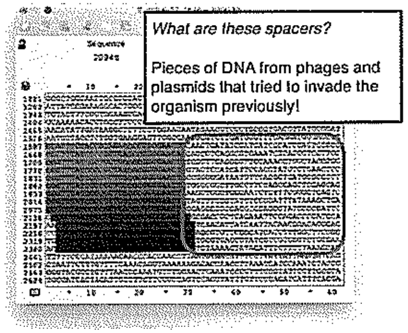


Barrangou, R. and Horvath P. *Nat. Microbiol.* 2, 17092 (2017)

Start of the Story

- 1987: arrays of DNA repeats observed in *E. coli* *(later other bacteria and archaea as well)*
- 2002: CRISPR sequences used to compare different strains of *S. thermophilus* in the food industry
- 2002: some (proto)spacers are conserved across strains; these strains shared phage resistance phenotypes!
- 2005: the spacers match DNA sequences from invasive viruses and plasmids!





Region of interest shown clipped from complete genome; numbers not accurate.

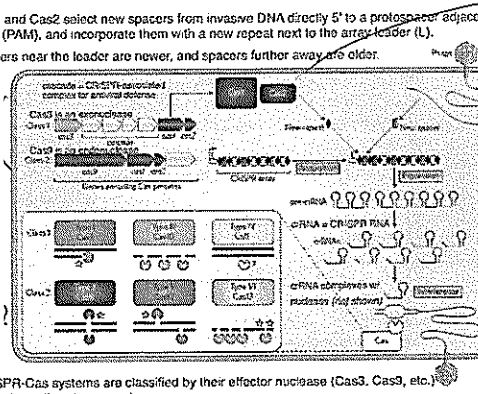
What is its function?

- CRISPR arrays stores pieces of foreign DNA.
- Genes near the array express proteins = CRISPR associated proteins **(Cas proteins)**
- These Cas proteins variously:
  1. acquire pieces of invader DNA
  2. incorporate new pieces as spacers in the CRISPR array, separated by repeats
  3. are guided to invading DNA, using the expressed spacer RNA as guides
  4. cleave invading DNA, protecting the organism
- 84% sequenced archaea and 45% bacteria encode a CRISPR-Cas system (Barrangou and van der Oost (2013) Springer Press, Heidelberg, pp. 1-129)

Barrangou, R. and Horvath P. *Nat. Microbiol.* 2, 17092 (2017)

Cas1 and Cas2 select new spacers from invasive DNA directly 5' to a protospacer adjacent motif (PAM), and incorporate them with a new repeat next to the array leader (L).

Spacers near the leader are newer, and spacers further away are older.

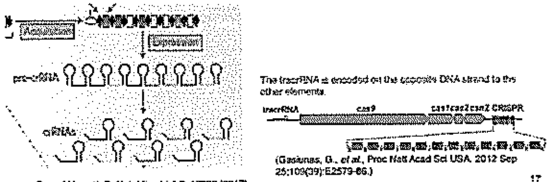


CRISPR-Cas systems are classified by their effector nuclease (Cas3, Cas9, etc.)
Class 1: multi-protein complex
Class 2: single protein

Barrangou, R. and Horvath P. *Nat. Microbiol.* 2, 17092 (2017)

CRISPR RNA (crRNA)

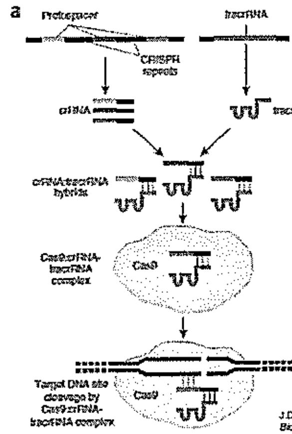
- The CRISPR array is expressed as a whole **(pre-crRNA)** before being processed into separate, smaller crRNAs
- Each crRNA contains the spacer and part of the CRISPR repeat
- In **Type II systems**, a transactivating crRNA **(tracrRNA)** is also expressed
- The tracrRNA hybridizes with the different crRNAs
- Each hybrid then complexes w/ the effector nuclease (Cas9), and guides it to the DNA targeted by the crRNA.



(Gasiunas, G., *et al.*, Proc Natl Acad Sci USA. 2012 Sep 25;109(39):E2579-86.)

Barrangou, R. and Horvath P. *Nat. Microbiol.* 2, 17092 (2017)

Type II System

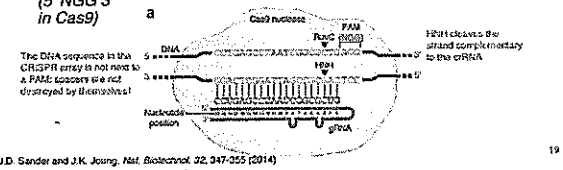


The tracrRNA hybridizes with the repeat portion of the crRNA
(not the spacer portion)

J.D. Sander and J.K. Joung. *Nat. Biotechnol.* 32, 347-355 (2014)

Cas9

- Of all the CRISPR-Cas systems, Type II is generating the most interest.
- The effector nuclease is the single **Cas9 protein**
- Cas9 is an **endonuclease** that creates blunt DNA double-stranded break (**DSB**)
- It uses an RuvC and an HNH motif to cut the target DNA
- Cleavage (3 nt upstream) only happens if the target site is immediately 5' to a **protospacer adjacent motif (PAM)** *(5'NGG 3' in Cas9)*

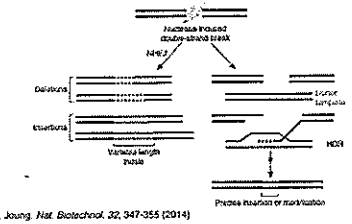


J.D. Sander and J.K. Joung, *Nat. Biotechnol.* 32, 347-355 (2014)

19

Double-stranded break repair

- Cells can repair DSBs using endogenous systems for:
  - Nonhomologous end-joining (NHEJ)
    - Easily introduces insert/deletion mutations of variable length
    - Can disrupt reading frames or binding sites
  - Homology-directed repair (HDR)
    - Recombines with a donor template to correctly repair break



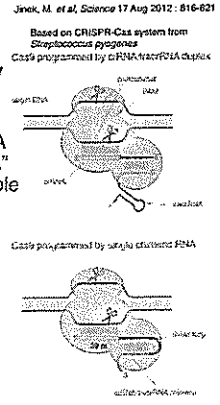
J.D. Sander and J.K. Joung, *Nat. Biotechnol.* 32, 347-355 (2014)

20

The Breakthrough(s)

- CRISPR-Cas systems are guided by short pieces of RNA to specific locations to make cuts.
- Can use a fusion (chimera) of crRNA and tracrRNA to make a "guide RNA" (**gRNA**) of our choice = programmable

Jinek, M. *et al*, *Science* 17 Aug 2012 : 816-821

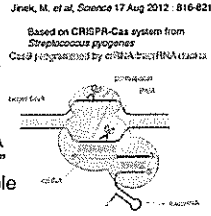


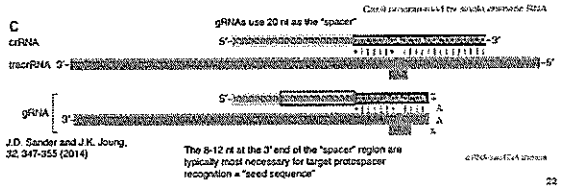
21

The Breakthrough(s)

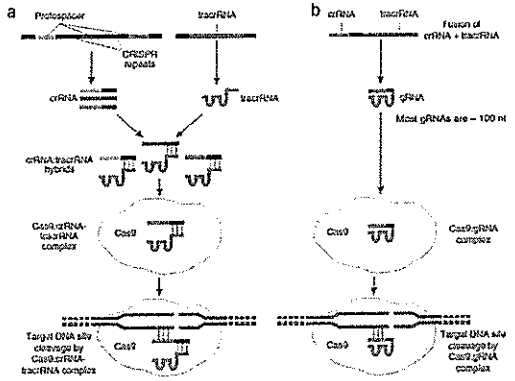
- CRISPR-Cas systems are guided by short pieces of RNA to specific locations to make cuts.
- Can use a fusion (chimera) of crRNA and tracrRNA to make a "guide RNA" (**gRNA**) of our choice = programmable

Jinek, M. *et al*, *Science* 17 Aug 2012 : 816-821





22

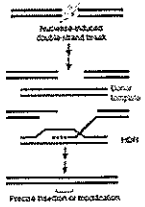


J.D. Sander and J.K. Joung, *Nat. Biotechnol.* 32, 347-355 (2014)

23

The Breakthrough(s)

- By supplying a DNA donor template, can use homology-directed repair to insert DNA of our choice
- In 2013, it was shown that CRISPR-Cas9 can be used in human and mouse cells. (Cong, L. *et al*. *Science* 15 Feb 2013: Vol. 339, Issue 6121, pp. 819-823)
- In the 4 years since, many CRISPR technologies have been developed for more precise genome editing, transcription control, epigenetic alterations, etc.



J.D. Sander and J.K. Joung, *Nat. Biotechnol.* 32, 347-355 (2014)

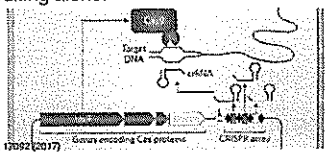
24

What does CRISPR refer to?

- Strictly speaking, CRISPR refers to the repeats.

- However, the spacers between the repeats are taken from invasive DNA sequences, and are of greater interest.
- A CRISPR array refers to the repeats and spacers together.

- In popular science, CRISPR has come to refer to the entire CRISPR-Cas system, or even only to the Cas proteins and their editing alone.

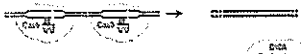


Barrangou, R. and Horvath P. *Nat. Microbiol.* 2, 17092 (2017)

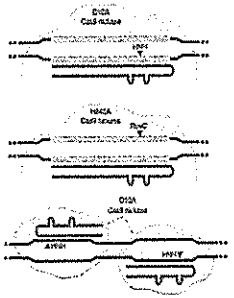
25

Techniques (Nickases)

- Using multiple gRNAs allows **large deletions and inversions** between DSBs



- **Nickases** can be created by mutations in the cleavage domains of the nucleases. *(D10A in RuvC and H840A in HNH)*
- Using two nickases w/ different gRNAs allows even greater specificity.
- Mutations in both domains gives "dead" Cas9 (**dCas9**), that binds w/out cleaving.

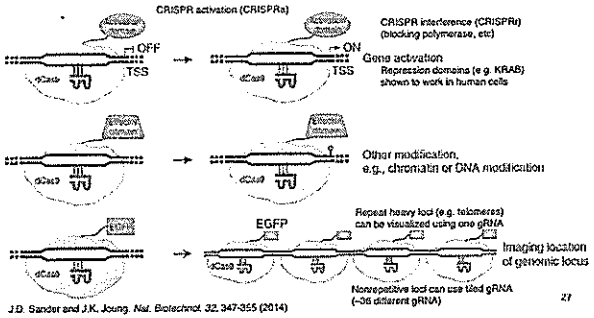


J.D. Sander and J.K. Joung, *Nat. Biotechnol.* 32, 347-355 (2014)

26

Techniques (dCas9)

- Mutations in both domains gives "dead" Cas9 (**dCas9**), that binds w/out cleaving. Fusion to heterologous effector domains possible.



J.D. Sander and J.K. Joung, *Nat. Biotechnol.* 32, 347-355 (2014)

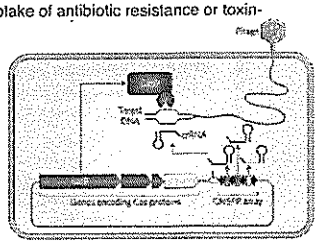
27

Techniques (Modifying Resistance)

- Valuable strains of bacteria can be given artificial CRISPR arrays providing resistance to bacteriophages, transposons, or different plasmids. Barrangou, J. E. *et al* Nature 468, 67–71 (04 November 2010)

- May allow us to prevent uptake of antibiotic resistance or toxin-encoding genes

- Used to resist viruses in eukaryotic cells (HIV infections, etc)
  Hu, W. *et al. Proc. Natl Acad. Sci. USA* 111, 11461–11466 (2015)



DNA infection.

J.D. Sander and J.K. Joung, *Nat. Biotechnol.* 32, 347-355 (2014)

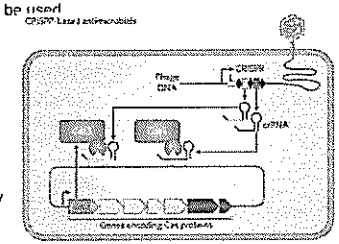
Techniques (Antimicrobials)

- Introducing self-targeting CRISPR arrays into bacteria can utilize endogenous CRISPR systems to result in cell death.
  Vercoe, R. B. *et al.* PLoS Genet. 9: e1003454 (2013).

- Engineered phages may be used

- Type I systems (e.g. Cas3) are particularly effective, due to the destructiveness of exonucleases.
  Gomaa, A.A., *et al.* 2014. *mBio* 5(1):e00928-13.

- Many pathogens natively encode Type I systems.



J.D. Sander and J.K. Joung, *Nat. Biotechnol.* 32, 347-355 (2014)

PAM Limitations

- One limitation of the CRISPR-Cas9 system is the need for PAM sequences near editing sites.

- Other CRISPR types have different PAM sequences, but are not as adaptable and efficient as Type II systems, like Cas9

- Luckily, new CRISPR-Cas systems still being discovered:
  - First Cas9 from archaea (unknown subtype)
  - CRISPR-CasX and –CasY from bacteria in groundwater (compact systems)
    - New PAMs upstream of protospacer (5' TCCN 3' in CasX and 5' TA 3' in CasY)

  Burstein, D., *et al*, Nature 542, 237–241 (09 February 2017)