

CIBCB 2017 Tutorial

An Introduction to CRISPR for Bioinformaticists

Wendy Ashlock

Ashlock & McGuinness Consulting, Inc.

August 23, 2017

Tutorial Outline

- 1 CRISPR Craze
- 2 What is CRISPR?
- 3 Bioinformatic Challenges
- 4 Resources

Tutorial Outline

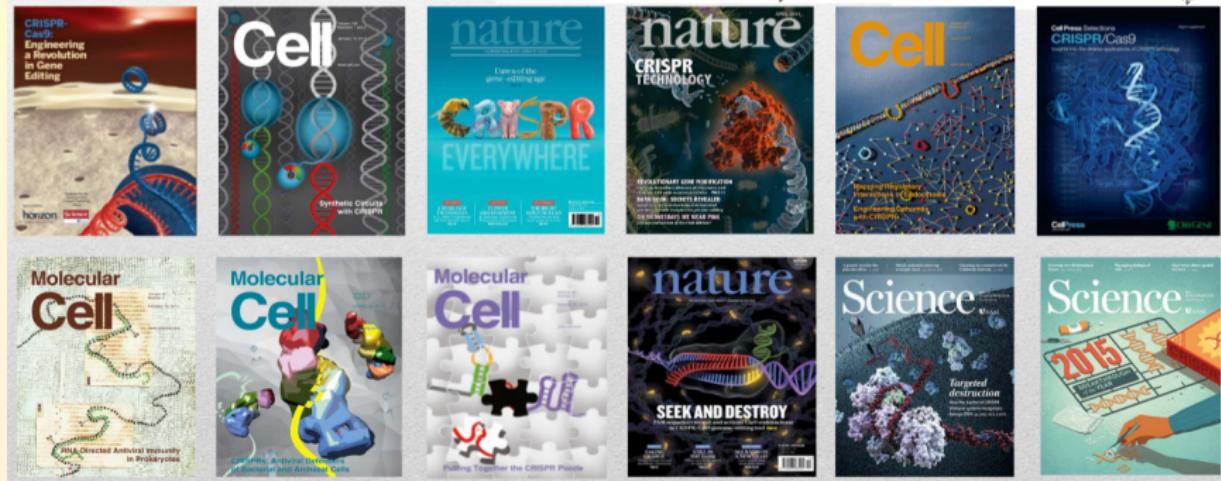
1 CRISPR Craze

2 What is CRISPR?

3 Bioinformatic Challenges

4 Resources

Scientific interest



- 100's of publications; 1000's of citations
- Supplanting other technologies (TALEN, ZFN)
- Scientific democratization (easy and inexpensive)

Industrial interest



- Commercial products (agriculture, biofuels, food, drug development, etc.)
- Large venture capital investments
- All scales (start-ups, mid-size, large, non-profits)

2016 Canada Gairdner International Awards



- Emmanuelle Charpentier (Germany/Sweden), Jennifer Doudna (USA), and Feng Zhang (USA) for development of CRISPR-CAS as a genome editing tool for eukaryotic cells



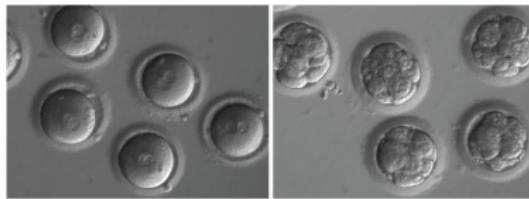
- Rodolphe Barrangou (USA) and Philippe Horvath (France) for establishing and characterizing CRISPR-Cas bacterial immune defence system

CRISPR in the New York Times this month

In Breakthrough, Scientists Edit a Dangerous Mutation From Genes in Human Embryos

[查看简体中文版](#) | [查看繁体中文版](#) | [Leer en español](#)

By PAM BELLUCK AUG. 2, 2017



Newly fertilized eggs before gene editing, left, and embryos after gene editing and a few rounds of cell division. A study published on Wednesday announced that edited human embryos can repair common and serious disease-causing gene mutations. Shoukhrat Mitalipov

Gene Editing Spurs Hope for Transplanting Pig Organs Into Humans

By GINA KOLATA AUG. 10, 2017



Piglets whose genes were edited to remove retroviruses, which could help clear the way for pig organs to be transplanted to humans. eGenesis

Tutorial Outline

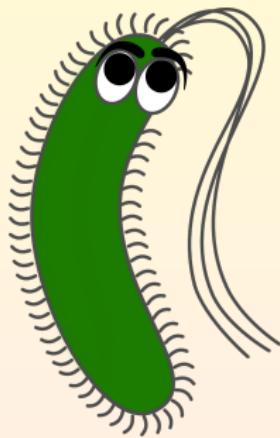
1 CRISPR Craze

2 What is CRISPR?

3 Bioinformatic Challenges

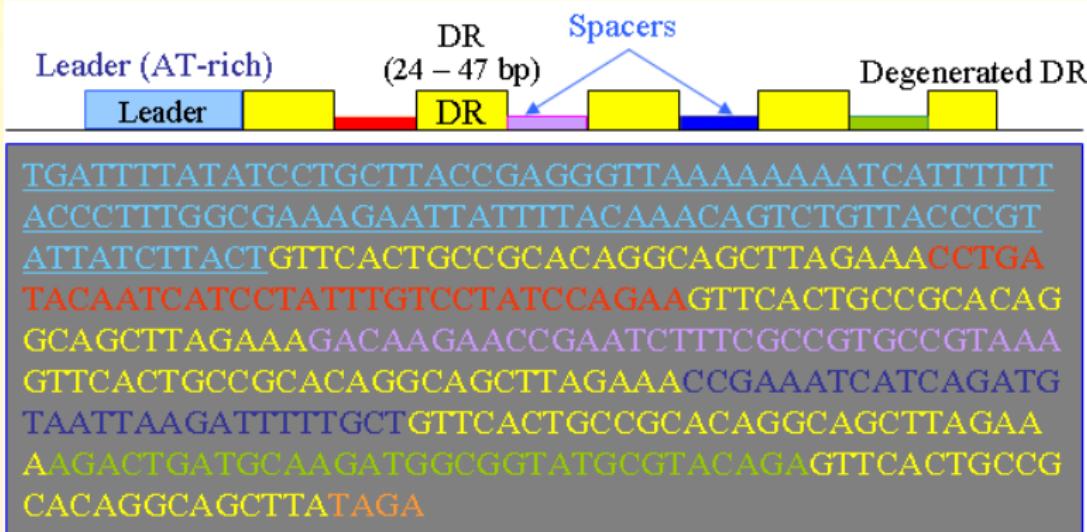
4 Resources

What is CRISPR?



Clustered
Regularly
Interspaced
Short
Palindromic
Repeats

What is CRISPR?



from CRISPRs web server: <http://crispr.i2bc.paris-saclay.fr/index.php?page=FAQs>

Palindromic

GTTC₁ACTGCCGCACAGGCAGCTT₂AGAAA (direct repeat)

TTT₁CTAAGCTGC₂CTGTGCCGGCAGTGAAC (reverse complement)

What is CRISPR?



<http://www.the-scientist.com/?articles.view/articleNo/41676/title/There-s-CRISPR-in-Your-Yogurt/>

Philippe Horvath and Rodolphe Barrangou, 2007

How CRISPR works – bacterial immunity

- **Adaptation**

- short pieces of DNA homologous to virus sequences are integrated into CRISPR loci
- repeat is duplicated

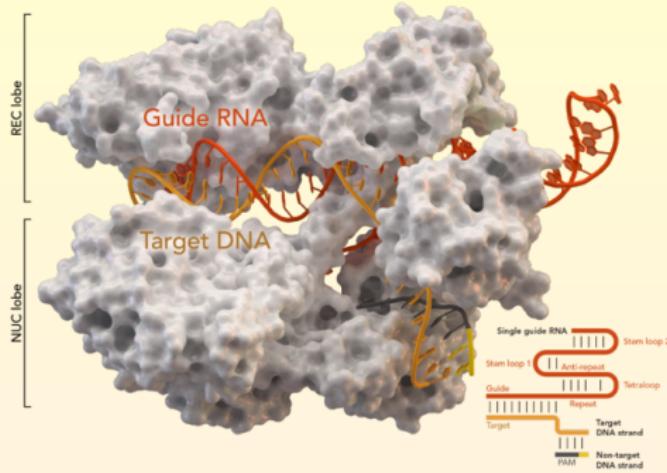
- **Expression**

- long transcript of a CRISPR locus is generated
- transcript is processed into short CRISPR RNAs (crRNAs)

- **Interference**

- crRNAs guide Cas proteins to the target sequences that match the spacers
- foreign DNA or RNA is targeted and cleaved

How CRISPR works – gene editing



by Thomas Splettstoesser (www.scistyle.com) [CC BY-SA 4.0] via Wikimedia Commons

- location for editing is specified by guide RNA
- sequence must be followed by PAM sequence
- Cas protein cuts the DNA
- cut is “repaired” using cell’s own machinery

PAM

- **Proto-spacer Adjacent Motif**
- 2-6 base pair DNA sequence immediately following target DNA
- Cas proteins will not successfully bind to target if it is not followed by the appropriate PAM sequence
- selection of spacer from viral DNA is determined by location of PAM
- Canonical PAM is 5'-NGG-3'
- PAMs differ between CRISPR-Cas variants

Bioinformatic problem: Find good CRISPR-Cas variant for gene target

Types of CRISPR



Type II :
precisely
target DNA
with Cas9 or
Cpf1



Type I: hit and
destroy DNA with
Cas3



Type III: cleave
either DNA or
RNA with Cas10

Uses for CRISPR technology

- **Disrupt** (knock out genes)
 - useful for studying gene function
- **Correct** (repair genes)
 - introduce a replacement sequence for sequence that was cut out
- **Gene Regulation** (CRISPRi)
 - use a dead Cas protein that doesn't cut
 - repress transcription by blocking transcription
 - activate transcription by fusing a transcriptional activator to Cas protein
- **Gene drive**

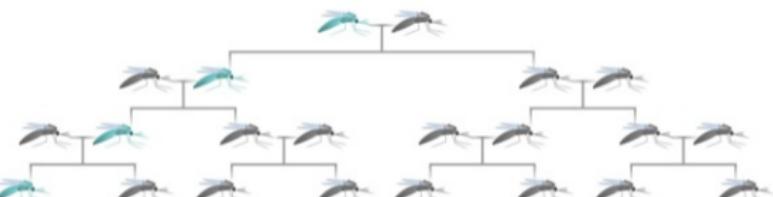
Gene Drive

Altered Gene Wild-Type



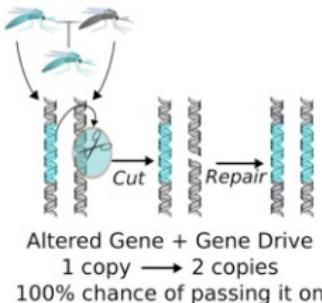
Altered Gene Only
1 copy inherited from 1 parent
50% chance of passing it on

Normal Inheritance



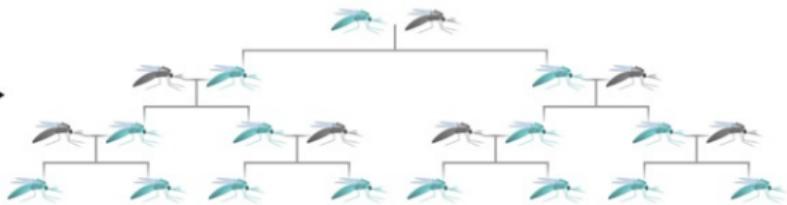
Altered gene does not increase

Gene Drive Wild-Type



Altered Gene + Gene Drive
1 copy → 2 copies
100% chance of passing it on

Gene Drive Inheritance

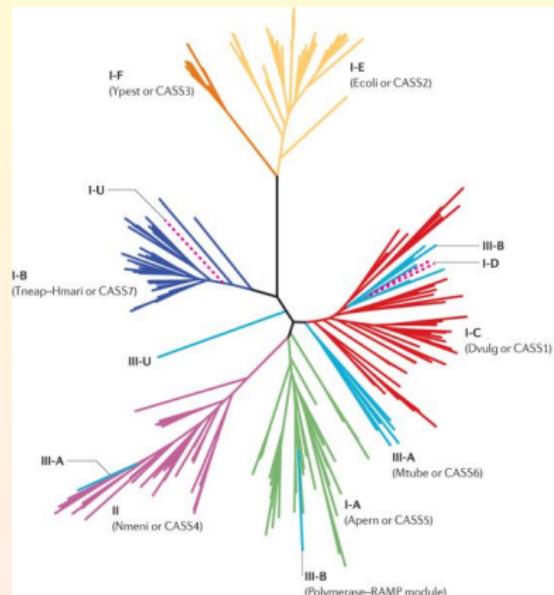


Altered gene is always inherited due to gene drive

from Scientific American blog, “Gene Drives” and CRISPR could revolutionize ecosystem management, Kevin Esvelt, George Church and Jeantine Lunshof

CRISPR-Cas systems

CRISPR-Cas systems are diverse and evolve quickly.



- 65 distinct Cas proteins, 23-45 families
- 8 distinct subtypes of CRISPR-Cas systems
- 44% of bacterial and archael genomes studied had CRISPR-Cas systems
- can be multiple CRISPR-Cas systems in a single genome
- can be different CRISPR-Cas systems in different strains of the same species

S. Makarova et al, Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*. 2011;9(6):467-477.

Tutorial Outline

1 CRISPR Craze

2 What is CRISPR?

3 Bioinformatic Challenges

4 Resources

Understanding natural CRISPR systems

- Origin – Casposons? (Krupovic et al, BMC Biology 2014 12:36)
- What drives spacer acquisition? Are the systems pruned?
 - Do CRISPR systems serve as a historic record of the cell?
- What about spacers that don't match any known viruses? Do CRISPRs have other functions? (DNA repair, gene expression, virulence)
- What don't all microbes have CRISPRs? (diversity of viruses?)
- Categorizing all CRISPR systems

Molecular Cell



Volume 62, Issue 1, 7 April 2016, Pages 137–147

Technology

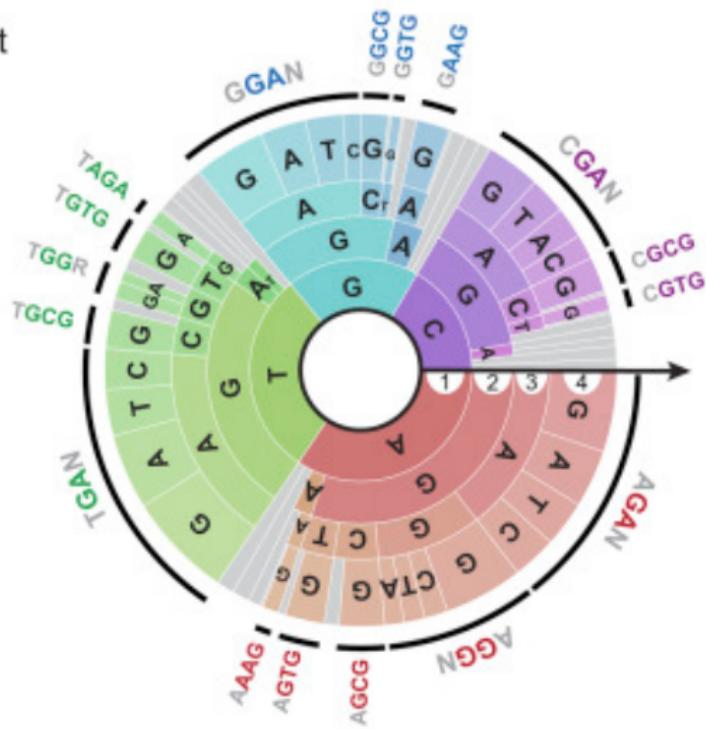
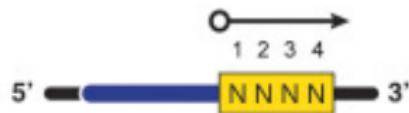
Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems

Ryan T. Leenay¹, Kenneth R. Maksimchuk¹, Rebecca A. Slotkowski¹, Roma N. Agrawal¹, Ahmed A. Gomaa^{1, 2}, Alexandra E. Briner³, Rodolphe Barrangou³, , , Chase L. Beisel¹, ,

PAM wheels

B

S. pyogenes
Type II-A Cas9 VQR variant



Determining gene function

20

Crutfeldt-Jakob disease
Gerstmann-Straussler disease
Insomnia, fatal familial
Pantothenate kinase associated neurodegeneration
Alagille syndrome
Corneal dystrophy
Inhibitor of DNA binding, dominant negative
Facial anomalies syndrome
Gigantism
Retinoblastoma
Rous sarcoma
Colon cancer
Galactosialidosis
Severe combined immunodeficiency
Hemolytic anemia
Obesity/hyperinsulinism
Pseudohypoparathyroidism, type Ia
McCune-Albright polyostotic fibrous dysplasia
Somatotrophinoma
Pituitary ACTH secreting adenoma
Shah-Waardenburg syndrome

63 million base pairs



Diabetes insipidus, neurohypophyseal
McKusick-Kaufman syndrome
Cerebral amyloid angiopathy
Thrombophilia
Myocardial infarction, susceptibility to
Huntington-like neurodegenerative disorder
Anemia, congenital dyserythropoietic
Acromesomelic dysplasia, Hunter-Thompson type
Brachydactyly, type C
Chondrodysplasia, Grebe type
Hemolytic anemia
Myeloid tumor suppressor
Breast cancer
Maturity Onset Diabetes of the Young, type 1
Diabetes mellitus, noninsulin-dependent
Graves disease, susceptibility to
Epilepsy, nocturnal frontal lobe and benign neonatal, type 1
Epiphyseal dysplasia, multiple
Electro-encephalographic variant pattern
Pseudohypoparathyroidism, type IB

Poster from Human Genome Project – Chromosome 20

Classifying CRISPR-Cas systems

"It should be emphasized that a **robust family classification of the Cas proteins**, many of which diverge rapidly, is not only a matter of convenient description but also a **basis for experimental validation** of the respective functional predictions. Therefore, it is important that this classification be continuously updated and revised when necessary, using new sequence and structure information combined with **state-of-the-art computational methods.**"

- S. Makarova et al, Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*. 2011;9(6):467-477.

Tutorial Outline

1 CRISPR Craze

2 What is CRISPR?

3 Bioinformatic Challenges

4 Resources

CRISPRs web server

- <http://crispr.i2bc.paris-saclay.fr/>
- CRISPRs database: includes 8939 CRISPRs from 7014 genomes
- CRISPRs finder: detect CRISPRs in sequences up to 67 Mbp
- CRISPR utilities: list of DR and spacers and genomes with and without CRISPRs
- BLAST CRISPR: BLASTs a sequence against direct repeats and spacers
- FlankAlign: useful for finding CRISPR homologs or validating questionable CRISPRs
- CRISPRtionary: helpful for analyzing sequences from multiple alleles derived from the same locus

CRISPI: a CRISPR interactive database

- <http://crispi.genouest.org/>
- database of genomes with CRISPRs, spacers, and direct repeats
- BLAST against database
- utility to find CRISPRs

WTSI Genome Editing

Find CRISPRs in our genome browser:



Find CRISPRs by gene using our table:

Pair			
Exon ID	Spacer	Status	Summary
ENSE000003666217	20	Complete	closest: None total_pairs: 1 max_distance: 1000
	3	Complete	closest: None total_pairs: 1 max_distance: 1000

Find CRISPRs by 20bp gRNA:

Sequence: AATAGTAGACATAAAAGTCT

Species: Human (GRCh37) Human (GRCh38) Mouse (GRCm38)

EnsEMBL In gene In exon

Find CRISPRs in genomic sequence:

IAAGGAATGTT**CCC**AATAGTAGACATAAAAGTCTTCG

Crispr ID	EnsEMBL	In gene
1106710403	13:32325087-32325109	No
1106710404	13:32325088-32325110	No
1106710405	13:32325110-32325132	No

Find off-targets by sequence:

Mouse (GRCm38)

Orientation

PAM Right (NGG)
 PAM Left (CCN)

Sequence	Ori
GTGTCACTGAACTTACTCT	par
GTGTCCCCAGAAACTTACTCT	par



[www.sanger.ac.uk/
htgt/wge/](http://www.sanger.ac.uk/htgt/wge/)

Cas-Database

- <http://www.rgenome.net/cas-database/>
- tool for designing crRNAs that will work for specific genes
- works with 5 vertebrate species; 5 plant species; 1 insect species; and 1 other species

Family: Cas_Cas6 (PF01881)

2 architectures 512 sequences 1 interaction 400 species 9 structures

Summary**Domain organisation****Clan****Alignments****HMM logo****Trees****Curation & model****Species****Interactions****Structures****Jump to... ↻**

enter ID/acc

**Summary: CRISPR associated protein Cas6**

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

No Wikipedia article **Pfam** InterPro

This tab holds the annotation information that is stored in the Pfam database. As we move to using Wikipedia as our main source of annotation, the contents of this tab will be gradually replaced by the Wikipedia tab.

CRISPR associated protein Cas6 [Provide feedback](#)

This group of families is one of several protein families that are always found associated with prokaryotic CRISPRs, themselves a family of clustered regularly interspaced short palindromic repeats, DNA repeats found in nearly half of all bacterial and archaeal genomes. These DNA repeat regions have a remarkably regular structure: unique sequences of constant size, called spacers, sit between each pair of repeats [1]. It has been shown that the CRISPRs are virus-derived sequences acquired by the host to enable them to resist viral infection. The Cas proteins from the host use the CRISPRs to mediate an antiviral response. After transcription of the CRISPR, a complex of Cas proteins termed Cascade cleaves a CRISPR RNA precursor in each repeat and retains the cleavage products containing the virus-derived sequence. Assisted by the helicase Cas3, these mature CRISPR RNAs then serve as small guide RNAs that enable Cascade to interfere with virus proliferation [2]. Cas5 contains an endonuclease motif, whose inactivation leads to loss of resistance, even in the presence of phage-derived spacers [3].

Literature references

- Haft DH, Selengut J, Mongodin EF, Nelson KE; PLoS Comput Biol. 2005;1:e60.: A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. [PUBMED:16292354](#) [EPMC:16292354](#)

**Example structure**

PDB entry [3QJL](#): One RAMP protein binding different RNA substrates

[View a different structure:](#)

3QJL

Questions?

Thank you for your attention!

Creative Commons Attribution ShareAlike 4.0 license



<http://creativecommons.org/licenses/by-sa/4.0/>

You are free to:

- **Share** – copy and redistribute the material in any medium or format
- **Adapt** – remix, transform, and build upon the material for any purpose, even commercially

Under the following terms:

- **Attribution** - You must give **appropriate credit**, provide a link to the license, and **indicate if changes were made**. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **No additional restrictions** - You may not apply legal terms or **technological measures** that legally restrict others from doing anything the license permits.