Imperial College London

What are the roles of CRISPR and Anti-CRISPR system in the dynamic of bacteria and bacteriophage in the infant gut?

Author: Jiqiu Wu
01618530
j.wu18@imperial.ac.uk

Supervisors:
Tim Barraclough t.barraclough@imperial.ac.uk Imperial College London
Stineke Van Houte c.van-houte@exeter.ac.uk The University of Exeter
Edze Westra e.r.westra@exeter.ac.uk The University of Exeter

MRes in Computational Methods in Ecology and Evolution, 2018-2019
Department of Life Science
Silwood Park Campus
Imperial College London

1 Title

What are the roles of CRISPR and Anti-CRISPR system in the dynamic of bacteria and bacteriophage in the infant gut?

2 Abstract

Gut microbiota, composed by bacteria, fungi and viruses, play an essential role in the health of human and animals. Gut microbiota in infant can take effort to the gut development in adult. There is a dynamic process between bacteria and bacteriophages in early life, however, the interaction between the two important composition is unclear. In this project, we will use computational methods to study if CRISPR and Anti-CRISPR take part in the process and what the role it takes based on data from public database.

3 Introduction

3.1 Gut microbiota has been associated with a great of diseases

The mammalian microbiome consists of bacteria, archaea, fungi and viruses.[1] They make up to 100 trillion cells, tenfold the number of human cells. They encode 100-fold genes than human genome. They have a profound influence on human physiology, nutrition and disease.[2] They form a bioreactor and produce bioactive compounds. These microbiota-derived compounds signal to distant organs including brain, liver, lung and so forth.[3]

In cecum and colon, dietary fibers, proteins and peptides escape digestion by host enzymes in the upper gut, and then metabolized by the microbes. Their production of nutrients and vitamins can defense against pathogens and maintain the immune system. In particularly, short chain fatty acids (SCFAs) help maintain gut barrier function, by reducing luminal pH and inhibiting some pathogenic microorganisms.[4]

In heart, gut micro-derived metabolites are recognized as contributors to atherogenesis. Specific trimethylamine (TMA)-containing dietary nutrients, such as choline, carnitine and phosphatidylcholine, can be used by gut microbes as a carbon fuel source. A waste product is TMA, which is carried via the portal circulation to the liver, where it is rapidly converted by a family of enzymes, host hepatic flavin monooxygenases (FMOs), into TMA N-oxide (TMAO).[5] Researches show TMAO play an obligatory role in atherosclerosis development.[6]

Microbiomes play a role in neurodevelopmental and mood disorders.[7] The beta-diversity of the gut microbiome in major depressive disorder (MDD) patients is significantly different from that of healthy controls, with significantly more Actinobacteria and less Bacteroidetes in MDD-associated microbial populations.[8] Differences in species richness and their diversity between autism spectrum disorder (ASD) and controls have been reported.[9] The gut microbiome of ASD patients have increased abundance and diversity of Clostridia species and a general increase in non-spore-forming anaerobes and microaerophilic bacteria compared to neurotypical

controls.[10]

As one part of the gut microbiome, the virome has a significant implication in health and disease.[11] Current studies suggest that the human gut virome is a highly individual but temporally stable collective, dominated by phages exhibiting a temperate lifestyle.[12] Although both eukaryotic and prokaryotic viruses are part of the microbiota, prokaryotic viruses, known as phages, dominate.[13] However, during inflammatory bowel diseases(IBD) the intestinal phage population is altered and transitions from an ordered state to a stochastic dysbiosis in murine.[14]

3.2 The microbiota of early life is important in adulthood

The microbiota of infants is a key factor for the development of the microbiome.[15] Epidemiological studies have shown that factors that alter bacterial communities in infants during childhood increase the risk for several diseases.[16] Childhood overweight is associated with the microbiota of early life in Denmark[17] and Norway[18]. Early gut microbiota was associated with infant growth rates.[19]

As for immune system, a diverse early life intestinal microbiota can inhibit the pathways that lead to allergic sensitization by multiple mechanisms.[20] Absence of crucial species within the microbiota results in an incomplete or altered maturation of the mammalian immune system, thus driving allergens.[21] A study in the United Kingdom suggests that colonization by particular Bifidobacterium species in newborns as young as 1-week-old was indicative of later allergy risk.[22] Another study in Canada shows that the absence of colonization by particular microbial species early in life can be a driving factor for allergic airway disease. A low abundance of the *Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia* (FLVR) genera at 3 months of age was associated with a higher risk of allergy and asthma development.[23] It has been revealed that children born by C-section are at a higher risk for autoimmune diseases, because of the lack of maternal vaginal microbiome.[24] However, some restoration toward vaginal-derived microbial status can be achieved by exposing newborns to vaginal microorganisms derived from their mothers.[25]

Besides, researches over the past few years reveal that the gut microbiome plays a role in basic neurogenerative processes and modulates many aspects of animal behavior in early life.[26] Partial depletion of rats associated microbiota for a short time with vancomycin can result to visceral hypersensitivity during adulthood.[27] Jochen et al. revealed significant social impairments in GF mice, which is similar with social cognition deficits observed in patients with neurodevelopmental disorders, indicating microbiota are crucial to the genesis of neurodevelopmental disorders of altered sociability.[28]

3.3 There is a dynamic process of microbiota in infants

The ecology of the bacterial microbiome increases in richness and diversity towards an adult-like composition.[29] Soon after birth, the bacterial microbiome rapidly switches

from predominantly facultative anaerobic bacteria to a diverse community of anaerobes.[30] Similarly, the virome is highly dynamic during infant development.[31] During the first months of life, the early infant gut bacteriophage virome is composed of a rich community of bacteriophages, the majority of which derive from the Caudovirales order. Subsequently, the bacteriophages decrease in richness and shifts towards a Microviridae-dominated community over the first 2 years of life. Thus, the infant virome and bacterial microbiome evolves in a dynamic trajectory during the early years of life.[32] Recent study indicates that early life is a unique window when the gut microbiome is most influential on immune development and other disease outcomes.[33] Many factors can affect this process, such as diet[34], antibiotic[35], and mode of delivery[36].

3.4 What is the role of CRISPR and Anti-CRISPR in the dynamic bacteria-bacteriophage process?

As the Red Queen hypothesis proposes that organisms must continually evolve new mechanisms of resistance to parasites to avoid extinction.[37] Bacteria have evolved a great of diverse strategies to defend themselves against phage, including restriction–modification enzymes that inactivate target DNA by cleavage, toxin–antitoxin modules that lead to phage abortive infection and CRISPR–Cas systems that target and inactivate specific nucleic acid sequences by cleavage. In response, phages have evolved various mechanisms to overcome these defenses, including expression of proteins that modify restriction sites[38] or degrade restriction– modification cofactors[39], antitoxin molecules that inhibit the activity of toxin-antitoxin abortive infection systems[40] and proteins that directly bind to and inactivate CRISPR–Cas machinery[41].

Clustered regularly interspaced short palindromic repeats (CRISPR), encoded by most archaea and many bacteria, are involved in the resistance to viruses.[42] The organization of CRISPR loci is very diverse.[43] However, generally it consists of two main components, the CRISPR locus whose function is genetic memory, and the Cas proteins, which functions as the catalytic core of the system. The CRISPR array is composed of repetitive sequences(repeats) that are separated by variable sequences(spacers) derived from invading mobile genetic elements, such bacteriophages and plasmids.[44] Beyond the adaptive immunity of bacteria and archaea, CRISPR/Cas modulates other processes[45], such as, Endogenous gene regulation[46], regulation of bacterial virulence[47] and DNA repair[48].

The CRISPR/Cas systems mediate immunity to invading genetic elements via a process of three stages, adaption, expression and interference.[49] During the first stage, proto-spacer-adjacent motifs (PAMs) is likely to play a role in the spacer acquisition.[50] Short pieces of DNA homologous to virus or plasmid sequences are integrated into the CRISPR loci as precursors(proto-spacers).[51] During the second stage, expression, the long primary transcript of a CRISPR locus (pre-crRNA) is generated and processed into short crRNAs. The last stage is interference, the crRNAs guide the complexes of Cas proteins to the complementary virus or plasmid target sequences that match the spacers. And then, the cleavage of the sequences will be catalyzed by various enzymes based on

the classification that the bacteria contain.[52]

The classification scheme encompasses two classes and six types. Class 1 CRISPR-Cas systems are defined by the presence of a multi-subunit crRNA–effector complex, including type I (I-A, I-B, I-C, I-D, I-E, I- F, and I-U subtypes) and type III (III-A, III-B, III-C, and III-U subtypes) CRISPR-Cas systems, as well as the putative new type IV.[53] Class 2 CRISPR-Cas systems are defined by the presence of a single subunit crRNA–effector module, including type II (II-A, II-B, and II-C subtypes), as well as two putative new types V (V-A, V-B, V-C, and V-U subtypes) and VI (VI-A, VI-B, and VI-U subtypes).[54]

Anti-CRISPR was demonstrated to allow that phage to overcome the CRISPR/Cas system[55]. In 2013, Bondy-Denomy et al. first found anti-CRISPR proteins. Five distinct proteins (AcrF1, AcrF2, AcrF3, AcrF4 and AcrF5) were shown to inactive the type I-F CRISPR/Cas system in *Pseudomonas* spp. phages.[55] Subsequently, more anti-CRISPR proteins (AcrE1, AcrE2, AcrE3 and AcrE4) with anti-I-F[56], anti-I-E[57], anti-II-A[58], anti-II-C[41], and anti-VI-B[59] activities were identified.

There are numerous mechanisms that anti-CRISPR proteins could inhibit the CRISPR-Cas activity. For example, anti-CRISPRs could prevent the acquisition of new CRISPR spacers, block expression of Cas proteins[60], inhibit crRNA transcription or processing, prevent the assembly of the active CRISPR-Cas complex, inhibit binding to the foreign DNA element, or block cleavage activity.[61]

Thus, CRISPR-Cas and anti-CRISPR of the metagenomic and 16S rRNA sequences from eight healthy infants (four twin pairs) between 0-2 years old[32] will be studied. Computational approaches will be used to detect CRISPR-Cas and anti-CRISPR sequences. What kind of CRISPR-Cas and anti-CRISPR are contained and during the dynamic process of bacteria-bacteriophage, if the CRISPR and Anti-CRISPR systems change, if they play a role will be answered. If the switch of lysogenic and lytic life style drives the dynamic process will be studied. Besides, in terms of previous researches, there is no CRISPR system in Bacteroidetes and no anti-CRISPR in Caudovirales[62] which are dominant in gut, new insight is wished.

Extra:
Maybe we can study the known interactions between anti-CRISPR proteins and their binding site of CRISPR-Cas, and then predict the novel anti-CRISPR proteins based on the numerous CRISPR-Cas data.[63]

4.Materials and Methods (all the methods need further consideration)
4.1 Device
Laptop with Linux operation system and Server from Samraat's Lab(potential).
4.2 Data
Customized metagenomic and 16S rRNA sequences from eight healthy infants (four

twin pairs) between 0-2 years old[32] can be downloaded at:
https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP058399;
http://pathology.wustl.edu/virusseeker/data/VirusDBNT_20131107_ID98.tgz;
http://pathology.wustl.edu/virusseeker/data/VirusDBNR_20131107_ID98.tgz;
CRISPR-Cas data[64] can be downloaded at:
http://www.microbiome-bigdata.com/CRISPRminer/;
Anti-CRISPR data[65] can be downloaded at:
http://bioanalysis.otago.ac.nz/Anti-CRISPRdb/
Prophage sequence data[66] can be downloaded at:
http://phaster.ca/.

4.3 Operations have been done previous researchers

Sample and sequencing have been done. Overlapping reads have been generated. Low quality nucleotides were trimmed and discarded at a quality filter of Q30 Phred quality score.

4.4 Virome sequence processing

The methods were described[32], but the newest database will be used.

Blast the clean data to the newest NCBI NT and NR database. CD-HIT[67] will be used to minimize the sequence redundancy (98% and over 98% of the sequence length). Sequencing reads were queried against the customized viral database sequentially using BLASTn (e-value cutoff 1E−10), followed by BLASTx (e-value cutoff 1E−3). False positive viral sequences were filtered by sequentially querying the candidate viral reads against the NCBI NT database using MegaBLAST (e-value cutoff 1E−10), BLASTn (e-value cutoff 1E−10), and the NCBI NR database using BLASTx (e-value cutoff 1E−3) to remove sequences that have a top BLAST hit corresponding to a nonviral sequence (for example, human, fungal, etc.) as previously described.[68] The taxonomic assignment for sequencing reads was determined by the taxonomy ID of the top BLAST result. Bacteriophage species taxonomic assignment was determined using the lowest-common-ancestor algorithm implemented in Megan[69] with the following parameters: Min Support: 1, Min Score: 40.0, Max Expected: 0.01, Top Percent: 10.0, Min-Complexity filter: 0.44.

4.5 Virome analysis

The methods were described[32], but the newest database will be used.

Sequencing reads have been rarefied. Heatmap will be plotted by R package[70] to show the number of reads detected for a given viral taxon. Ecological analyses including richness and diversity measurements (Shannon index, Bray-Curtis dissimilarity), agglomerative hierarchical clustering and rarefaction curve analyses were performed using the vegan R package[71]. Rarefaction curves were performed using 500 permutations. Principal-coordinate analyses (PCoA) was performed using Emperor[72].

4.6 Bacterial 16S rRNA gene analysis

16S analysis was performed with QIIME (Quantitative Insights Into Microbial Ecology[73]. Sequences were quality filtered at Q20 Phred quality score and demultiplexed. Sequences were assigned to closed reference operational taxonomic units (OTUs) at a 97% identity threshold using the Greengenes database[74]. Investigators were blinded to the group allocation (i.e., age, twin pair, time point) during processing

of 16S rRNA gene sequences up to OTU taxonomic assignment. To account for inter-sample sequencing depth variability, all samples were rarefied to 10,000 reads per sample (10 iterations), which exceed the generally accepted minimum sequencing depth previously described[75]. Consistent results were obtained in the 16S analyses across all iterations. Thus, results obtained from a representative iteration are shown. Alpha diversity (Faith's phylogenetic diversity), OTU richness and UniFrac distance were calculated using QIIME. Rarefaction curves were performed using the vegan R package using 500 permutations. PCoA plots were visualized using Emperor.

4.7 CRISPR analysis

The methods were described[76], but the newest database will be used.

4.7.1 Identify CRSIPR spacers using newest CRISPR spacer database

The newest spacer database[64] will be used to query all the contigs from the metagenomic data using BLASTN with e-value threshold of 1E-4. To obtain a nonredundant set, contigs <50 kb will be clustered using uclust[77] requiring 80% identity over at least 80% of the contig length. Contigs >50 kb will be clustered using BLASTN with the same parameters.

4.7.2 Identification of phage contigs

Gene prediction using Glimmer[78] was performed on the nonredundant MGE contigs. Predicted genes were then searched against the Conserved Domains Database (CDD)[79] using rpsblast with e-value threshold 0.05. An MGE contig was deemed a phage contig if its size was 10 kb or more and the CDD description of at least one of its annotated genes contained the word “phage,” “holin,” or “tail.”

4.8 Anti-CRISPR analysis

There are two strategies for detecting Anti-CRISPR. First, using the Anti-CRISPR database[65] to query the metagenomic data using BLASTN with e-value threshold of 1E-4. Second, using the aca gene sequences combined from all the previous researches to query the metagenomic data using the same parameters.

4.9 Detection of the lysogenic phages

There are two principles[76] that can lead to the Python script. First, the lysogenic phages should be combined with integrase and recombinase. Second, the lysogenic phages should be integrated into bacteria with flanking sequences. And for the left phages, we regard them as lytic phages.

4.10 Alternative tools

PHASTER[66], CRISPRFinder[80], and CRISPI[81].

4.11 Gantt chart

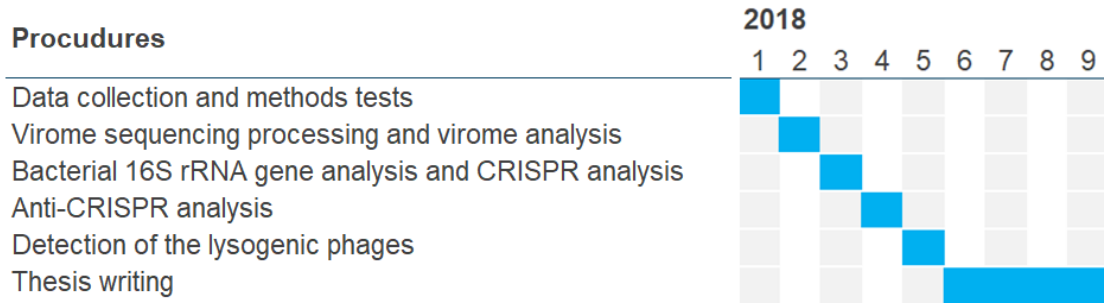

Figure1: The Gantt chart for the project

5 Expected results

5.1 The bacteria and bacteriophage development in early life

At different timepoint, the abundance and richness of bacteria (maybe at genus level) and bacteriophage (maybe at family level) can be calculated. And heatmaps can be plotted by R based on there results.

5.2 The change of CRISPR system in the gut of infants

We can get the number, class, and subtype of CRISPR system at every timepoint. During the time, we can see the change of them.

5.3 The change of Anti-CRISPR system in the gut of infants

We can get the number, class, and subtype of Anti-CRISPR system at every timepoint. During the time, we can see the change of them.

5.4 The switch condition of lysogenic and lytic phages

We can get the number of lysogenic and lytic phages every timepoint. During the time, we can see the change of them.

6 Discussion

Metagenomics analysis provide a chance for us to study the CRISPR and Anti-CRISPR system dynamic process behind the bacteria and bacteriophages dynamic process in the early life. If the CRISPR and Anti-CRISPR drive the bacteria and bacteriophages dynamic process will be discussed. If they affect the health of human will be discussed. Or there is other role they play in the gut community. Maybe there will be some other interesting results we can discuss during the data mining process, who knows?

Still thinking……

References:

[1] Lynch S V, Pedersen O. The human intestinal microbiome in health and disease[J]. New England Journal of Medicine, 2016, 375(24): 2369-2379.
[2] Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing[J]. nature, 2010, 464(7285): 59.
[3] Schroeder B O, Bäckhed F. Signals from the gut microbiota to distant organs in physiology and disease[J]. Nature medicine, 2016, 22(10): 1079.
[4] Macfarlane G T, Macfarlane S. Bacteria, colonic fermentation, and gastrointestinal health[J]. Journal of AOAC International, 2012, 95(1): 50-60.
[5] Zhu W, Gregory J C, Org E, et al. Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk[J]. Cell, 2016, 165(1): 111-124.
[6] Wang Z, Klipfell E, Bennett B J, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease[J]. Nature, 2011, 472(7341): 57.
[7] De Theije C G M, Bavelaar B M, Lopes da Silva S, et al. Food allergy and food-based therapies in neurodevelopmental disorders[J]. Pediatric Allergy and Immunology, 2014, 25(3): 218-226.
[8] Zheng P, Zeng B, Zhou C, et al. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism[J]. Molecular psychiatry, 2016, 21(6): 786.
[9] Williams B L, Hornig M, Parekh T, et al. Application of novel PCR-based methods for detection, quantitation, and phylogenetic characterization of Sutterella species in intestinal biopsy samples from children with autism and gastrointestinal disturbances[J]. MBio, 2012, 3(1): e00261-11.
[10] De Angelis M, Piccolo M, Vannini L, et al. Fecal microbiota and metabolome of children with autism and pervasive developmental disorder not otherwise specified[J]. PloS one, 2013, 8(10): e76993.
[11] Virgin H W. The virome in mammalian physiology and disease[J]. Cell, 2014, 157(1): 142-150.
[12] Ogilvie L A, Jones B V. The human gut virome: a multifaceted majority[J]. Frontiers in microbiology, 2015, 6: 918.
[13] Feiner R, Argov T, Rabinovich L, et al. A new perspective on lysogeny: prophages as active regulatory switches of bacteria[J]. Nature Reviews Microbiology, 2015, 13(10): 641.
[14] Duerkop B A, Kleiner M, Paez-Espino D, et al. Murine colitis reveals a disease-associated bacteriophage community[J]. Nature microbiology, 2018, 3(9): 1023.
[15] Stewart C J, Ajami N J, O'Brien J L, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study[J]. Nature, 2018, 562(7728): 583.
[16] Tamburini S, Shen N, Wu H C, et al. The microbiome in early life: implications for health outcomes[J]. Nature medicine, 2016, 22(7): 713.
[17] Ajslev T A, Andersen C S, Gamborg M, et al. Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics[J]. International journal of obesity, 2011, 35(4): 522.
[18] Stanislawski M A, Dabelea D, Wagner B D, et al. Gut Microbiota in the First 2 Years of Life and the Association with Body Mass Index at Age 12 in a Norwegian Birth Cohort[J]. mBio, 2018, 9(5): e01751-18.
[19] White R A, Bjørnholt J V, Baird D D, et al. Novel developmental analyses identify longitudinal patterns of early gut microbiota that affect infant growth[J]. PLoS

computational biology, 2013, 9(5): e1003042.
[20] Reynolds L A, Finlay B B. Early life factors that affect allergy development[J]. Nature Reviews Immunology, 2017, 17(8): 518.
[21] Noverr M C, Huffnagle G B. The 'microflora hypothesis' of allergic diseases[J]. Clinical & Experimental Allergy, 2005, 35(12): 1511-1520.
[22] Ismail I H, Boyle R J, Licciardi P V, et al. Early gut colonization by Bifidobacterium breve and B. catenulatum differentially modulates eczema risk in children at high risk of developing allergic disease[J]. Pediatric Allergy and Immunology, 2016, 27(8): 838-846.
[23] Arrieta M C, Stiemsma L T, Dimitriu P A, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma[J]. Science translational medicine, 2015, 7(307): 307ra152-307ra152.
[24] Sevelsted A, Stokholm J, Bønnelykke K, et al. Cesarean section and chronic immune disorders[J]. Pediatrics, 2015, 135(1): e92-e98.
[25] Dominguez-Bello M G, De Jesus-Laboy K M, Shen N, et al. Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer[J]. Nature medicine, 2016, 22(3): 250.
[26] Sharon G, Sampson T R, Geschwind D H, et al. The central nervous system and the gut microbiome[J]. Cell, 2016, 167(4): 915-932.
[27] O'Mahony S M, Felice V D, Nally K, et al. Disturbance of the gut microbiota in early-life selectively affects visceral pain in adulthood without impacting cognitive or anxiety-related behaviors in male rats[J]. Neuroscience, 2014, 277: 885-901.
[28] Desbonnet L, Clarke G, Shanahan F, et al. Microbiota is essential for social development in the mouse[J]. Molecular psychiatry, 2014, 19(2): 146.
[29] Yatsunenko T, Rey F E, Manary M J, et al. Human gut microbiome viewed across age and geography[J]. nature, 2012, 486(7402): 222.
[30] Koenig J E, Spor A, Scalfone N, et al. Succession of microbial consortia in the developing infant gut microbiome[J]. Proceedings of the National Academy of Sciences, 2011, 108(Supplement 1): 4578-4585.
[31] Lim E S, Wang D, Holtz L R. The bacterial microbiome and virome milestones of infant development[J]. Trends in microbiology, 2016, 24(10): 801-810.
[32] Lim E S, Zhou Y, Zhao G, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants[J]. Nature medicine, 2015, 21(10): 1228.
[33] Cahenzli J, Köller Y, Wyss M, et al. Intestinal microbial diversity during early-life colonization shapes long-term IgE levels[J]. Cell host & microbe, 2013, 14(5): 559-570.
[34] Minot S, Sinha R, Chen J, et al. The human gut virome: inter-individual variation and dynamic response to diet[J]. Genome research, 2011, 21: 1616-1625.
[35] Yassour M, Vatanen T, Siljander H, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability[J]. Science translational medicine, 2016, 8(343): 343ra81-343ra81.
[36] Dominguez-Bello M G, Costello E K, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns[J]. Proceedings of the National Academy of Sciences, 2010, 107(26): 11971-11975.
[37] Van Valen L. A new evolutionary law[J]. Evol Theory, 1973, 1: 1-30.
[38] Krüger D H, Bickle T A. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts[J]. Microbiological reviews, 1983, 47(3): 345.
[39] Studier F W, Movva N R. SAMase gene of bacteriophage T3 is responsible for

overcoming host restriction[J]. Journal of virology, 1976, 19(1): 136-145.
[40] Otsuka Y, Yonesaki T. Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins[J]. Molecular microbiology, 2012, 83(4): 669-681.
[41] Pawluk A, Amrani N, Zhang Y, et al. Naturally occurring off-switches for CRISPR-Cas9[J]. Cell, 2016, 167(7): 1829-1838. e9.
[42] Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes[J]. Science, 2007, 315(5819): 1709-1712.
[43] Godde J S, Bickerton A. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes[J]. Journal of Molecular evolution, 2006, 62(6): 718-729.
[44] Karginov F V, Hannon G J. The CRISPR system: small RNA-guided defense in bacteria and archaea[J]. Molecular cell, 2010, 37(1): 7-19.
[45] Westra E R, Buckling A, Fineran P C. CRISPR–Cas systems: beyond adaptive immunity[J]. Nature reviews Microbiology, 2014, 12(5): 317.
[46] Kaiser D, Robinson M, Kroos L. Myxobacteria, polarity, and multicellular morphogenesis[J]. Cold Spring Harbor perspectives in biology, 2010: a000380.
[47] Louwen R, Horst-Kreft D, De Boer A G, et al. A novel link between Campylobacter jejuni bacteriophage defence, virulence and Guillain–Barré syndrome[J]. European journal of clinical microbiology & infectious diseases, 2013, 32(2): 207-226.
[48] Babu M, Beloglazova N, Flick R, et al. A dual function of the CRISPR–Cas system in bacterial antivirus immunity and DNA repair[J]. Molecular microbiology, 2011, 79(2): 484-502.
[49] Makarova K S, Haft D H, Barrangou R, et al. Evolution and classification of the CRISPR–Cas systems[J]. Nature Reviews Microbiology, 2011, 9(6): 467.
[50] Mojica F J M, Díez-Villaseñor C, García-Martínez J, et al. Short motif sequences determine the targets of the prokaryotic CRISPR defence system[J]. Microbiology, 2009, 155(3): 733-740.
[51] Garneau J E, Dupuis M È, Villion M, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA[J]. Nature, 2010, 468(7320): 67.
[52] Van der Oost J, Jore M M, Westra E R, et al. CRISPR-based adaptive and heritable immunity in prokaryotes[J]. Trends in biochemical sciences, 2009, 34(8): 401-407.
[53] Makarova K S, Zhang F, Koonin E V. SnapShot: class 1 CRISPR-Cas systems[J]. Cell, 2017, 168(5): 946-946. e1.
[54] Shmakov S, Smargon A, Scott D, et al. Diversity and evolution of class 2 CRISPR–Cas systems[J]. Nature Reviews Microbiology, 2017, 15(3): 169.
[55] Bondy-Denomy J, Pawluk A, Maxwell K L, et al. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system[J]. Nature, 2013, 493(7432): 429.
[56] Pawluk A, Staals R H J, Taylor C, et al. Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species[J]. Nature microbiology, 2016, 1(8): 16085.
[57] Pawluk A, Bondy-Denomy J, Cheung V H W, et al. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa[J]. MBio, 2014, 5(2): e00896-14.
[58] Rauch B J, Silvis M R, Hultquist J F, et al. Inhibition of CRISPR-Cas9 with bacteriophage proteins[J]. Cell, 2017, 168(1-2): 150-158. e10.
[59] Smargon A A, Cox D B T, Pyzocha N K, et al. Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27

and Csx28[J]. Molecular cell, 2017, 65(4): 618-630. e7.
[60] Wang X, Yao D, Xu J G, et al. Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3[J]. Nature structural & molecular biology, 2016, 23(9): 868.
[61] Maxwell K L. The anti-CRISPR story: a battle for survival[J]. Molecular cell, 2017, 68(1): 8-14.
[62] Pawluk A, Davidson A R, Maxwell K L. Anti-CRISPR: discovery, mechanism and function[J]. Nature Reviews Microbiology, 2018, 16(1): 12.
[63] Alipanahi B, Delong A, Weirauch M T, et al. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning[J]. Nature biotechnology, 2015, 33(8): 831.
[64] Zhang F, Zhao S, Ren C, et al. CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions[J]. Communications biology, 2018, 1(1): 180.
[65] Dong C, Hao G F, Hua H L, et al. Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins[J]. Nucleic acids research, 2017, 46(D1): D393-D398.
[66] Arndt D, Grant J R, Marcu A, et al. PHASTER: a better, faster version of the PHAST phage search tool[J]. Nucleic acids research, 2016, 44(W1): W16-W21.
[67] Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences[J]. Bioinformatics, 2006, 22(13): 1658-1659.
[68] Zhao G, Krishnamurthy S, Cai Z, et al. Identification of novel viruses using VirusHunter--an automated data analysis pipeline[J]. PLoS One, 2013, 8(10): e78470.
[69] Huson D H, Mitra S, Ruscheweyh H J, et al. Integrative analysis of environmental sequences using MEGAN4[J]. Genome research, 2011: gr. 120618.111.
[70] McMurdie P J, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data[J]. PloS one, 2013, 8(4): e61217.
[71] OHara R B, Simpson G L, Solymos P, et al. Vegan: community ecology package. R package version 2.0-10[J]. 2013.
[72] Vázquez-Baeza Y, Pirrung M, Gonzalez A, et al. EMPeror: a tool for visualizing high-throughput microbial community data[J]. Gigascience, 2013, 2(1): 16.
[73] Caporaso J G, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data[J]. Nature methods, 2010, 7(5): 335.
[74] McDonald D, Price M N, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea[J]. The ISME journal, 2012, 6(3): 610.
[75] Caporaso J G, Lauber C L, Walters W A, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample[J]. Proceedings of the national academy of sciences, 2011, 108(Supplement 1): 4516-4522.
[76] Stern A, Mick E, Tirosh I, et al. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome[J]. Genome research, 2012: gr. 138297.112.
[77] Edgar R C. Search and clustering orders of magnitude faster than BLAST[J]. Bioinformatics, 2010, 26(19): 2460-2461.
[78] Salzberg S L, Delcher A L, Kasif S, et al. Microbial gene identification using interpolated Markov models[J]. Nucleic acids research, 1998, 26(2): 544-548.
[79] Marchler-Bauer A, Lu S, Anderson J B, et al. CDD: a Conserved Domain Database for the functional annotation of proteins[J]. Nucleic acids research, 2010, 39(suppl_1): D225-D229.
[80] Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered

regularly interspaced short palindromic repeats[J]. Nucleic acids research, 2007, 35(suppl_2): W52-W57.

[81] Rousseau C, Gonnet M, Le Romancer M, et al. CRISPI: a CRISPR interactive database[J]. Bioinformatics, 2009, 25(24): 3317-3318.