Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation

Devaki Bhaya,[1] Michelle Davison,[1,2] and Rodolphe Barrangou[3]

[1]Carnegie Institution for Science, Department of Plant Biology, Stanford, California 94305; email: dbhaya@stanford.edu

[2]Department of Biology, Stanford University, Stanford, California 94305; email: shellblu@stanford.edu

[3]DANISCO, USA, Inc., Madison, Wisconsin 53716; email: rodolphe.barrangou@danisco.com

Annu. Rev. Genet. 2011. 45:273–97

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi: 10.1146/annurev-genet-110410-132430

Copyright © 2011 by Annual Reviews. All rights reserved

0066-4197/11/1201-0273$20.00

Keywords

coevolution, immunity, interference, RNAi, bacteriophage

Abstract

Bacteria and archaea have evolved defense and regulatory mechanisms to cope with various environmental stressors, including virus attack. This arsenal has been expanded by the recent discovery of the versatile CRISPR-Cas system, which has two novel features. First, the host can specifically incorporate short sequences from invading genetic elements (virus or plasmid) into a region of its genome that is distinguished by clustered regularly interspaced short palindromic repeats (CRISPRs). Second, when these sequences are transcribed and precisely processed into small RNAs, they guide a multifunctional protein complex (Cas proteins) to recognize and cleave incoming foreign genetic material. This adaptive immunity system, which uses a library of small noncoding RNAs as a potent weapon against fast-evolving viruses, is also used as a regulatory system by the host. Exciting breakthroughs in understanding the mechanisms of the CRISPR-Cas system and its potential for biotechnological applications and understanding evolutionary dynamics are discussed.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPR): hallmark of CRISPR-Cas systems

CRISPR-associated proteins (Cas): diverse types of proteins encoded by *cas* genes in the vicinity of CRISPRs

Bacteria and archaea have evolved to cope and thrive as communities in dynamic environments that are stressful and fluctuating. These environmental stressors can be both abiotic (e.g., nonoptimal temperatures or nutrient levels, redox stress) and biotic (e.g., toxins, viruses, transmissible genetic elements) in nature. The abundant presence of viruses in almost all environments is a constant threat to the survival of bacteria and archaea (2, 87, 93, 107). Furthermore, viruses can have high rates of mutation and recombination, so a successful defense system has to be multilayered and have the ability to deal with variable and fast-evolving predators (2, 41, 66). In this context, the discovery of a novel adaptive defense system in bacteria and archaea has generated great interest and encouraged research into the molecular basis of the mechanism of action, consequently there have been several recent breakthroughs. The CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated proteins) defense comprises a multistep process by which specific small fragments of foreign nucleic acids are first recognized as being nonself and incorporated into the host genome between short DNA repeats. Subsequently, these fragments or spacers, in conjunction with host Cas proteins, are used as a surveillance and adaptive immune system by which incoming foreign nucleic acids are recognized and destroyed or possibly silenced. The CRISPR-Cas system has primarily been investigated in its defensive role against foreign DNA (viruses and plasmids), but its versatile, modular architecture may allow it to play a regulatory role in host cells. The apparent parallels to the eukaryotic RNA interference (RNAi) system (see sidebar, CRISPRs and RNAi: Common Denominators and Differences) and the potential uses in biotechnology have also generated interest. The CRISPR-Cas system provides a unique opportunity to observe and model coevolution between host and virus in natural environments or in controlled settings because acquisition and immunity occur on short time scales and evidence of past genetic aggressions can be deduced in some cases. Finally, the ability to dynamically acquire foreign DNA and subsequently use it to fight off invading genetic material has elements of an acquired and heritable immunity system, reflecting a Lamarckian mode of evolution (64).

We begin with a personal perspective of the discovery of the CRISPR-Cas system and then synthesize results in this fast-moving field, including an attempt to categorize the bewildering variety of Cas proteins and their functions. We focus on bacteria, although studies with archaea, most of which contain a CRISPR-Cas system, have provided important insights into its mode of action, diversity and unique

CRISPRS AND RNAI: COMMON DENOMINATORS AND DIFFERENCES

The hypothesis that CRISPR/Cas systems might be an adaptive immune system was based on in silico analyses that also hinted at an analogy to the eukaryotic RNA interference (RNAi) mechanism (74). Although there are common denominators between crRNA and RNAi-based mechanisms of action, there are also significant differences. The molecular commonalities between CRISPR and RNAi primarily consist of the fact that both are mediated by small noncoding RNAs which in conjunction with a ribo-nucleoprotein-complex target sequence-specific cleavage of nucleic acids. There are functional similarities between the proteins involved in the biogenesis of small interfering RNAs, as well as mechanistic and structural commonalities between the eukaryotic RNA-induced silencing complex (RISC), and the CRISPR-associated complex for antiviral defense (CASCADE). One notable difference is the fact that the primary target for CRISPR-interference is dsDNA, although mRNA can be targeted by some CRISPR/Cas systems. Some key proteins do not have direct functional/ structural equivalents, notably the eukaryotic Argonaute (ARO), and some of the CRISPR universal and "signature" proteins. Overall, although CRISPR is arguably most similar to PIWI-interacting RNAs (piRNAs), which protect genome integrity from parasites such as transposons (103), further studies that examine the molecular basis for CRISPR-based interference will likely expand the list of mechanistic idiosyncrasies and notable differences. Nevertheless, evidence of small RNA-based defense and regulatory systems in all three domains of life should foster research that transcend these boundaries.

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

characteristics (for reviews with a focus on Archaea, see 32, 69, 99, 108). We present new evidence of the regulatory role of CRISPR-Cas systems in bacteria, which functions via small noncoding RNAs. The evolutionary implications of this rapidly evolving, heritable immune system in prokaryotes and the opportunity it affords to study the coevolution of viruses and hosts are discussed. We include a section on progress in engineering the CRISPR-Cas system for biotechnological and epidemiological applications. Several recent reviews have described the experiments that led to the discovery of this new defense system (5, 24, 49, 58, 104, 113) or have focused on the mechanism of action of the CRISPR-Cas system in bacteria (56, 76) or in archaea (32, 99, 108); others have examined the CRISPR-Cas system from an evolutionary perspective (63, 112) or placed it in the context of the burgeoning small RNA world (58).

MÉNAGE A TROIS: BIOINFORMATICS, BIOCHEMISTRY, AND BACTERIOPHAGES

Brief History of the CRISPR-Cas System

A timeline of unrelated observations made over the past twenty years provides a compelling example of the slow but satisfying path from hypotheses generated solely from sequence and genome context predictions (73, 80) to biochemical, structural, and genetic data that substantiated these initial ideas (11, 30). An unintended consequence of this trajectory has been the use of several confusing acronyms and synonyms [see Makarova (72) and Deveau et al. (24) for clarification]. Fifteen years elapsed between the initial report of the presence of DNA repeat arrays in the intergenic region adjacent to the alkaline phosphatase (*iap*) gene in *Escherichia coli* K12 (52) and the coining of the CRISPR acronym in 2002, following the observation that such arrays of repeats were common in bacteria and archaea (53, 54, 81). In the 1990s, bioinformatics-based tools allowed for easy identification of these highly conserved arrays of palindromic repeats as bacterial and archaeal genome sequencing/annotation projects greatly expanded (16, 47, 48, 59, 60, 62, 77, 81, 84, 98). In hindsight, the meager viral and plasmid sequence information available hindered the interpretation of the genetic content and potential function(s) of the CRISPRs and even today remains a limitation. Initial predictions based on bioinformatic analyses were made suggesting involvement in chromosome partitioning (81) and DNA repair (73), and the bold conjecture was put forward by Makarova and colleagues that the CRISPR-Cas system might be a defense system akin to eukaryotic RNAi (74).

Spacer: small, variable sequences (flanked by repeats) in the host genome that are acquired from foreign nucleic acids and play a role in defense, so the name is inaccurate but widely adopted

Bacteriophage: viruses that infect bacteria, also known as phage

The year 2005 marked a turning point when three groups independently reported that the hypervariable spacers showed sequence homology to viruses (or bacteriophages) or plasmids and hypothesized that CRISPRs and associated proteins could play a role in immunity against transmissible genetic elements (13, 80, 90). Research groups working with genetically tractable bacterial systems and available host and/or viral genomes were soon able to provide experimental evidence. A report in 2007 documenting the ability of the CRISPR-Cas system to provide viral resistance (11) and a publication in 2008 showing the ability of CRISPRs to prevent plasmid transfer (75) provided the impetus to investigate the mechanism of action, leading to important new advances. Researchers also quickly realized the potential of these hypervariable and rapidly evolving genetic loci as a valuable tool for genotyping closely related pathogenic or environmentally relevant strains and this is an area of active research.

THE CAST OF CHARACTERS

Novel Features of the CRISPR-Cas Defense System

The CRISPR-Cas defense system has the novel ability to incorporate short sequences of nonself genetic material known as spacers at specific locations within CRISPRs in the

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

CRISPR locus/array: region on genome or plasmid containing CRISPRs

Leader: 5′ end of crRNA preceding the first CRISPR repeat, which can contain long AT-rich tracts and include a promoter

Protospacer: sequence in foreign DNA which is acquired as a spacer into the host genome

Protospacer-associated motif (PAM): short conserved sequence in the immediate vicinity of a protospacer that is required for acquisition

pre-CRISPR RNA (pre-crRNA): full-length transcript produced from CRISPR array

CRISPR RNA (crRNA): small noncoding RNA produced by cleavage of pre-crRNA (also known as psiRNA or guide RNA)

Interference: the process by which incoming foreign DNA or RNA is targeted for destruction; also known as adaptive immunity

CASCADE (CRISPR-associated complex for antiviral defense): multisubunit protein complex required for interference

host genome. Spacers are transcribed and processed into small noncoding RNAs, which in conjunction with specific Cas protein complexes can bind to incoming foreign genetic material if there is a close or absolute sequence match between the small RNA and incoming nucleic acid. This sequence-specific recognition process culminates in destruction of the invading nucleic acid and requires several Cas proteins. The surveillance and attack process exploits previous exposure to a virus or plasmid to target incoming foreign DNA (or RNA). This provides the host heritable immunity to recently detected foreign DNA and hence has been termed an adaptive or acquired immune system. However, some obvious distinctions between the CRISPR-Cas system and the classical vertebrate immune response include the fact that the CRISPR-Cas system can readily acquire new spacers (or conversely, lose old spacers) and this time-resolved activity allows it to respond dynamically to a viral predator that is also evolving at high rates. Furthermore, spacer-derived immunity is inherited by daughter cells, reminiscent of a Lamarckian mode of evolution, which does not occur in eukaryotes.

A functional CRISPR-Cas system has two distinguishable components required for activity (**Figure 1**). The first easily recognizable feature is the CRISPR locus/array located on the genome (either chromosome or plasmid), which contains the hypervariable spacers acquired from virus or plasmid DNA. The second feature is a diverse group of *cas* genes located in the vicinity of a CRISPR locus, which encode proteins (generically called Cas proteins) required for the multistep defense against invasive genetic elements.

The CRISPR-Cas defense process can be separated into either two or three stages (**Figure 1**). The first stage, referred to as adaptation (30, 76), immunization (49), or spacer acquisition (58, 113), involves the recognition and subsequent integration of spacers between two adjacent repeat units within the CRISPR locus. Spacers appear to be integrated primarily at one end (the leader end) of the CRISPR locus; thus, positional information represents a timeline of spacer acquisition events. To indicate the sequence on the viral genome that corresponds to a spacer, the term protospacer was coined (23). In several, but not all, cases, a very short stretch of conserved nucleotides in the immediate vicinity of the protospacer, referred to as the protospacer adjacent motif (PAM), or CRISPR motif, appears to be a recognition motif required for acquisition of the DNA fragment (24, 79) (**Figure 2**). This first phase minimally requires two nucleases, Cas1 and Cas2, both of which are universally present in genomes that have a functional CRISPR/Cas system and can be considered hallmarks of the system.

In the second stage, referred to as CRISPR expression, a primary transcript, or pre-CRISPR RNA (pre-crRNA) is transcribed from the CRISPR locus by RNA polymerase. Next, specific endoribonucleases cleave the pre-crRNAs into small CRISPR RNAs (crRNAs). Based on their function, these small RNAs have also been referred to as prokaryotic silencing (psiRNAs) (40, 74) or guide RNAs (14, 19). In the third and final stage, described as interference (24) or immunity (30), the crRNAs within a multiprotein complex, [called CASCADE (CRISPR-associated complex for antiviral defense) in particular organisms, such as *E. coli*] can recognize and base-pair specifically with regions of incoming foreign DNA (or RNA) that have perfect (or almost perfect) complementarity (14). This initiates cleavage of the crRNA–foreign nucleic acid complex (30). On the other hand, if there are mismatches between the spacer and target DNA or if there are mutations in the PAM, then cleavage is not initiated. In this case, DNA is not targeted for attack, replication of the virus proceeds, and the host is not immune to virus attack (**Figure 2**). This leads to host lysis, and the released virus can attack other susceptible host cells. To operate as a defense system, all three phases (i.e., spacer acquisition, expression, and interference) must be functional, but it is important to note that each of these processes can work independently, both mechanistically and temporally.

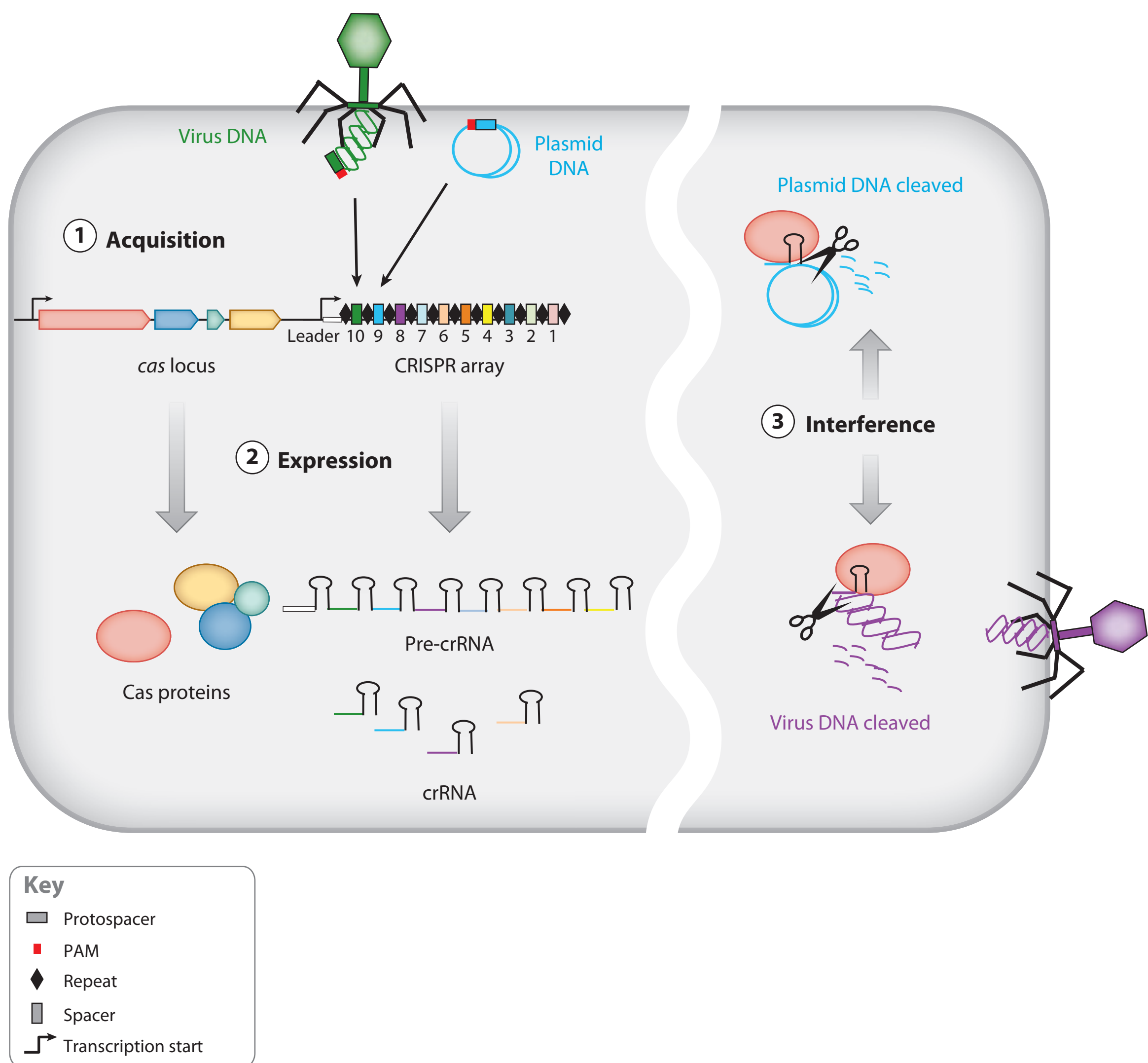

Figure 1

Features of the CRISPR-Cas adaptive immune system. Stage 1: CRISPR spacer acquisition. Specific fragments or protospacers (with an adjacent protospacer-associated motif; *shown as red bar*) of double-stranded DNA from a virus or plasmid are acquired at the leader end of a CRISPR array on host DNA. A CRISPR array consists of unique spacers (*colored boxes; spacers are numbered sequentially with the most recently acquired spacer having the highest number*) interspaced between repeats (*black diamonds*). Acquisition occurs by a process that minimally requires Cas1 and Cas2, encoded in the *cas* locus, usually located in the vicinity of the CRISPR array. Stage 2: CRISPR expression. Pre-CRISPR RNA (Pre-crRNA) is transcribed from the leader region by RNA polymerase and further cleaved into smaller crRNAs that contain a single spacer and a partial repeat (*hairpin structures with colored spacers*) by Cas proteins. Stage 3: CRISPR interference. crRNA containing a spacer that has a strong match to incoming foreign nucleic acid (plasmid or virus) initiates a cleavage event (*shown by scissors*); Cas proteins are required for this process. DNA cleavage interferes with virus replication or plasmid activity and imparts immunity to the host. Interference can be mechanistically and temporally separated from CRISPR acquisition and expression (*depicted by wavy white bar across the cell*). This figure is based on the CRISPR-Cas system in *Streptococcus thermophilus*, which represents a well-studied and relatively simple CRISPR-Cas system.

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

Figure 2

A closeup of the CRISPR-Cas system. Some of the basic steps in CRISPR-Cas defense are depicted here based on current understanding of the CRISPR Type I and II systems; other variations are shown in **Figure 4**. (*a*) Double-stranded viral DNA (*gray*) with a protospacer (*green*) and adjacent protospacer associated motif (PAM, *red*). (*b*) Host DNA with the newly acquired spacer (*blue*) flanked by CRISPR repeats (*black*). Acquisition minimally requires Cas1 and Cas2. (*c*) RNA polymerase transcribes pre-crRNA (transcription start site not shown); orange arrows mark where cleavage of pre-CRISPR RNA (crRNA) by endoribonucleases can occur to create mature crRNAs. This step is accomplished by different Cas proteins, depending on the particular CRISPR type. The putative secondary structure of crRNA is shown below, with a 5′ handle (*black*) followed by the spacer (*blue*) and a potential hairpin structure at the 3′ end (*hairpin in black*). (*d*): A perfect or almost match between crRNA and foreign nucleic acid, initiates a cleavage event within the protospacer. A number of Cas-encoded proteins, e.g., CASCADE and Cas3, are required for this process. (*e*) Mismatches (*purple letters*) between crRNA and foreign nucleic acid (either in the protospacer or PAM) prevent cleavage, and interference does not occur. Note that crRNA basepairs with the complementary strand of target foreign DNA during the interference step (not shown here).

CRISPRdb: CRISPR database; includes several tools to identify and analyze CRISPRs. Maintained by Universite Paris Sud (**http://crispr.u-psud.fr/**)

CRISPI: database that identifies Cas proteins and CRISPRs; includes many features complementary to CRISPRdb (**http://crispi.genouest.org/**)

The CRISPR Locus

Although these loci were christened in several ways, the term CRISPR locus (or CRISPR array) has now been widely adopted and refers to a genomic region containing CRISPRs (24). Dedicated databases (CRISPRdb and CRISPI) (35, 94) that identify CRISPRs and Cas proteins on sequenced genomes indicate that they are present on most archaeal (~90%) and many bacterial (~50%) genomes or on resident plasmids. The number of spacers in a particular CRISPR locus can vary widely, from as few as one to several hundred (as many as 587 spacers at a specific CRISPR locus in the myxobacterium *Haliangium ochraceum* DSM 14365; NC_013440) (35). Among different species, the length of the repeat can vary from 21 bp to 48 bp, whereas spacers are typically between 26 bp and 72 bp (33, 35). The sequence of the repeat units in different CRISPR loci is not conserved, although there are partially conserved sequences such as a GTTTg/c motif at the 5′ end and a GAAAC motif at the 3′ end (24, 33, 54, 65). Because of the partially palindromic nature of the repeats, it was hypothesized that transcripts from these regions may form stable, highly conserved RNA secondary

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org by Universidad Nacional de la Plata on 11/17/11. For personal use only.

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

structures (65, 74), and recent structure/function data supports this (see section entitled Interference).

Although the majority of genomes contain a single CRISPR-Cas locus, there are several examples of bacterial or archaeal species that harbor multiple CRISPR loci on the chromosome. Some bacterial genomes contain as many as 13–15 CRISPR loci, and archaeal genomes also harbor numerous CRISPR loci, based on information available in CRISPRdb (35) and CRISPI (94). Not all CRISPR loci have adjoining *cas* genes; it is possible that only the subset of CRISPR loci that have adjacent *cas* genes are functionally active, whereas the others represent inactive loci, or that one set of Cas proteins suffices for the activity of related CRISPR loci in *trans* (25, 50). In cases where there are multiple CRISPR loci on a genome, the sequence of some of the CRISPR repeats may be identical or very similar, and these may utilize the same set of Cas proteins (44, 50, 65). CRISPR repeats have been classified into at least 12 groups, and there appears to be some correspondence between certain repeats and groups (or subtypes) of Cas proteins associated with them (65, 72).

Cas Proteins

CRISPR loci often have groups of conserved protein-encoding genes, named *cas* genes, in their vicinity (73). Based on computational analyses, Cas proteins were predicted to contain identifiable domains characteristic of helicases, nucleases, polymerases, and RNA-binding proteins, which led to the initial speculation that they may be part of a novel DNA repair system (73). The order, orientation, and groupings of *cas* genes appear to be extremely variable, and this picture grows ever more complex as the number of annotated genomes increases. Attempts to classify Cas proteins have been made, but this has proven difficult because of the diversity of the proteins involved (38, 72, 74). Initially, Jansen's group identified four gene families, *cas1–4* (53), which were then extended to include *cas*5 and *cas6* (13, 38). Haft and colleagues (38) defined eight subtypes of Cas proteins based on the phylogeny of the highly conserved Cas1 protein and the operonic organization of *cas* genes, which were named after eight representative organisms that contained a single CRISPR-Cas locus [e.g., *E. coli* Cas proteins were designated cse1 (CRISPR system of *E. coli* gene1); other subtypes included *Aeropyrum* (csa), *Desulfovibrio* (csd), *Haloarcula* (csh), *Mycobacterium* (csm), *Neisseria* (csn), *Thermotoga* (cst), and *Yersinia* (csy)].

These initial categories, although useful, cannot easily handle the relationships between homologous but distantly related Cas proteins, the extensive variability that exists in *cas* operons, or organisms that contain multiple CRISPR loci. In a new and unified classification system based on multiple criteria, including evolutionary relationships of conserved proteins and *cas* operon organization, several groups (72) working on CRISPR-Cas systems have proposed a consensus view that the CRISPR-Cas system can be divided into two partially independent subsystems. The first consists of an information processing module and requires the universally present core proteins, Cas1 and Cas2, which are involved in new spacer acquisition. The second, or executive, subsystem is required for processing of primary CRISPR transcripts (crRNA) and recognition and degradation of invading foreign nucleic acid, and is quite diverse. For instance, in certain CRISPR sub-types, the multisubunit CASCADE is involved in the processing of the crRNA, whereas in other types a single multifunctional protein may play this role. In addition, there are several repeat-associated mysterious proteins (RAMPs) that constitute a large superfamily of Cas proteins. RAMPs contain at least one RNA recognition motif (RRM; it is also called the ferredoxin-fold domain), and some have been shown to be involved in pre-crRNA processing (14, 19, 26, 42). Based on this classification that integrates phylogeny, sequence, locus organization, and content, three types have been distinguished, Type I, Type II, and Type III CRISPR-Cas systems (**Table 1, Figure 3**, **Figure 4**).

Repeat-associated mysterious protein (RAMP): proteins with RNAse activity involved in crRNA processing

Type I, II, and III CRISPR-Cas system: new classification of CRISPR-Cas systems based on multiple criteria

Table 1 Major Cas proteins (Cas 1–Cas 10). For other Cas proteins, please refer to Makarova (72). Blue, red, and purple designate universal, signature, and type-specific Cas proteins (as in Figure 3).

Protein	Distribution	COG	Process	Function
Cas1	Universal	COG1518	Spacer acquisition	DNAse, not sequence specfic, can bind RNA; present in all Types; structure available for several Cas 1 proteins
Cas2	Universal	COG1343, COG3512	Spacer acquisition	Small RNAse specific to U-rich regions; present in all Types; structure available from *Thermus thermophilus* and *Sulfolobus solfataricus* and others
Cas3	Type I signature	COG1203,COG2254	Target interference	DNA helicase; most proteins have a fusion to HD nuclease
Cas4	Type I, II	COG1468	Spacer acquisition	RecB-like nuclease with exonuclease activity homologous to RecB
Cas5	Type I	COG1688, RAMP	crRNA expression	RAMP protein, endoribonuclease involved in crRNA biogenesis; part of CASCADE
Cas6	Type I, III	COG1583, COG5551, RAMP	crRNA expression	RAMP protein, endoribonuclease involved in crRNA biogenesis; part of CASCADE; structure available from *P. furiosus*
Cas7	Type I	COG1857, COG3649, RAMP	crRNA expression	RAMP protein, endoribonuclease involved in crRNA biogenesis; part of CASCADE
Cas8	Type I	Not determined	crRNA expression	Large protein with McrA/HNH-nuclease domain and RuvC-like nuclease; part of CASCADE
Cas9	Type II signature	COG3513	Target interference	Large multidomain protein with McrA-HNH nuclease domain and RuvC-like nuclease domain; necessary for interference and target cleavage
Cas10	Type III signature	COG1353	crRNA expression and interference	HD nuclease domain, palm domain, Zn ribbon; some homologies with CASCADE elements

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org by Universidad Nacional de la Plata on 11/17/11. For personal use only.

Type I CRISPR-Cas System

In addition to the presence of the conserved Cas1 and Cas2 proteins, Type I is defined by the ubiquitous presence of a signature protein, the Cas3 helicase/nuclease. Cas3 is a large multidomain protein with distinct DNA nuclease and helicase activities (102). In addition, there are multiple Cas proteins that form CASCADE-like complexes that are involved in the interference step (**Figures 4** and **5**). Many of these proteins are in distinct RAMP superfamilies (Cas5, Cas6, Cas7). Of the three systems, Type I, thus far, is the most diverse with six different subtypes (Type I-A through Type II-F) (72). The Type I CRISPR system is believed to target DNA, and cleavage requires Cas3 [which has a histidine, aspartic acid (HD) nuclease domain] or Cas4, a RecB-family nuclease (102). The Type I CRISPR-Cas system in *E. coli* is one of the best characterized (**Figures 4** and **5**) and recent experiments using *E.coli* are described in later sections (14) (57). Multiple studies in *Pseudomonas aeruginosa* have also shed light on the Type I CRISPR-Cas mechanism of action (42, 118, 119). For DNA interference, CASCADE associates with processed crRNA to form a ribonucleoprotein complex that drives the formation of R-loops in invasive double-stranded DNA (dsDNA) (**Figure 5**) through seed sequence–driven base

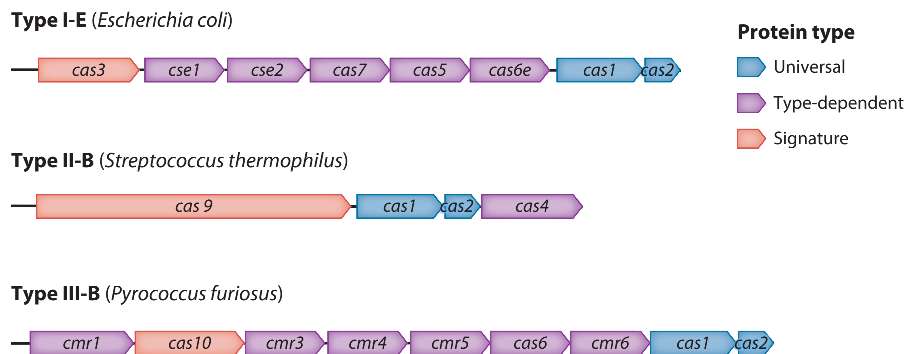

Figure 3

Cas proteins in Type I, II, and III CRISPR-Cas systems. Archetypal Type I, II, and III systems are represented by the operon structure from *Escherichia coli*, *Streptococcus thermophilus*, and *Pyrococcus furiosus*, respectively. The universally present *cas1* and *cas2* genes required for acquisition are shown in blue. Signature genes for each type (Type I, *cas3*; Type II, *cas9;* and Type III, *cas10*) are shown in red. Type-dependent genes (i.e., *cas4*, *5*, *6*, *7*) are in purple, *cas8*, which is not shown here, is found in Type I-A, I-B, and I-C. In Type III-B, *cmr1*, *3*, *4*, *5*, *6* are type-dependent genes (Type III A has a different set of type-dependent genes, several of which are repeat-associated mysterious proteins). Type I has six subtypes (Type IA-IF) and both Type II and III have two subtypes. Type-dependent proteins are typically involved in expression and/or interference; signature genes are involved in interference. However, there are exceptions to these categories; see **Table 1** for Cas protein functions and Makarova et al. (72) for further details.

pairing. It was recently shown in *E. coli* (97) and *P. aeruginosa* (119) that CASCADE facilitates dsDNA target recognition by sequence-specific hybridization between crRNA and the target DNA over a 7–8 bp sequence (the seed sequence) at the 5′ end of the spacer (**Figure 5**).

Type II CRISPR-Cas System

This system is typified by the Cas9 signature protein, a large multifunctional protein with the ability to generate crRNA, as well as target phage and plasmid DNA for degradation (30). Cas9 appears to contain two nuclease domains, one at the N terminus (RuvC-like nuclease) and an HNH (McrA-like) nuclease domain in the middle section (which might be involved in target cleavage based on its endonuclease activity) (**Table 1**). Type II is the simplest of the three CRISPR-Cas types, with only four genes that compose the operon (this includes *cas9*, *cas1*, *cas2*, and either *cas4* or *csn2*).There are two subtypes, Type IIA (or CASS4 that includes *csn2*) and Type IIB (or CASS4a that includes *cas4*). The best-studied Type II system is that of *Streptococcus thermophilus*, which has been shown to provide defense against bacteriophage and plasmid DNA (11, 30). It was also recently established that a *trans*-encoded small CRISPR RNA (tracrRNA) is involved in the processing of pre-crRNA into crRNA in Type II systems through the formation of a duplex with the CRISPR repeat sequence (22). Mature crRNA, together with Cas9, interferes with matching invasive dsDNA by homology-driven cleavage within the protospacer sequence, in the direct vicinity of the PAM (30). Mismatches at the 3′ end of the protospacer and/or in the PAM allow phages and plasmids to circumvent CRISPR-encoded immunity (24, 30).

tracrRNA: Trans-encoded small RNA required for crRNA maturation

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org by Universidad Nacional de la Plata on 11/17/11. For personal use only.

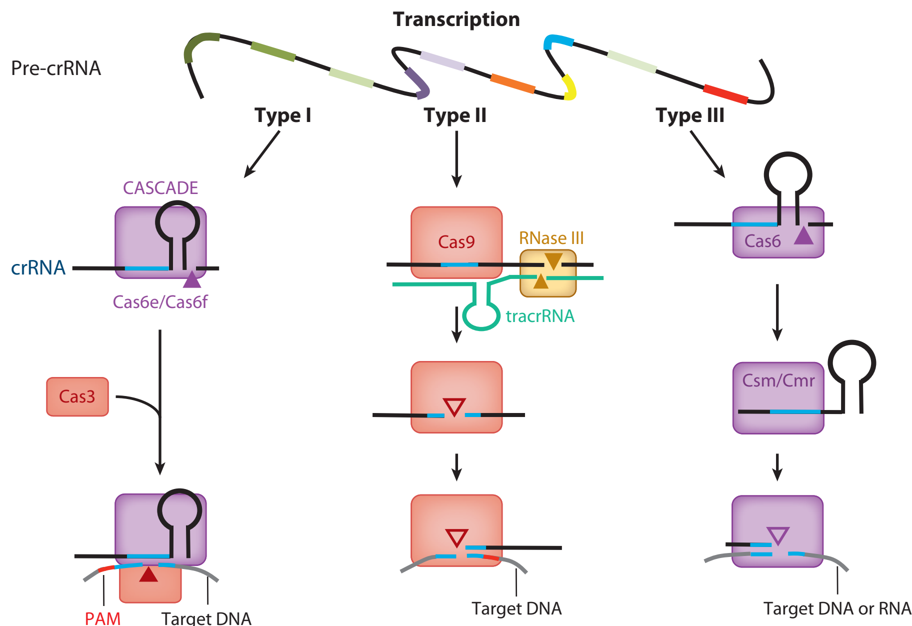

Figure 4

Model of Type I, II, and III CRISPR-Cas mechanism of action. Transcription of the primary transcript (pre-crRNA) by RNA polymerase is followed by the production of mature crRNAs. The interference process is different in the Type I, II, and III systems. In Type I, the multisubunit CASCADE binds pre-crRNA, which is cleaved by Cas6e in subtype I-E or by Cas6f in subtype I-F, to create crRNAs with a typical 8-nt extension or handle at the 5′ end, followed by the spacer (*blue*) and part of the repeat region, which can form a hairpin structure at the 3′ end. In Type II, a *trans*-encoded small RNA (tracrRNA) base-pairs with the repeat region and is cleaved by host RNase III. Additionally, Cas9 is required for this step and for subsequent maturation of the crRNA. In Type III, processing of crRNA requires Cas6, but the crRNAs appear to be transferred to a specific Cas complex (Cmr in subtype III-B and Csm in subtype III-A). In subtype III-B, the 3′ end of crRNA is further trimmed. The final step results in cleavage of targeted foreign nucleic acid and proceeds differently in all systems. In Type I, crRNA with CASCADE along with the Cas3 subunit can recognize (via the PAM, *shown in red*) complementary target DNA and is responsible for cleavage of target DNA. In Type II, Cas9 along with crRNA can probably target DNA for cleavage (*open orange triangle*) in a process that requires the PAM. Subtype IIIA can target DNA, whereas subtype III-B can target RNA and a PAM does not appear to be required for the activity of Type III systems. This figure is modified from Makarova et al. (72). Filled triangles represent nuclease activity that has been experimentally demonstrated; open triangles represent activity that has not yet been identified.

Type III CRISPR-Cas System

This system has a number of recognizable features, including the signature RAMP protein, Cas10, which is likely involved in the processing of crRNA and possibly also in target DNA cleavage (6), and is somewhat functionally analogous to the Type I CASCADE. The Type III system also contains the signature Cas6, involved in crRNA processing and additional RAMP proteins likely to be involved in crRNA trimming. The universal *cas1* and *cas2* genes are mostly in operon-like structures with the rest of the *cas* genes but are not always in the same operon as the RAMP proteins in the Type III systems. So far, two type III systems have been distinguished (Type IIIA and IIIB). In *Pyrococcus furiosus*, a Type IIIA system, the target of CRISPR interference is mRNA (40), whereas in *Staphylococcus epidermidis*, a Type IIIB system, the target is DNA (75). This highlights

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

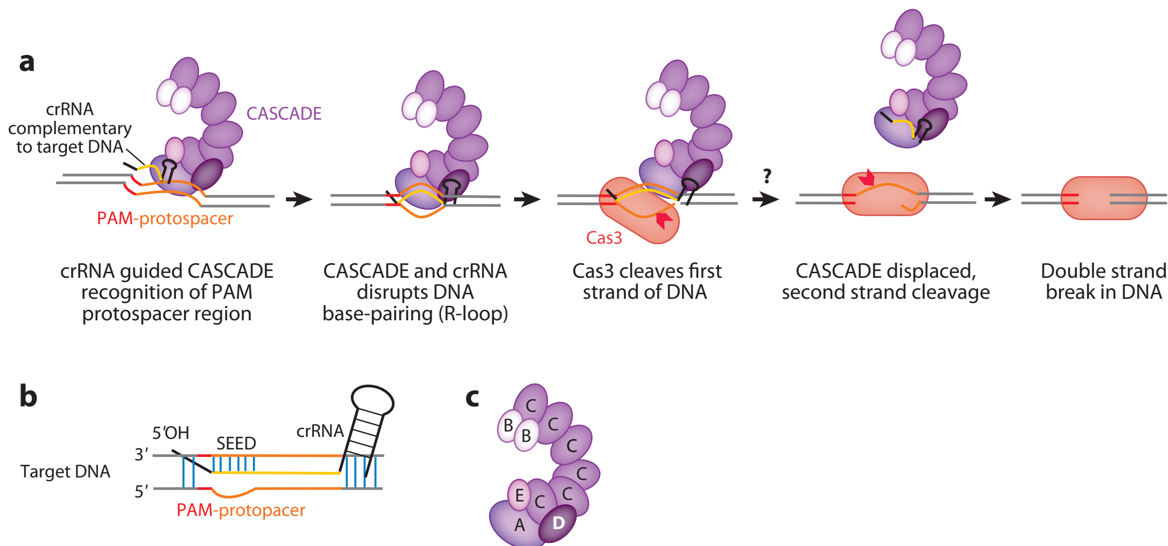

Figure 5

Model of CASCADE function and involvement of seed sequences. (*a*) Initially, the CASCADE-crRNA complex identifies a protospacer-PAM region on target DNA. This promotes strand separation, and crRNA hybridizes to the complementary DNA strand leading to R-loop formation. The single-DNA strand in the R-loop structure is the target of Cas3 nuclease activity resulting in single-strand breaks in the protospacer. Following cleavage of the first DNA strand, Cas3 helicase activity can possibly displace the CASCADE-crRNA complex and permit second DNA strand cleavage. This would result in both strands of target DNA being cleaved, and interference with virus replication or plasmid function. This model is adapted from Sinkunas et al. (102), but also incorporates information from Wiedenheft et al. (119), Jore et al. (57), Semenova et al. (97). (*b*) Model of base pairing between crRNA spacer and target DNA that results in R-loop formation. The process is initiated at the seed sequence adjacent to the PAM and then propagated along the protospacer region in a 5′→3′ direction over the complete protospacer region (97, 119). (*c*) Structural model of CASCADE based on Jore et al. (57) showing the stoichiometry and the unusual seahorse architecture of the subunits of CASCADE which includes CasA, B, C, D, and E.

the polymorphic nature of CRISPR-Cas systems, even within the Type III systems.

The distribution of the three CRISPR-Cas systems has some notable features, with the Type I system being found in both bacteria and archaea. In contrast, Type II is exclusively present in bacteria, whereas the Type III systems appear more commonly in archaea, although it is also found in bacteria (72, 108). Although no large scale detailed distributional or functional analysis is yet available, there are several examples of species that contain more than one CRISPR-Cas type. Horizontal gene transfer (via plasmids that harbor CRISPR-Cas loci or by other gene transfer mechanisms such as transposon activity) has been implicated in the movement of CRISPR-Cas loci across widely diverged lineages (33, 50, 88).

A PLAY IN THREE ACTS: MECHANISM OF DEFENSE VIA CRISPR SPACER ACQUISITION, EXPRESSION, AND INTERFERENCE

Current understanding of the mechanism of action, which is divided into three phases, CRISPR spacer acquisition, CRISPR locus expression (including transcription and processing), and CRISPR activity or interference are described below.

CRISPR Spacer Acquisition

Currently, the best-studied model organisms to investigate the mechanistic aspects of CRISPR acquisition include *E. coli* and *P. aeruginosa* (Type I), *S. thermophilus* (Type II)

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

and *P. furiosus* (Type III). The ability to acquire novel spacers has been experimentally shown in vivo in *S. thermophilus*, CRISPR1 and CRISPR3, which represent Type II systems, and in *Streptococcus mutans* (114). In *Streptococcus* sp. during the natural generation of phage-resistant strains, one or more spacers were incorporated into their CRISPR loci (11, 114). These loci evolve via addition of new spacers at the leader end, following exposure to lytic phages or plasmid transformation and may be derived from both sense and antisense DNA strands (11, 23, 30). Concurrent internal addition and deletion of spacers appear to be rare events, whereas iterative additions of spacers increase both the level and spectrum of phage resistance in the host (11, 23). Internal deletions of repeat-spacer units likely occur via homologous recombination between CRISPR direct repeats that are physically close on the chromosome as shown in the archaeon *Sulfolobus* sp. (37). When *Sulfolobus* sp. are challenged with various plasmid or viral genes, surviving mutants carried either partial or whole deletions of CRISPR loci (37). Intuitively, it seems disadvantageous and unlikely that CRISPR arrays grow ad infinitum, and the balance of polarized additions and internal deletions has been documented (23), although the mechanisms by which this occurs have not been identified. The greater likelihood of internal deletions occurring toward the trailer or farther away from the leader end of a CRISPR spacer array would preferentially delete spacers that target historically older phages. Conversely, this allows preferential conservation of spacers that provide immunity against contemporary phages. Internal deletions, which typically consist of the removal of several consecutive repeat-spacer units, have been observed in metagenomic studies (111) and when the spacer contents of multiple closely related strains spanning 11 genera were compared (50).

Functionally, the process of spacer acquisition can be divided into distinct steps involving (*a*) recognition of the invasive nucleic acid and scanning foreign DNA for potential PAMs that identify protospacers, (*b*) the generation of a new repeat spacer by processing of the nucleic acid, and (*c*) the integration of the new CRISPR repeat spacer unit at the leader end of the CRISPR locus. Of these processes, only the first step has been characterized, and the mechanism by which a new spacer is integrated into the host genome is very poorly understood. PAMs, also called CRISPR motifs (23), have been recognized in the direct vicinity of some protospacers (79). These conserved regions are short (typically only 2 to 5 nt long) and occur within 1 to 4 bp of the protospacer sequence, on either side, depending on the system (**Figure 2**). For the CRISPR1 and CRISPR3 systems of *S. thermophilus*, AGAAW and GGNG have been identified as PAMs at the 3′ end of the protospacer, respectively (23, 50). Equivalent motifs were identified adjacent to protospacers of *S. mutans*, either downstream (NGG, NAA) or immediately upstream (TTC) of the protospacer (114). A similar TTC sequence immediately upstream of the protospacer was identified in *Xanthomonas* (96). In the archaeon *Sulfolobus* sp., a CC motif has been identified (69). Further bioinformatics analyses of CRISPR spacer–repeat units from a larger dataset of bacterial and archaeal species and the corresponding protospacers should provide a better understanding of the repertoire of PAMs; whether specific PAMs are correlated with CRISPR repeats and/or particular Cas proteins or whether some CRISPR-Cas systems do not require PAMs, such as some Type III systems. Currently, such an analysis is hampered by the limited range of host-virus systems under investigation and the lack of an extensive database of viral genomes.

The Cas1 protein has repeatedly been linked with involvement in the acquisition and/or integration of novel spacers in the CRISPR locus during the acquisition process (11, 14). This correlation is also consistent with the coevolutionary pattern and inherent functional linkage observed between Cas1 and specific CRISPR repeat sequences. The crystal structure of Cas1 is available from several organisms (**Table 2**), including *P. aeruginosa* where studies have demonstrated that Cas1 is

Table 2 Available structures of Cas proteins

Protein	Function	Organism	PDB code (Gene)
Cas1	Metal-dependent DNA-specific endonuclease	*Pseudomonas aeruginosa*	3god (*cas1*)
Cas1	Metal-dependent DNA-specific endonuclease	*Thermotoga maritima*	3lfx (T1797)
Cas1	Metal-dependent DNA-specific endonuclease	*Aquiflex aeolicus*	2yzs
Cas1	Metal-dependent DNA-specific endonuclease	*Pyrococcus horikoshii*	3pv9 (PH1245)
Cas1	Metal-dependent DNA-specific endonuclease	*Escherichia coli*	3nke (*ygbT*)
Cas2	Metal-dependent ssRNA-specific endoribonuclease; ferredoxin fold	*Sulfolobus solfataricus*	2i8e;2ivy;3exc (Sso1404)
Cas2	Metal-dependent ssRNA-specific endoribonuclease; ferredoxin fold	*Desulfovibrio vulgaris*	3oq2 (DvuCas2)
Cas2	Metal-dependent ssRNA-specific endoribonuclease; ferredoxin fold	*Pyrococcus furiosus*	2i0x (PF1117)
Cas2	Metal-dependent ssRNA-specific endoribonuclease; ferredoxin fold	*Thermus thermophilus*	1zpw (Tth1823)
Cas3	Endonuclease (HD domain)	*Thermus thermophilus*	3sk9 (TTHB187)
Cas3	Endonuclease (HD domain)	*Methanocaldococcus jannaschii*	3S4L
Cse2 (CasB)	Alpha helical protein, in CASCADE complex	*Thermus thermophilus*	2zca (TTHB189)
Cas6	Endoribonuclease that generates guide RNA	*Pyrococcus furiosus*	3i4h, 3pkm (*cas6*)
Cas6e (CasE)	RNA-binding protein with ferredoxin fold, RAMP protein	*Thermus thermophilus*	1wj9 (TTHB192)
Cas6f (Csy4)	Endoribonuclease processing crRNA	*Pseudomonas aeruginosa*	2xli, 2xlj, 2xlk (*csy4*)
Cmr5	Cmr complex, Type IIIA	*Thermus thermophilus*	2zop (ttCmr5)
Csm6/Csa3	HTH-type transcriptional regulator	*Sulfolobus solfataricus*	2wte (Sso1445)
Csa3	HTH-type transcriptional regulator	*Sulfolobus solfataricus*	3QYF(Sso1393)

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

a homodimeric, metal-dependent DNAse that can process double-stranded DNA (dsDNA) to a size of ~80 bp (42, 118). Cas1 also has the ability to interact with proteins involved in DNA recombination and repair, and can resolve Holliday junctions, implicating its dual involvement in CRISPR function and in DNA repair (10). This is consistent with the proposed role of Cas1 in addition or removal of CRISPR repeat–spacer units and reconciles the initial prediction that Cas proteins were involved in DNA repair (73). The structure of Cas2 (which is often genetically associated with Cas1) from *Sulfolobus solfataricus* (12) and other organisms has also been resolved and was reported to have a RNA recognition domain and exhibit endoribonucleic activity (**Table 2**). It has been implicated in new repeat-spacer acquisition and integration, although this requires further experimental validation. In addition to Cas1, it was shown that in *S. thermophilus*, Csn2 is necessary for the acquisition of novel spacers following exposure to phages (11) or plasmids (30).

CRISPR Locus Expression

We have differentiated expression into two steps: CRISPR locus transcription/regulation and crRNA processing, both of which are required for interference to occur.

CRISPR locus transcription and regulation. Transcription of a CRISPR locus into a primary transcript, or pre-crRNA, has been examined in a few organisms. These include the workhorse gram-negative bacterium *E. coli* (14, 89, 92), the plant pathogen *Xanthomonas oryzae* (89), the thermophilic bacterium *T. thermophilus* (3), and two archaeal species, *P. furiosus* (39) and the crenarchaeon *Sulfolobus* (69).

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

In *E. coli*, it appears that CRISPR loci are transcribed at constitutively low levels (89), but it is not yet known if this is generally true in most organisms. In *T. thermophilus*, CRISPR expression levels can be induced in the presence of a phage toward the peak of infection near the beginning of host cell lysis. This is corroborated by proteomic characterizations in which peak amounts of Cas proteins are concurrent with phage proteins synthesized following phage challenge (70). Microarray analysis of the 12 CRISPR loci in *T. thermophilus* demonstrated a complex pattern of induction that was partly dependent on the small molecule cAMP in conjunction with the catabolite regulator protein (3, 101). In the few bacterial species that have been examined so far, unidirectional transcription occurs from the 5′ leader end and promoters lie upstream (3, 92). In contrast, detailed transcript analyses of CRISPR loci from the archaeon *Sulfolobus* have shown that transcription initiates just upstream from the first repeat, but that bidirectional transcription of the CRISPR locus also occurs (68, 69). Generally, cells appear to exhibit constitutive levels of Cas proteins and pre-crRNA, but under certain conditions it is possible to regulate these levels, suggesting that there is the option of a background defensive monitoring of invasive nucleic acid occurrence, as well as the flexibility to mount a more concerted counter-attack when necessary (see section titled The Expanding Repertoire).

crRNA processing. Once pre-crRNA has been transcribed, it is processed by endonucleolytic cleavage into smaller units that typically contain a single spacer flanked by partial CRISPR repeats. In the archaea *P. furiosus* (39) and *Sulfolobus* sp. (69), it has been demonstrated that the processed crRNA units, after undergoing complete processing, consist of a single spacer flanked by partial repeats on both sides. The cleavage of pre-crRNA occurs at the base of the hairpin formed by the palindromic CRISPR repeats, typically yielding a crRNA with an 8-nt tag or handle at the 5′ end and a less well-defined boundary at the 3′ end (14, 19, 42). In *Pyrococcus furiosus*, Cas6 is involved in the processing of pre-crRNA into crRNA units (19). In *E. coli*, a multimeric complex CASCADE consisting of CasABCDE processes pre-crRNA (14), whereas in *P. aeruginosa* the protein Csy4 is responsible for cleavage (42). Several studies are currently investigating the molecular basis for crRNA biogenesis across all three CRISPR/Cas system types, as well as the patterns that drive constitutive expression, regulated transcription and overall abundance of crRNA in bacteria and archaea. Preliminary results suggest that CRISPR loci are constitutively expressed, can be induced by viral challenge, and often constitute quantitatively dominant amounts of small RNAs in the cell (22).

CRISPR Interference

The processed crRNA, together with specific Cas proteins, form a CRISPR ribonucleoprotein (crRNP) complex that facilitates spacer base pairing to the target or matching invasive nucleic acid. The crRNA serves as a guide (hence the term guide RNA has also been used) to allow for specific base pairing between the exposed crRNA within the ribonucleoprotein interference complex and the corresponding protospacer on the foreign DNA (14, 86). It is likely that crRNA interacts directly with complementary sequences in the target. The unique occurrence of the PAM sequence on the invading foreign DNA (and conversely, its absence in the host spacer sequence) is likely to play a dual role: first, in spacer selection and acquisition and second, in the interference process for discrimination of self versus nonself, which highlights its importance. Indeed, it has been demonstrated that despite perfect matches between spacer and protospacer sequences, mutations in the PAM can circumvent CRISPR-encoded immunity (23, 30, 95). In archaea, PAMs are likely to be involved in crRNA-mediated targeting (37), which is consistent with crRNA-directed RNA cleavage reported in *P. furiosus* (19). Expanding the spectrum of identified PAMs will be an important step forward that requires greater sequence information about host spacers and corresponding

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

viral genomes. Given the current paucity of complete viral genomes and the fact that they evolve quickly, this is a challenge for future studies.

The CRISPR-Cas system is versatile and has the ability to interfere with foreign dsDNA and single-stranded (ss) mRNA. This is reflected by the bewildering diversity of Cas proteins and their enzymatic activity in various species (72) (**Table 1**). Most early evidence suggested that dsDNA was the primary target of CRISPR-encoded defense or immunity in bacteria, and in *S. thermophilus*, specific CRISPR spacers were found to match coding or template strands of dsDNA phages (11, 13, 23). Marraffini & Sontheimer (75) substantiated this observation when they demonstrated that inserting a self-splicing intron into a protospacer had an impact on CRISPR-encoded immunity in *S. epidermidis*. The CRISPR system prevented uptake of the native plasmid, whereas the intron-containing variant could be conjugated into the host (75). This was later confirmed in the *S. thermophilus* Type II system, with compelling biochemical evidence showing that dsDNA from phages and plasmids were directly cleaved by Csn1 (Cas9) in the vicinity of the PAM (30). On the other hand, it was also established that ssRNA was the primary target of the CRISPR-Cas system in the Type III system of the archaeon *P. furiosus* (40) and that mRNA was also the likely target in the archaeon *Sulfolobus* (32). Both these Type III CRISPR-Cas systems contain genes encoding RAMP proteins, suggesting that these systems can act on RNA. However, a large number of archaeal species contain more than one CRISPR-Cas system, which would expand the repertoire of targets that can be recognized by an organism and add more levels of regulatory control.

The strong link between the activity of various Cas proteins and the sequence of CRISPR repeats appears to be so specific that there is little apparent crosstalk between the different CRISPR-Cas systems; inactivation of a specific *cas* encoding gene cannot be rescued by the activity of other Cas proteins. This is also consistent with observed coevolutionary patterns between the sequences of CRISPR repeats and *cas* genes (50), and appears to hold for both the acquisition of spacers and for the interference process (11, 30). Thus, it seems likely that the diversity in Cas proteins is likely responsible for differential DNA versus RNA targeting and that some RAMPs might be specifically responsible for target RNA interference.

Initial experiments indicated that perfect sequence identity was required between spacer and protospacer sequences for CRISPR-encoded immunity to occur because the presence of a single nucleotide polymorphism in the protospacer or in the PAM sequence abrogated the defense response of the host (11, 23) (**Figure 2**). However, followup experiments in other systems have shown that in certain cases even several mismatches between spacer and protospacer still allowed for the immune response to occur (30, 96). Thus, in some cases, degeneracy in recognition can be tolerated, whereas in others, single nucleotide mutations in the virus genome can thwart CRISPR-encoded defense. The key appears to lie in the location of potential mismatches relative to the cleavage site. Mutations that are distant from the cleavage site do not impact activity, whereas mismatches occurring in the PAM or in the direct vicinity of the cleavage site have a strong impact (18, 30, 97). One can theorize that the short PAM sequences may be easily eroded by mutations; however, the rate at which this happens in the environment has not been tested. Preliminary results in *S. thermophilus* seem to indicate that mutations in phage genomes that circumvent CRISPR-encoded immunity may be costly given that the majority of mutations are either nonsynonymous or deleterious (23). Likewise, some circumstantial evidence points to independent acquisition of effective spacers targeting critical and highly conserved regions with stringent and sequence-dependent functionalities. Although PAM-dependent sampling of protospacers seems implicated in spacer selection, it remains to be determined whether there is strand specificity or other features that guide spacer selection and acquisition rates in different organisms.

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

THE EXPANDING REPERTOIRE: CRISPR-MEDIATED ACTIVITY IN REGULATION AND DEVELOPMENT

To date, the majority of research has focused on the novel ability of the CRISPR-Cas system to provide immunity against invasive genetic elements (virus or plasmid) via small noncoding RNAs that are derived from CRISPR loci. Yet, with the discovery of RNA interference (RNAi) in eukaryotes and the ever-expanding repertoire of roles of small noncoding RNAs in eukaryotes, it has become obvious that small RNAs can be multifunctional and versatile (29, 55, 61). Likewise, there is growing evidence of the diverse regulatory roles that small noncoding RNAs play in bacteria (34, 116). The flexibility and diversity of the CRISPR-Cas system as well as its proven ability to target RNA suggest that the system has the ability to regulate or silence transcript levels within the cell, although to date, such evidence is sparse. The presence of a transcribed CRISPR spacer that matches histidyl-tRNA synthetase was shown to contribute to attenuated histidyl-tRNA pools in the cell, with consequences for the synthesis of histidine-rich proteins (4), but it is unclear if this is a widely used regulatory mechanism in bacteria. A bioinformatics-based approach searching for evidence of the presence of self DNA in CRISPR arrays located a few examples, albeit at a very low frequency, which could represent errors in acquisition rather than a widely used regulatory system (106).

Recent reports indicate that the Cas1 protein (YgbT) of *E. coli* exhibits nuclease activity against both single-stranded and branched DNAs, replication forks, and 5′ flaps, and interacts with RecB, RecC, and RuvB of the DNA repair system, suggesting it may have a dual role and functions in the CRISPR-Cas defense system but can also function in DNA repair (10). It was also shown in *E. coli* that the CRISPR-Cas system was triggered under specific conditions in which misfolded proteins accumulated in the membrane (86), so it was speculated that the CRISPR-Cas system could provide a defense against defective protein accumulation. It is unclear if this is a special case or whether this could hint at the potential use of the CRISPR-Cas system for a more widespread surveillance system in bacteria undergoing stress conditions (either via virus attack or because of other environmental stresses). Regulation of the multiple CRISPR-Cas systems in *E. coli* has also been investigated, and a recent study demonstrated that in *E. coli* K12 transcription from the *casA* and CRISPR I promoters is repressed by heat-stable nucleoid-structuring protein (H-NS) (92), which is a global repressor of transcription in many gram-negative bacteria. In a follow-up study, Westra et al. (117) provided experimental evidence that when LeuO (a LysR-type transcription factor) binds to two sites flanking the *casA* promoter and the H-NS nucleation site, it results in derepression. Thus, in *E. coli* H-NS and LeuO appear to be antagonistic regulators of CRISPR-based immunity (117), although the implications of this regulatory loop in the context of phage infections in natural environments awaits further study (25). In *P. aeruginosa*, the CRISPR loci appear to be involved in lysogeny-dependent inhibition of biofilm formation (18, 119), and CRISPR activity has also been implicated in swarming of myxobacteria (115), indicating that CRISPR-Cas systems may be involved in functions other than defense. We anticipate that there will be other examples of how this versatile system has been coopted for other cellular functions as other organisms are examined.

ALL THE WORLD'S A STAGE: BIOTECHNOLOGICAL APPLICATIONS AND EVOLUTIONARY IMPLICATIONS OF THE CRISPR-CAS SYSTEM

Strain Typing and Epidemiological Studies

Incoming spacers are rapidly acquired at the leader end of the CRISPR array, so they provide

a time-resolved window of spacer acquisition. Therefore, these hypervariable and rapidly evolving genetic loci provide a valuable tool for genotyping of strains, also known as spoligotyping. This feature has been used for genotyping and epidemiological studies of pathogenic *Mycobacterium tuberculosis* (1, 15, 36, 121), *Yersinia pestis* (20, 90), *Corynebacterium diphtheriae* (82, 83), *P. aeruginosa* (18), *Legionella* (21) *Streptococcus pyogenes* (48, 78), and *Salmonella* sp. (71), and for industrially relevant organisms such as lactobacilli and streptococci (50, 51). CRISPR loci provide the ability to segregate nearly identical strains over time or within clonal populations (8, 11, 51). The degree of spacer polymorphism in terms of both number of unique spacers and spacer arrangements usually correlates with the activity level of a given locus; thus, in cases when there are multiple CRISPR loci in the genome, it is important to choose a genetically polymorphic locus.

The temporal and spatial hypervariability of CRISPR spacers can also be exploited to resolve population level genotypes in complex samples where strain diversity is difficult to determine, as was shown in *Leptospirillum* population analyses in acid mine drainage acidophilic biofilm samples (8, 111). Similar observations were made in natural samples containing mixed and dynamic populations of *Sulfolobus* sp. (45, 46) and *Synechococcus* sp., (44) and also in human subjects, where streptococci population, exposed to phage predation, showed significant changes (91). Such approaches could monitor complex, dynamic systems over time to identify ancestral relationships or infer events such as blooms, selective sweeps, and bottlenecks. It is likely that active and hypervariable CRISPR loci will be increasingly used in complex metagenomic studies to genetically characterize microbial population content and dynamics (7, 105). As the understanding of the role viruses play in shaping bacterial and archaeal communities grows, we foresee that CRISPRs will not only assist in resolving population dynamics of the microbes, but also shed light on the coevolutionary dynamics between host and virus (7, 31, 44).

Microbial Populations, Population Dynamics, and the Modeling of the CRISPR-Associated Arms Race

CRISPR loci provide information about the exposure to foreign genetic elements and insight into the relationships between bacteria or archaea and their biotic environments. A key feature of the CRISPR defense system is that CRISPR loci can rapidly acquire novel spacers. Phages also have the ability to mutate their genomes, allowing them to circumvent the host CRISPR-encoded immunity system, which relies on close matches between spacers and incoming nucleic acid. Phages may escape CRISPR spacers by either mutating or deleting bases in the protospacer and/or the PAM (23, 30), or by shuffling sequences targeted by CRISPR spacers (8). These rapid evolutionary dynamics can provide important insights into genome evolution of both the host and phage populations (but see References 109, 110 for a different perspective). They set the stage for mathematical modeling of their evolutionary interplay (43) and short- or long-term experimental analyses of phage-host coevolution. These experiments can be carried out in closed, controlled laboratory conditions or in open environmental systems (67, 112).

Spoligotyping: a modified form of genotyping based on strain-dependent hybridization patterns of in vitro amplified DNA with multiple spacer oligonucleotides

Natural Defense and Immunity Against Phages and Plasmids

The ability of the CRISPR-Cas system to target plasmids that contain antimicrobial resistance markers and to target sequences from antibiotic resistance genes has been documented (30). Consequently, the CRISPR-Cas system provides a natural means to develop strains refractory to the uptake of plasmids that carry undesirable genes. There is potential to develop CRISPRs in strains so as to preclude uptake and prevent dissemination of undesirable genetic elements such as prophages, antibiotic resistance markers, and pathogenicity islands (27, 85, 100). Indeed, given the negative correlation between the occurrence of CRISPR and acquired antibiotic resistance in multidrug-resistant enterococci, the CRISPR-Cas system

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org by Universidad Nacional de la Plata on 11/17/11. For personal use only.

appears able to mitigate the spread of mobile elements, notably plasmids and prophages, which account for up to 25% of pathogenic *Enterococcus faecalis* (85). This functionality could be exploited to reduce the dissemination of antimicrobial resistance genes and virulence factors in widely distributed bacterial pathogens, such as methicillin-resistant *Staphylococcus aureus* (76).

The negative effect of CRISPR-encoded immunity on plasmid occurrence and dissemination is apparent in the relative scarcity of plasmids in dairy *S. thermophilus*, as compared with *Lactococcus lactis*. Both species have comparable genome size, and identical environments where phages are a recurring problem, but the former relies primarily on chromosomally encoded CRISPR loci, whereas the latter is mostly dependent on CRISPR-independent strategies that are widely encoded on plasmids, such as restriction modification systems and abortive infection systems (2, 24, 66). As the mechanistic understanding of CRISPR-Cas defense rapidly grows, it is important to place it in the context of other important host defense systems, such as abortive infection, toxin-antitoxin, and restriction-modification (66). Furthermore, the little-appreciated role of phage ecology (2) and the fledging field of viral ecogenomics will need to be integrated into our understanding of the evolution of the CRISPR-Cas systems (9, 28, 87, 93).

Prospects for Developing CRISPR-Enabled Technologies

The potential to harness the natural ability of the CRISPR-Cas immune system to develop increased phage resistance in vitro provides an experimental framework to iteratively build up phage resistance for perennial use of valuable cultures and domesticated industrial microbes. Further, the unique spacer combination obtained through several consecutive rounds of CRISPR mutant screening can be seen as a natural genetic tag for valuable proprietary strains. Although there are several examples in the literature that highlight the functional features of the CRISPR-Cas defense system in a variety of organisms, it is important to note that their propensity for mutations, deletions, and loss of *cas* genes has resulted in loss of function in a large array of CRISPR loci (50). There is evidence that the CRISPR-Cas system can be moved by horizontal transfer and conversely that they can also be rapidly lost (or reorganized) from an organism (33, 44, 88). There are examples of CRISPR loci located on plasmids, and transposons and insertion sequences are known to flank CRISPR loci (33, 44, 50), so these loci also represent compelling examples of various driving forces in the evolution of genomes, including horizontal gene transfer.

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

SUMMARY POINTS

1. Bacteria and archaea have evolved several mechanisms to deal with abundant, evolving virus populations and invasive genetic elements such as plasmids. The recently discovered CRISPR-Cas system utilizes exposure to foreign nucleic acids to subsequently target and destroy incoming related viruses or plasmids. This provides the host with heritable resistance, and hence it has been termed an adaptive immune response system.
2. This novel and widely occurring defense system can incorporate specific small fragments of foreign nucleic acids into the host genome between conserved short DNA repeats known as CRISPRs. Later, these fragments or spacers are transcribed and processed into small RNAs, which, in conjunction with Cas proteins, are used to recognize and destroy nonself nucleic acid.

3. CRISPR-Cas systems have been identified in most archaeal (~90%) and many bacterial (~50%) genomes and on resident plasmids. Most genomes contain a single CRISPR-Cas locus, but can also harbor multiple CRISPR loci.
4. The CRISPR-Cas system has recently been categorized into three types (Type I, II, and III) based on phylogeny, sequence, locus organization, and content of the CRISPRs and associated *cas* genes (which encode various DNases, RNases, and other proteins).
5. The CRISPR-Cas defense process can be operationally distinguished into three phases: spacer acquisition, CRISPR expression, and CRISPR interference, which are temporally separated. Each step requires one or more Cas proteins, and deciphering the role and structure of specific Cas proteins is an area of active research.
6. In addition to defense against incoming viruses or plasmids, CRISPRs may play a role in host regulatory and developmental processes.
7. The hypervariable CRISPR spacers have been effectively used in pathogen and environmental genotyping and in the future may have broader applications.

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

FUTURE ISSUES

1. Many mechanistic aspects of the defense system remain unclear. Of these, the details of the acquisition process and whether PAMs are universally required for recognition still need significant experimental input. What controls the size of CRISPR loci? Are specific proteins involved in orchestrating loss of CRISPR spacers, or is it a stochastic process? A clearer understanding of how both RNA and DNA can be targeted by the Cas proteins is work for the future.
2. A few well-developed model systems have been successfully used for the characterization of the basic mechanisms of CRISPR-Cas mediated defense, but to appreciate the full scope of CRISPR-Cas mediated function will require exploring a wide range of genetically tractable model systems.
3. Several open questions pertain to the distribution of the CRISPR loci in varied environmental niches. For example, many genomes contain multiple functional CRISPR-Cas loci. What are the evolutionary driving forces for the concurrent maintenance of multiple functional systems? Do they provide additive immunity, and are CRISPR-Cas loci more common in environments with certain characteristics, e.g., high population sizes or high temperatures? Conversely, why do certain bacterial species lack this versatile and robust defense system? Some of these questions will benefit from targeted metagenomic studies as well as theoretical modeling approaches.
4. The features of CRISPR-Cas systems readily lend themselves to a broad range of applications, notably leveraging hypervariability for strain typing and epidemiological studies, resolving the content and dynamics of complex microbial communities, building natural defenses against viruses and plasmids, engineering immunity against undesirable genetic elements, and expanding the shelf life of industrial workhorses and other organisms being developed for biotechnological applications.

DISCLOSURE STATEMENT

R.B. is a coinventor on several patent applications related to the use of CRISPR for various applications.

ACKNOWLEDGMENTS

We thank Philippe Horvath for a critical reading of the manuscript and several members of the CRISPR community for discussing and sharing their data. D.B. and M.D. were supported by the National Science Foundation and the Carnegie Institution for Science, and R.B. by DANISCO.

LITERATURE CITED

1. Abadia E, Zhang J, dos Vultos T, Ritacco V, Kremer K, et al. 2010. Resolving lineage assignation on *Mycobacterium tuberculosis* clinical isolates classified by spoligotyping with a new high-throughput 3R SNPs based method. *Infect. Genet. Evol.* 10:1066–74
2. Abedon ST. 2009. Phage evolution and ecology. *Adv. Appl. Microbiol.* 67:1–45
3. Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, Shinkai A. 2010. Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J. Mol. Biol.* 395:270–81
4. Aklujkar M, Lovley DR. 2010. Interference with histidyl-tRNA synthetase by a CRISPR spacer sequence as a factor in the evolution of *Pelobacter carbinolicus*. *BMC Evol. Biol.* 10:230
5. Al-Attar S, Westra ER, van der Oost J, Brouns SJ. 2011. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.* 392:277–89
6. Anantharaman V, Iyer LM, Aravind L. 2010. Presence of a classical RRM-fold palm domain in Thg1-type 3′–5′ nucleic acid polymerases and the origin of the GGDEF and CRISPR polymerase domains. *Biol. Direct* 5:43
7. Anderson RE, Brazelton WJ, Baross JA. 2011. Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* 77:120–33
8. Andersson AF, Banfield JF. 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320:1047–50
9. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, et al. 2006. The marine viromes of four oceanic regions. *PLoS Biol.* 4:e368
10. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, et al. 2011. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol. Microbiol.* 79:484–502
11. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–12
12. Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, et al. 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* 283:20361–71
13. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151:2551–61
14. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, et al. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321:960–64
15. Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, et al. 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* 6:23
16. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, et al. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273:1058–73
17. Cady KC, White AS, Hammond JH, Abendroth MD, Karthikeyan RS, et al. 2011. Prevalence, conservation and functional analysis of *Yersinia* and *Escherichia* CRISPR regions in clinical *Pseudomonas aeruginosa* isolates. *Microbiology* 157:430–37

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

18. Cady KC, O'Toole GA 2011 Non-identity targeting of *Yersinia*-subtype CRISPR-prophage interaction requires the Csy and Cas3 proteins. *J. Bacteriol*. 93:3433–45
19. Carte J, Wang R, Li H, Terns RM, Terns MP. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 22:3489–96
20. Cui Y, Li Y, Gorge O, Platonov ME, Yan Y, et al. 2008. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS ONE* 3:e2652
21. D'auria G, Jimenez-Hernandez N, Peris-Bondia F, Moya A, Latorre A. 2010. *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC Genomics* 11:181
22. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, et al. 2011. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* 471:602–7
23. Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, et al. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190:1390–400
24. Deveau H, Garneau JE, Moineau S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* 64:475–93
25. Diez-Villasenor C, Almendros C, Garcia-Martinez J, Mojica FJ. 2010. Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* 156:1351–61
26. Ebihara A, Yao M, Masui R, Tanaka I, Yokoyama S, Kuramitsu S. 2006. Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Sci.* 15:1494–99
27. Edgar R, Qimron U. 2010. The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *J. Bacteriol.* 192:6291–94
28. Edwards RA, Rohwer F. 2005. Viral metagenomics. *Nat. Rev. Microbiol.* 3:504–10
29. Farazi TA, Juranek SA, Tuschl T. 2008. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 135:1201–14
30. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, et al. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468:67–71
31. Garrett RA, Prangishvili D, Shah SA, Reuter M, Stetter KO, Peng X. 2010. Metagenomic analyses of novel viruses and plasmids from a cultured environmental sample of hyperthermophilic neutrophiles. *Environ. Microbiol.* 12:2918–30
32. Garrett RA, Shah SA, Vestergaard G, Deng L, Gudbergsdottir S, et al. 2011. CRISPR-based immune systems of the *Sulfolobales*: complexity and diversity. *Biochem. Soc. Trans.* 39:51–57
33. Godde JS, Bickerton A. 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* 62:718–29
34. Gottesman S, Storz G. 2010. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb. Perspect. Biol.* doi: 10.1101/cshperspect.a003798
35. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35:W52–57
36. Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD. 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*: application for strain differentiation by a novel typing method. *Mol. Microbiol.* 10:1057–65
37. Gudbergsdottir S, Deng L, Chen Z, Jensen JV, Jensen LR, et al. 2011. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* 79:35–49
38. Haft DH, Selengut J, Mongodin EF, Nelson KE. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* 1:e60
39. Hale C, Kleppe K, Terns RM, Terns MP. 2008. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14:2572–79
40. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, et al. 2009. RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex. *Cell* 139:945–56
41. Hambly E, Suttle CA. 2005. The viriosphere, diversity, and genetic exchange within phage communities. *Curr. Opin. Microbiol.* 8:444–50
42. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329:1355–58

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org by Universidad Nacional de la Plata on 11/17/11. For personal use only.

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

43. He J, Deem MW. 2010. Heterogeneous diversity of spacers within CRISPR (clustered regularly interspaced short palindromic repeats). *Phys. Rev. Lett.* 105:128102
44. Heidelberg JF, Nelson WC, Schoenfeld T, Bhaya D. 2009. Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS ONE* 4:e4169
45. Held NL, Herrera A, Cadillo-Quiroz H, Whitaker RJ. 2010. CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLoS ONE* 5:e12988
46. Held NL, Whitaker RJ. 2009. Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environ. Microbiol.* 11:457–66
47. Hermans PW, van Soolingen D, Bik EM, de Haas PE, Dale JW, van Embden JD. 1991. Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect. Immun.* 59:2695–705
48. Hoe N, Nakashima K, Grigsby D, Pan X, Dou SJ, et al. 1999. Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. *Emerg. Infect. Dis.* 5:254–63
49. Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327:167–70
50. Horvath P, Coute-Monvoisin AC, Romero DA, Boyaval P, Fremaux C, Barrangou R. 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food Microbiol.* 131:62–70
51. Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, et al. 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190:1401–12
52. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 169:5429–33
53. Jansen R, Embden JD, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43:1565–75
54. Jansen R, van Embden JD, Gaastra W, Schouls LM. 2002. Identification of a novel family of sequence repeats among prokaryotes. *OMICS* 6:23–33
55. Jinek M, Doudna JA. 2009. A three-dimensional view of the molecular machinery of RNA interference. *Nature* 457:405–12
56. Jore MM, Brouns SJ, van der Oost J. 2011. RNA in defense: CRISPRs protect prokaryotes against mobile genetic elements. *Cold Spring Harb. Perspect. Biol.* doi: 10.1101/cshperspect.a003657
57. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, et al. 2011. Structural basis for CRISPR RNA-guided DNA recognition by CASCADE. *Nat. Struct. Mol. Biol.* 18:529–36
58. Karginov FV, Hannon GJ. 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* 37:7–19
59. Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, et al. 1999. Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* 6:83–101, 145–52
60. Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, et al. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3 (supplement). *DNA Res.* 5:147–55
61. Ketting RF. 2011. The many faces of RNAi. *Dev. Cell* 20:148–61
62. Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, et al. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390:364–70
63. Koonin EV, Makarova KS. 2009. CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 Biol. Rep.* 1:95
64. Koonin EV, Wolf YI. 2009. Is evolution Darwinian or/and Lamarckian? *Biol. Direct.* 4:42
65. Kunin V, Sorek R, Hugenholtz P. 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8:R61
66. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8:317–27
67. Levin BR. 2010. Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria. *PLoS Genet.* 6:e1001171
68. Lillestol RK, Redder P, Garrett RA, Brugger K. 2006. A putative viral defence mechanism in archaeal cells. *Archaea* 2:59–72

69. Lillestol RK, Shah SA, Brugger K, Redder P, Phan H, et al. 2009. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol. Microbiol.* 72:259–72
70. Lintner NG, Frankel KA, Tsutakawa SE, Alsbury DL, Copie V, et al. 2011. The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. *J. Mol. Biol.* 405:939–55
71. Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, et al. 2011. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl. Environ. Microbiol.* 77:1946–56
72. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, et al. 2011. Evolution and classification of the CRISPR/Cas systems. *Nat. Rev. Microbiol.* 9:467–77
73. Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV. 2002. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* 30:482–96
74. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* 1:7
75. Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322:1843–45
76. Marraffini LA, Sontheimer EJ. 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11:181–90
77. Masepohl B, Gorlitz K, Bohme H. 1996. Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium Anabaena sp. PCC 7120. *Biochim. Biophys. Acta* 1307:26–30
78. McShan WM, Ferretti JJ, Karasawa T, Suvorov AN, Lin S, et al. 2008. Genome sequence of a nephritogenic and highly transformable M49 strain of *Streptococcus pyogenes*. *J. Bacteriol.* 190:7773–85
79. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155:733–40
80. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60:174–82
81. Mojica FJ, Ferrer C, Juez G, Rodriguez-Valera F. 1995. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* 17:85–93
82. Mokrousov I, Limeschenko E, Vyazovaya A, Narvskaya O. 2007. Corynebacterium diphtheriae spoligotyping based on combined use of two CRISPR loci. *Biotechnol. J.* 2:901–6
83. Mokrousov I, Vyazovaya A, Kolodkina V, Limeschenko E, Titov L, Narvskaya O. 2009. Novel macroarray-based method of *Corynebacterium diphtheriae* genotyping: evaluation in a field study in Belarus. *Eur. J. Clin. Microbiol. Infect. Dis.* 28:701–3
84. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, et al. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399:323–29
85. Palmer KL, Gilmore MS. 2010. Multidrug-resistant enterococci lack CRISPR-cas. *mBio* 1:e00227–10
86. Perez-Rodriguez R, Haitjema C, Huang Q, Nam KH, Bernardis S, et al. 2011. Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. *Mol. Microbiol.* 79:584–99
87. Pignatelli M, Aparicio G, Blanquer I, Hernandez V, Moya A, Tamames J. 2008. Metagenomics reveals our incomplete knowledge of global diversity. *Bioinformatics* 24:2124–25
88. Portillo MC, Gonzalez JM. 2009. CRISPR elements in the Thermococcales: evidence for associated horizontal gene transfer in *Pyrococcus furiosus*. *J. Appl. Genet.* 50:421–30
89. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, et al. 2010. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* 77:1367–79
90. Pourcel C, Salvignol G, Vergnaud G. 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151:653–63
91. Pride DT, Sun CL, Salzman J, Rao N, Loomer P, et al. 2011. Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome Res.* 21:126–36

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

92. Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R. 2010. Identification and characterization of *E. coli* CRISPR-Cas promoters and their silencing by H-NS. *Mol. Microbiol.* 75:1495–512
93. Rohwer F. 2003. Global phage diversity. *Cell* 113:141
94. Rousseau C, Gonnet M, Le Romancer M, Nicolas J. 2009. CRISPI: a CRISPR interactive database. *Bioinformatics* 25:3317–18
95. Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides plasmid immunity in *Escherichia coli*. *Nucl. Acids Res.* doi: 10.1093/nar/gkr606
96. Semenova E, Nagornykh M, Pyatnitskiy M, Artamonova II, Severinov K. 2009. Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. *FEMS Microbiol. Lett.* 296:110–16
97. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, et al. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. USA* 108:10098–103
98. Sensen CW, Charlebois RL, Chow C, Clausen IG, Curtis B, et al. 1998. Completing the sequence of the *Sulfolobus solfataricus* P2 genome. *Extremophiles* 2:305–12
99. Shah SA, Garrett RA. 2011. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. *Res. Microbiol.* 162:27–38
100. Shimomura Y, Okumura K, Yamagata Murayama S, Yagi J, Ubukata K, et al. 2011. Complete genome sequencing and analysis of a Lancefield group G *Streptococcus dysgalactiae* subsp. *equisimilis* strain causing streptococcal toxic shock syndrome (STSS). *BMC Genomics* 12:17
101. Shinkai A, Kira S, Nakagawa N, Kashihara A, Kuramitsu S, Yokoyama S. 2007. Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8. *J. Bacteriol.* 189:3891–901
102. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* 30:1335–42
103. Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 12:246–58
104. Sorek R, Kunin V, Hugenholtz P. 2008. CRISPR: a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6:181–86
105. Sorokin VA, Gelfand MS, Artamonova II. 2010. Evolutionary dynamics of clustered irregularly interspaced short palindromic repeat systems in the ocean metagenome. *Appl. Environ. Microbiol.* 76:2136–44
106. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* 26:335–40
107. Suttle CA. 2005. Viruses in the sea. *Nature* 437:356–61
108. Terns MP, Terns RM. 2011. CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* 14:321–27
109. Touchon M, Charpentier S, Clermont O, Rocha EP, Denamur E, Branger C. 2011. CRISPR distribution within the *Escherichia coli* species is not suggestive of immune-associated diversifying selection. *J. Bacteriol.* 193:2460–67
110. Touchon M, Rocha EP. 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. *PLoS ONE* 5:e11126
111. Tyson GW, Banfield JF. 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* 10:200–7
112. Vale PF, Little TJ. 2010. CRISPR-mediated phage resistance and the ghost of coevolution past. *Proc. Biol. Sci.* 277:2097–103
113. van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ. 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* 34:401–7
114. van der Ploeg JR. 2009. Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* 155:1966–76
115. Viswanathan P, Murphy K, Julien B, Garza AG, Kroos L. 2007. Regulation of dev, an operon that includes genes essential for *Myxococcus xanthus* development and CRISPR-associated genes and repeats. *J. Bacteriol.* 189:3738–50
116. Waters LS, Storz G. 2009. Regulatory RNAs in bacteria. *Cell* 136:615–28

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

117. Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, et al. 2010. H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* 77:1380–93
118. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* 17:904–12
119. Wiedenheft B, van Duijn E, Bultema J, Waghmare S, Zhou K, et al. 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA.* 108:10092–97
120. Zegans ME, Wagner JC, Cady KC, Murphy DM, Hammond JH, O'Toole GA. 2009. Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. *J. Bacteriol.* 191:210–19
121. Zhang J, Abadia E, Refregier G, Tafaj S, Boschiroli ML, et al. 2010. *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of "spoligotyping" with new spacers and a microbead-based hybridization assay. *J. Med. Microbiol.* 59:285–94

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

Annual Review of Genetics

Volume 45, 2011

Contents

Comparative Genetics and Genomics of Nematodes: Genome Structure, Development, and Lifestyle
Ralf J. Sommer and Adrian Streit 1

Uncovering the Mystery of Gliding Motility in the Myxobacteria
Beiyan Nan and David R. Zusman 21

Genetics and Control of Tomato Fruit Ripening and Quality Attributes
Harry J. Klee and James J. Giovannoni 41

Toxin-Antitoxin Systems in Bacteria and Archaea
Yoshihiro Yamaguchi, Jung-Ho Park, and Masayori Inouye 61

Genetic and Epigenetic Networks in Intellectual Disabilities
Hans van Bokhoven 81

Axis Formation in Hydra
Hans Bode 105

The Rules of Engagement in the Legume-Rhizobial Symbiosis
Giles E.D. Oldroyd, Jeremy D. Murray, Philip S. Poole, and J. Allan Downie 119

A Genetic Approach to the Transcriptional Regulation of *Hox* Gene Clusters
Patrick Tschopp and Denis Duboule 145

V(D)J Recombination: Mechanisms of Initiation
David G. Schatz and Patrick C. Swanson 167

Human Copy Number Variation and Complex Genetic Disease
Santhosh Girirajan, Catarina D. Campbell, and Evan E. Eichler 203

DNA Elimination in Ciliates: Transposon Domestication and Genome Surveillance
Douglas L. Chalker and Meng-Chao Yao 227

Double-Strand Break End Resection and Repair Pathway Choice
Lorraine S. Symington and Jean Gautier 247

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.

CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation
Devaki Bhaya, Michelle Davison, and Rodolphe Barrangou 273

Human Mitochondrial tRNAs: Biogenesis, Function, Structural Aspects, and Diseases
Tsutomu Suzuki, Asuteka Nagao, and Takeo Suzuki 299

The Genetics of Hybrid Incompatibilities
Shamoni Maheshwari and Daniel A. Barbash 331

Maternal and Zygotic Control of Zebrafish Dorsoventral Axial Patterning
Yvette G. Langdon and Mary C. Mullins 357

Genomic Imprinting: A Mammalian Epigenetic Discovery Model
Denise P. Barlow 379

Sex in Fungi
Min Ni, Marianna Feretzaki, Sheng Sun, Xuying Wang, and Joseph Heitman 405

Genomic Analysis at the Single-Cell Level
Tomer Kalisky, Paul Blainey, and Stephen R. Quake 431

Uniting Germline and Stem Cells: The Function of Piwi Proteins and the piRNA Pathway in Diverse Organisms
Celina Juliano, Jianquan Wang, and Haifan Lin 447

Errata

An online log of corrections to *Annual Review of Genetics* articles may be found at http://genet.annualreviews.org/errata.shtml

Annu. Rev. Genet. 2011.45:273-297. Downloaded from www.annualreviews.org
by Universidad Nacional de la Plata on 11/17/11. For personal use only.