Impact of Small Repeat Sequences on Bacterial Genome Evolution

Nicholas Delihas*

Department of Molecular Genetics and Microbiology, School of Medicine, State University of New York, Stony Brook

*Corresponding author: E-mail: nicholas.delihas@stonybrook.edu.

Accepted: 19 July 2011

Abstract

Intergenic regions of prokaryotic genomes carry multiple copies of terminal inverted repeat (TIR) sequences, the nonautonomous miniature inverted-repeat transposable element (MITE). In addition, there are the repetitive extragenic palindromic (REP) sequences that fold into a small stem loop rich in G–C bonding. And the clustered regularly interspaced short palindromic repeats (CRISPRs) display similar small stem loops but are an integral part of a complex genetic element. Other classes of repeats such as the REP2 element do not have TIRs but show other signatures. With the current availability of a large number of whole-genome sequences, many new repeat elements have been discovered. These sequences display diverse properties. Some show an intimate linkage to integrons, and at least one encodes a small RNA. Many repeats are found fused with chromosomal open reading frames, and some are located within protein coding sequences. Small repeat units appear to work hand in hand with the transcriptional and/or post-transcriptional apparatus of the cell. Functionally, they are multifaceted, and this can range from the control of gene expression, the facilitation of host/pathogen interactions, or stimulation of the mammalian immune system. The CRISPR complex displays dramatic functions such as an acquired immune system that defends against invading viruses and plasmids. Evolutionarily, mobile repeat elements may have influenced a cycle of active versus inactive genes in ancestral organisms, and some repeats are concentrated in regions of the chromosome where there is significant genomic plasticity. Changes in the abundance of genomic repeats during the evolution of an organism may have resulted in a benefit to the cell or posed a disadvantage, and some present day species may reflect a purification process. The diverse structure, eclectic functions, and evolutionary aspects of repeat elements are described.

Key words: DNA repeat sequences, MITE, REP, CRISPR, nonautonomous transposable elements, genome evolution.

Introduction

Small DNA repeat sequences, less than approximately 400 bp, are present in genomes in a wide range of bacteria. These repeats are primarily in intergenic regions of the chromosome and are present in multiple copies, some as many as approximately 1,600 (Rocco et al. 2010). Many repeat units fall into two broad categories, the miniature inverted-repeat transposable element (MITE) (Siguier et al. 2006; Delihas 2008) and the repetitive extragenic palindromic (REP) sequence (Stern et al. 1984; Bachellier et al. 1999). Other repeats such as the REP 2-5 units (Parkhill et al. 2000), YPLA/RU2 (De Gregorio et al. 2006; Delihas 2007), and *bcr* elements (Kristoffersen et al. 2011) appear to constitute separate classes or are subclasses. The clustered regularly interspaced short palindromic repeats (CRISPRs) are in a category of their own in that they are found as an array with short spacer sequences and are associated with a complex family of protein genes. Most repeat sequences have the potential to fold into a stable secondary structure at the DNA and/or RNA level, and many are transcribed into RNA where the RNA secondary structure may be a factor in regulating gene expression (Croucher et al. 2011). Examples of predicted RNA secondary structures of repeat units are in figure 1. Repeats display diverse roles in terms of bacterial cell physiology and cell–host interactions. They are found pintegron units (Gillings et al. 2009; Poirel et al. 2009). REPs are implicated in stimulation of the mammalian immune system (Magnusson et al. 2007), and they can affect genomic plasticity by serving as sites for insertion of transposable elements (Tobes and Pareja 2006). CRISPR units function as an RNA-based mechanism of inhibition of invading DNA and represent a possible example of Lamarckian inheritance in prokaryotes (Koonin and Wolf 2009).

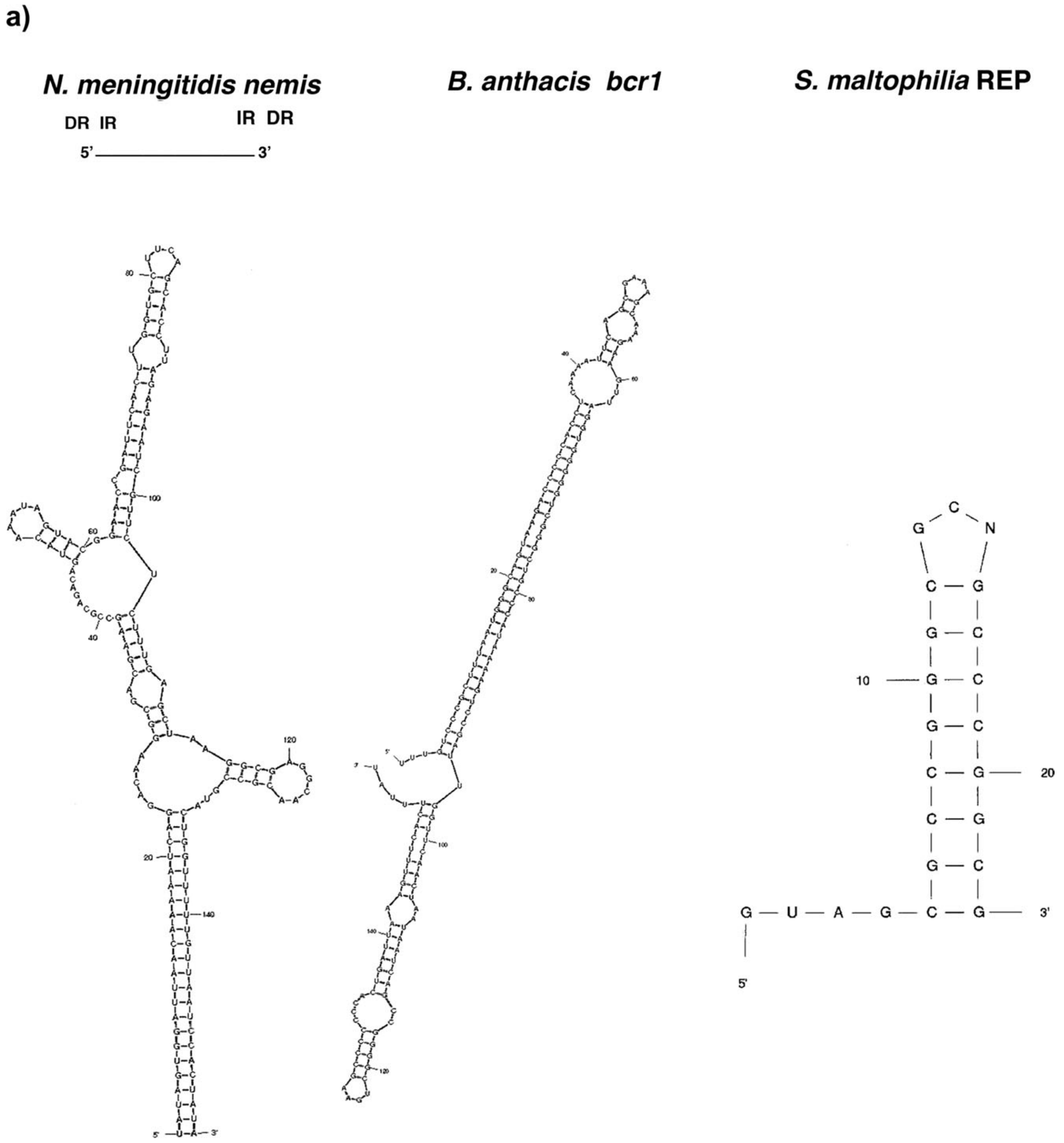

Fig. 1.—(*a*) Predicted secondary structures of repeat sequences at the RNA level. The Mfold program was used for RNA folding (Markham and Zuker 2005). The *Neisseria meningitidis* nemis (neisseria miniature ISs) is characteristic of MITEs, and the secondary structure shown is similar to that of Mazzone et al. (2001). The top schematic describes inverted repeats (IR) and DRs flanking the DNA strand. The *bcr1* structure is that of Bacillus anthracis 1R (Økstad et al. 2004) and is typical of the *Bacillus bcr1* RNA secondary structures (Klevan et al. 2007). These consist of a cruciform-like structure with two independent stem loops. The *Stenotrophomonas maltophilia* REP sequence and secondary structure shown is characteristic of the short high G–C content REPs found in these species; they are termed SMAG (Rocco et al. 2010). These SMAG units can carry an unpaired tetranucleotide sequence at one end. (*b*) Left, predicted RNA secondary structures of the REP2 sequence from *N. meningitidis* showing internal stem loops 1 and 2. The nt sequence is from Morelle et al. (2003). Upper schematic denotes the REP2 DNA strand with promoter, ribosome binding site (RBS), and ATG initiation codon. (*b*) Right, predicted secondary structural model of the *Borrelia burgdoferi* IR-A sequence from circular plasmid cp8.3/lp21 [nt sequence from Dunn et al. (1994)]. Stem loops 1 and 2 may be analogous to those of REP2; however, Dunn et al. (1994) show the two IR-A stem loops in DNA form. Top schematic depicts the DNA strand with promoter, RBS, and ATG sites on the IR-A segment.

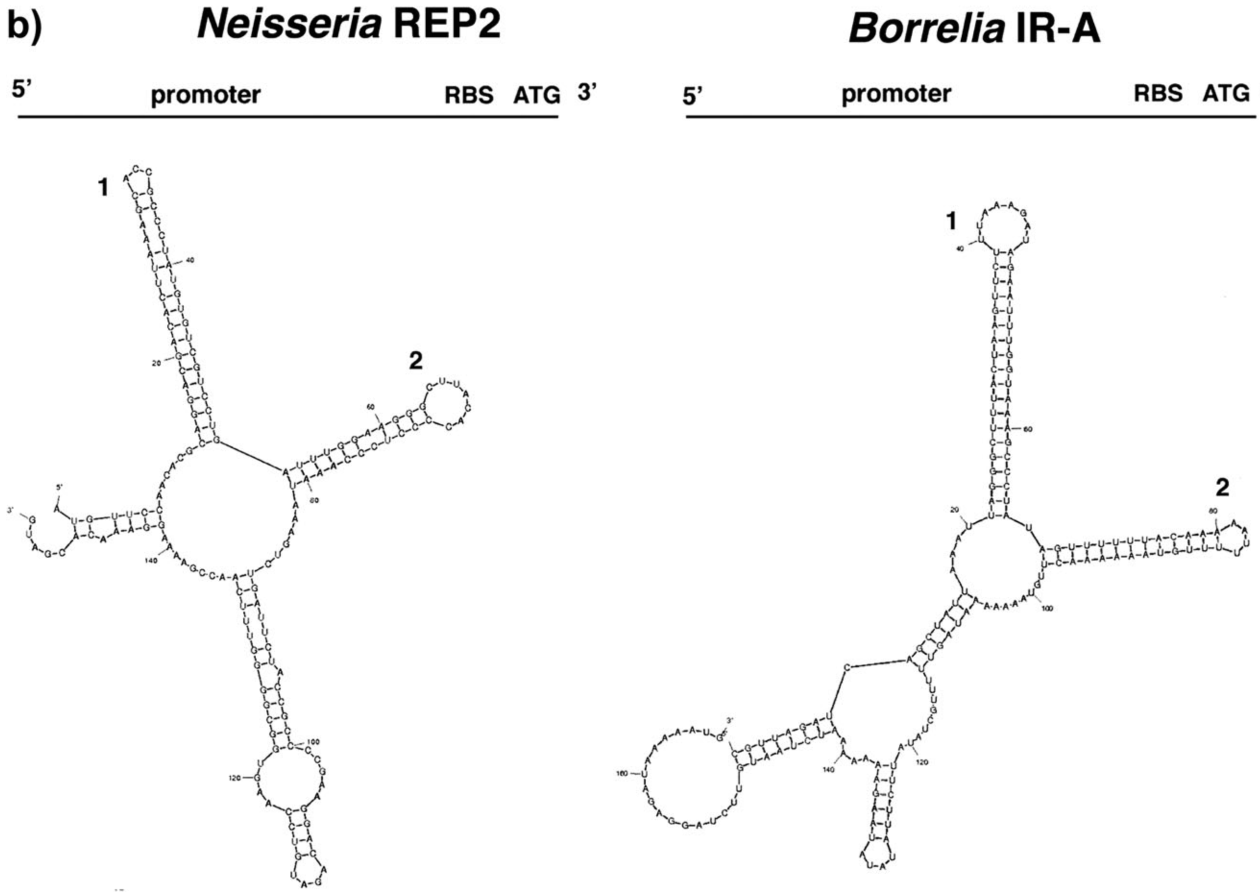

Fig. 1.— Continued

REP sequences were first discovered in *Escherichia coli* (Higgins et al. 1982; Gilson et al. 1984; Stern et al. 1984), the MITEs in *Neisseria* (Correia et al. 1986, 1988), and CRISPR palindromic repeats in *E. coli* (Ishino et al. 1987). Properties of several repeat elements have been reviewed in the past (Tobes and Pareja 2006; Brouns et al. 2008; Delihas 2008); however, with the advent of an array of whole-genome sequences and development of bioinformatics programs to identify these units (Chen et al. 2009), increased numbers of repeat elements are being discovered (table 1) and comparative genomics between closely related bacterial species can be done. Such analysis has yielded important aspects of evolutionary change occurring in genomes that may be related to repeat sequences, for example, the correlation between repeat element location and chromosomal plasticity (Mine et al. 2009; Silby et al. 2009; Ogier et al. 2010; Kristoffersen et al. 2011). In other studies, a comparison of changes in repeats during evolution has led to the concept that a high abundance of mobile repeats in genomes can be parasitic and a potential disadvantage to an organism; some current species are found to carry fewer mobile repeats than their ancestors (Croucher et al. 2011). On the other hand, phylogenetic comparisons of *Pelobacter carbinolicus* and its ancestors, together with the results from genetic experiments using transgenic strains of *Geobacter*, led Aklujkar and Lovley (2010) to propose that a CRISPR spacer sequence that contained a segment of the host gene *hisS* resulted in an evolutionary loss of ancestral genes that rely on the function of *hisS.*

Many repeat sequences also display open reading frames that are found fused to chromosomal reading frames. These fusions are discussed in terms of a possible formation of new proteins or the alteration of existing proteins. Jacob (1977) proposed the concept of "tinkering" during evolution in terms of the combination of two motifs to produce a different and more elaborate structure. We review here the diverse molecular, functional, and evolutionary aspects of recently discovered repeat elements.

MITE—A Repeat Element Found in a Broad Range of Bacteria

MITEs are termed nonautonomous as they are incapable of self-transfer and require a transposase acting in trans for transposition. Although MITEs were first discovered in bacteria (Correia et al. 1988), they were formalized as

Table 1
Bacterial Intergenic Repeat Elements

Repeat Element	Organisms	Class	Size (bp)	Reference
Correia	*Neisseria* sp.	MITE	∼104 to 157	Correia et al. (1988)
RUP	*Streptococcus pneumoniae*	MITE	107	Oggioni and Claverys (1999)
ERIC	Enterobacteriaceae	MITE	∼127	Sharples and Lloyd (1990) and Hulton et al. (1991)
MaeMITE	*Microcystis aeruginosa*	MITE	150–435	Kaneko et al. (2007)
Nezha	*Anabaena variabilis*, *Nostoc* sp.	MITE	∼130 to 170	Zhou et al. (2008)
MITE	*Anabaena* sp.	MITE	∼224	Wolk et al. (2010)
Chunjie	*Geobacter uraniireducens* Rf4	MITE	178 to 235	Chen et al. (2008)
Muzha	*A. variabilis*	MITE	∼154	Chen et al. (2009)
Duanwu	*Haloquadratum walsbyi*	MITE	257	Chen et al. (2009)
Qixi	*H. walsbyi*	MITE	165	Chen et al. (2009)
Chongyang	*H. walsbyi*	MITE	119	Chen et al. (2009)
MITE	*Anabaena* sp.	MITE	127–204	Fewer et al. (2011)
BOX	*Str. pneumoniae*	MITE-like	67–637	Martin et al. (1992)
R0[a]	*Pseudomonas fluorescens*	MITE-like	89	Silby et al. (2009)
R1	*Pse. fluorescens*	MITE-like	80	Silby et al. (2009)
R2	*Pse. fluorescens*	MITE-like	110	Silby et al. (2009)
R6	*Pse. fluorescens*	MITE-like	177	Silby et al. (2009)
IMU	*Enterobacter cloacae* CHE-2	IMU (MITE)	288	Poirel et al. (2009)
NFM2 MITE	*Acinetobacter* sp.	NFM2 (MITE)	439	Gillings et al. (2009)
SPRITE	*Str. pneumoniae*	Rho-independent terminator-like	∼105	Croucher et al. (2011)
CIR	*Caulobacter* + other sp.	CIR	∼110	Chen and Shapiro (2003)
RPE	*Rickettsia* sp.	RPE	∼105 to 146	Ogata et al. (2000)
YAPL/RU-2	*Yersinia* sp.	YAPL/RU-2	∼168	De Gregorio et al. (2006)and Delihas (2007)
RU-3	*Escherichia coli*, *Shigella* sp.	RU-3	103	Delihas (2007)
bcr1[b]	*Bacillus cereus* group	*bcr* Group A	∼155	Økstad et al. (2004)
bcr5[c]	*B. cereus* group	*bcr* Group B	310	Kristoffersen et al. (2011)
REP	Enterobacteriaceae	REP	∼35	Stern et al. (1984) and Gilson et al. (1984)
REP	*Pse. putida*[d]	REP	35	Aranda-Olmedo et al. (2002)
IR1_g	*Pse. fluorescens*	REP	∼25	Silby et al. (2009)
REP	*Stenotrophomonas* sp.	REP	∼35	Nunvar et al. (2010) and Rocco et al. (2010)
ATR	*Pse. fluorescens*	ATR	183	Silby et al. (2009)
R178	*Pse. fluorescens*	R178	101	Silby et al. (2009)
REP2	*Neisseria meningitidis*	REP2	∼134 to 154	Parkhill et al. (2000)
REP3	*N. meningitidis*	REP3	60	Parkhill et al. (2000)
REP4	*N. meningitidis*	REP4	26	Parkhill et al. (2000)
REP5	*N. meningitidis*	REP5	20	Parkhill et al. (2000)
RS (NIME)	*N. meningitidis*	RS (NIME)	70–200	Parkhill et al. (2000)
CRISPR	*E. coli* + other sp.	CRISPR	28–49	Ishino et al. (1987)
Borrelia IR	*Borrelia burgdorferi*	IR-A, IR-B	∼180	Dunn et al. (1994)
BRE	Beta-proteobacteria	BRE	∼90	Hot et al. (2011)
Stem loop left	*Borrelia* sp.	Stem loop left	34	Delihas (2009)
Stem loop right	*Borrelia* sp.	Stem loop right	32–51	Delihas (2009)

[a] Seven additional repeat elements without IR not shown.

[b] Two additional similar repeats not shown.

[c] Two additional similar repeats not shown.

[d] See Tobes and Pareja (2006) for additional species with REP sequences.

nonautonomous transposable sequences in plants (Bureau and Wessler 1992, 1994; Feschotte et al. 2002; Kikuchi et al. 2003). Experimentally, they have been transferred by transposases in vivo in both prokaryotes and eukaryotes (Poirel et al. 2009; Yang et al. 2009; Hancock et al. 2010).

Bacterial MITEs either are or once were mobile. They are generally less than 200 bp, but some are as larger as approximately 400 bp. MITE sequences have signatures typical of many insertion sequences (ISs), that is, they contain terminal inverted repeats (TIRs) that straddle a core sequence and

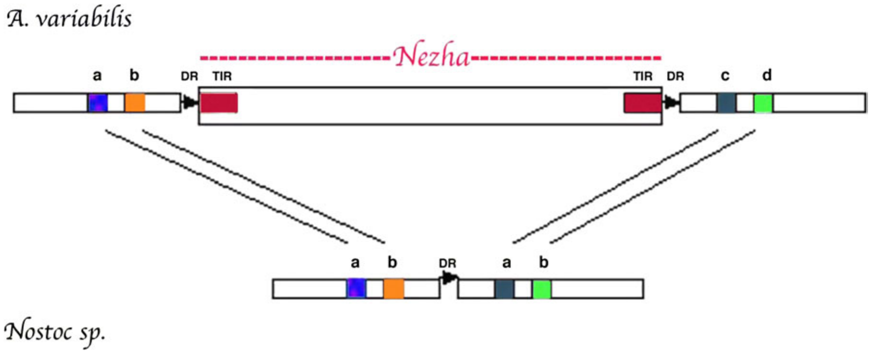

FIG. 2.—Diagrammatic representation of empty and filled site in homologous chromosomal regions in *Anabaena variabilis* and *Nostoc* sp. (based on Zhou et al. 2008) The *Nezha MITE* insertion is shown in *A. variabilis*. Shown also diagrammatically are the DRs and TIR. Genes depicted as "a" and "b" are orthologs between the two species. In another chromosomal region (not shown), *Nezha* can be found inserted into a site in *Nostoc* sp., while the same site is empty in *A. variabilis* (Zhou et al. 2008).

they are flanked by target site duplications (TDs), which consist of direct repeats (DRs). The core sequence of a MITE, however, lacks a transposase gene, although MITEs that carry open reading frames show amino acid sequences unrelated to transposase sequences (Delihas 2007). Another classic feature of MITEs is that they can fold into long stem loop structures at the RNA level (fig. 1*a*), and some are highly stable thermodynamically (Chen et al. 2008).

MITEs are multifaceted, for example, they can carry structure/function motifs, such as an integration host factor (IHF) binding site (Buisine et al. 2002), a methyltransferase binding site (Chen and Shapiro 2003), or promoter sequences (Black et al. 1995; Buisine et al. 2002; Snyder et al. 2003). Functionally, promoter strengths have been measured and RNA transcripts detected in transcriptional assays, but functional IHFs have not yet been observed (Siddique et al. 2011). Many repeats are found at 3′ end regions of genes and shown to be co-transcribed. Some regulate messenger RNA (mRNA) stability (De Gregorio et al. 2002, 2006). For example, the presence of an enterobacterial repetitive intergenic consensus (ERIC) sequence downstream of a gene may induce a conformational change in RNA transcripts and create a cleavage site for RNase E. This then can activate degradation of upstream mRNAs by 3′ to 5′ exoribonucleases (De Gregorio et al. 2005). There is an increased number of MITE and MITE-like units that are currently being discovered, and search programs such as MUST (Chen et al. 2009) can accelerate discovery of MITEs, for example, the newly found MITEs in cyanobacteria using the MUST program (Lin et al. 2011). With the availability of genome sequences from closely related organisms, the recent transposition of MITEs in some organisms has been proposed based on bioinformatics analyses (Zhou et al. 2008; Snyder et al. 2009).

As MITEs appear to be prevalent in cyanobacteria, we outline some recent findings. Kaneko et al. (2007) identified eight groups of putative MITE sequences in the cyanobacterium *Microcystis aeruginosa*. In a follow-up study, Lin et al. (2011) analyzed 17 cyanobacterial genomes and found several thousand MITE sequences. *Microcystis aeruginosa* also has a high abundance of IS elements, and a linear correlation was found between IS and MITE abundance. One group of MITEs is believed to be formed by a deletion within an IS element.

In other cyanobacteria, *Anabaena variabilis* and *Nostoc* sp., MITE sequences termed *Nezha* (approximately 130–170 bp) were characterized (Zhou et al. 2008). *Nezha* has signatures characteristic of MITEs, that is, TIRs that are similar in sequence to the TIRs of an intact transposon, DRs flanking the element, and predicted secondary structures that are highly stable thermodynamically. *Nezha* is predicted to be recently mobile based on analysis of empty and filled target sites in homologous chromosomal regions from closely related species (fig. 2). High percent identities and low *E* values show that adjacent genes in the empty and filled chromosomal sites are orthologous. *Nezha* shares the same TIR and nearly the same DR sequences as the IS IS*Npu3*. However, IS*Npu3* is only found in another species, *Nostoc punctiforme*. It is hypothesized that a similar IS transposase moved *Nezha* in *Nostoc* sp and *A. variabilis*. In a different study with *Anabaena* sp., five closely related MITE sequences have been detected (Wolk et al. 2010). As described for DNA repeat sequences in some Enterobacteriaceae species (Delihas 2007), several open reading frame fusions are found between open reading frames of *Anabaena* MITEs and chromosomal open reading frames (Wolk et al. 2010).

MITEs have also recently been characterized in additional bacteria. A repeat sequence called "Chunjie" also displays the classic signatures of a MITE. It was detected in *Geobacter uraniireducens Rf4*, a member of the deltaproteobacteria (Chen et al. 2008). The *Chunjie* sequences are 178–235 bp, contain 21 bp TIRs at each end, are A+T rich, and the terminal ends are flanked by 9 bp DRs. These

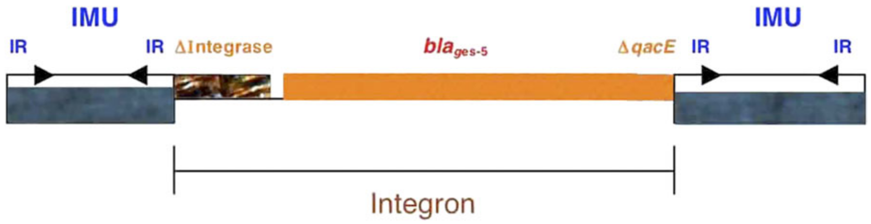

FIG. 3.—Diagrammatic representation of the defective integron flanked by identical IMU elements as found on *Enterobacter cloacae* plasmid pCHE-A (based on Poirel et al. 2009). The arrows represent the IMU inverted repeats (IR). Shown also are is the defective *int1* gene at the 5′ side (left), *bla*GES-5, the beta-lactamase gene cassette in the middle, and the defective quaternary ammonium salt gene *qacE* on the 3′ side. Lengths are not drawn to scale.

sequences can fold into highly stable secondary structures at the RNA level, for example, they show a delta G of approximately −98 to approximately 130 kcal/mol. Several *Chunjie* repeat sequences were found to overlap protein genes.

MITE-like sequences termed R0, R1, R2, R6, and R178 (table 1) were detected in *Pseudomonas fluorescens* (Silby et al. 2009). These range from 80 to 177 bp. Most have TIRs and can fold into stem loop structures. The inverted repeats of two MITE-like sequences R0 and R2 are identical to the inverted repeats found at the ends of IS elements present the same organism; thus, it is possible that the MITE-like sequences can be mobilized by these IS elements. *Pseudomonas fluorescens* also has regions devoid of repeats, which represents 40% of the genome. These regions are called "repeat deserts," which mostly have essential genes. There may have been an evolutionary selection process whereby cells that developed repeat sequence insertions in housekeeping genes could not survive.

Sequences comparable to the 127 bp ERIC MITE found in *E. coli* and related organisms (Hulton et al. 1991; Wilson and Sharp 2006) are particularly abundant in the chromosome of *Photorhabdus luminescens* (Duchaud et al. 2003). These MITEs are also found in *Xenorhabdus* (Ogier et al. 2010). Both *Photorhabdus* and *Xenorhabdus* belong to the Enterobacteriaceae family but are insect pathogens. These ERIC-type sequences have TIRs, TDs, and a $^{5'}TA^{3'}$ motif flanking both termini.

Snyder et al. (2009) provide evidence for the mobility of a Correia repeat (termed CREE) in *Neisseria gonorrhoeae* based on comparisons of chromosomal differences in locations of CREEs in two closely related strains of *N. gonorrhoeae.* The repeats are found in prophage regions in one strain and not in another, which indicates a recent transfer. In addition, many CREEs are found on the 5′ side of genes. Thus, the CREEs may influence gene expression at the transcriptional level. The same conclusion was reached by Siddique et al. (2011), who measured promoter strengths of the Correia repeat element in *N. meningitidis*.

Integron Mobilization

A special MITE termed integron mobilization unit (IMU) was detected in plasmid DNA of *Enterobacter cloacae* (Poirel et al. 2009). It encompasses a novel structure whereby two identical IMU sequences flank an intervening sequence that carries a defective class 1 integrase, a defective *qacE* gene and a beta-lactamase gene (bla_{ges-5}) that confers resistance to the antibiotic carbapenem (fig. 3). The integrase and *qacE* genes are features of class 1 integrons. The IMU sequence is 288 bp and contains TIRs. The spacer sequence is devoid of transposase sequences and displays no known motifs, but the IMU can fold into a predicted thermodynamically stable secondary structure at the RNA level. Importantly, transposition experiments show that the IMU-integron complex can be transposed in vivo to another plasmid by transposase acting in trans (Poirel et al. 2009). A five bp target site duplication (TD) is present at termini of the transposed IMU integron. The IMU TIR sequence is almost identical to the inverted repeat sequence of IS*Sod9* from *Shewanella oneidensis MR-1* (Poirel et al. 2009). The high similarity may be associated with recognition of the IMU by the transposase. The IMU represents the first MITE-like sequence found in plasmid DNA, and significantly, the first shown to be a nonautonomous transposable element using in vivo assays in prokaryotes (Poirel et al. 2009). The *Ent. cloacae* plasmid pCHE-A, which contains the IMU–integrase–antibiotic resistance gene complex, is nonself conjugative, thus the interesting question arises as to whether the IMU-containing integron can spread antibiotic resistance to other species via transposition.

A sequence similar to the *Ent. cloacae* IMU is found in a plasmid of another bacterial species but independent of an associated integron sequence. This IMU homolog is in *Aeromonas salmonicida* subsp. *salmonicida A449* plasmid 5 (as determined by a basic logic alignment search tool [Blast] search, Expect = $4e^{-30}$, Identity = 77%, found at nt positions 151346–151633, Accession number CP000646; N.D., unpublished). This putative IMU is in the intergenic region between locus ASA_P5G161, which encodes a truncated cobyrinic acid a,c-diamide synthase and locus ASA_P5G162, representing a hypothetical protein. Thus, the IMU may have a broader presence in genomes.

A defective Tn*402*-like integron is present in *Acinetobacter* sp. *str nfm2* (Gillings et al. 2009). This integron is flanked by identical copies of a 439 bp DR sequence, which appears to be MITE-like and is termed NFM2-MITE. The integron contains deletions at the 5′ and 3′ ends, which may have occurred when MITE sequences were fused with the integron. The outer ends of the MITE are flanked by a 5 bp DR. This MITE-like sequence is A+T rich, has TIRs, and has the potential to form a highly stable secondary structure. It may represent a special class of MITEs (Gillings et al. 2009); however, it has not been found to be transferable by transposase. On the other hand, experiments using polymerase chain reaction primers suggest that excision via homologous recombination is possible. Further analyses are needed to further define this interesting MITE-like-integron–associated element. Although the defective Tn*402*-MITE carries no antibiotic resistance genes, the Tn*402*-like integron is known to contribute to the proliferation of multi-antibiotic resistant genes (Gillings et al. 2009).

Repeat Sequences and Noncoding RNAs

Chinni et al. (2010) detected a small RNA transcript by northern blots from an intergenic sequence of *Salmonella typhi* that contains a heretofore uncharacterized repeat sequence of approximately 200 bp, a repeat that may be MITE-like. This repeat sequence and its overlapping RNA sequence map in a chromosomal region of *S. typhi* that represents a pathogenicity island. The RNA is growth regulated and appears during mid- to late-log phase. Sequences similar to the intergenic region in *S. typhi* are found in *E. coli*, but the RNA transcript has not been detected. Further detection of possible RNA transcripts in other *S. typhi* strains and nucleotide sequence comparisons of the repeat intergenic region in *S. typhi* strains and in *E. coli* may shed light on possible origins of the RNA and nature of the repeat sequence. For example, a comparison of sequences may show changes in the repeat sequence that formed a promoter for the putative RNA gene locus in *S. typhi*. A search for sequence changes that show upstream regulatory elements would also be useful as expression of the RNA is growth regulated.

Small RNA transcripts originating from intergenic chromosomal regions were detected in *N. meningitidis*. These transcripts are generated by an adjacent Correia element promoter (Siddique et al. 2011). In this case, the nt sequence downstream of the MITE Correia promoter is transcribed and not the Correia sequence.

Diverse Repeats in *Streptococcus pneumoniae*

Three repeat units, a tandem array of repeat sequences termed BOX, Repeat Unit of Pneumococcus (RUP), and *Streptococcus pneumoniae* Rho-Independent Terminator-like Element (SPRITE) were identified in *Streptococcus* sp. (Martin et al. 1992; Oggioni and Claverys 1999; Croucher et al. 2011). Predicted secondary structures of these repeats suggest possible roles at the RNA level (Croucher et al. 2011). For example, the SPRITE structure shows a motif similar to a Rho-independent termination motif, and its location in the genome has a bias in regions close to the 3′ ends of convergent genes. One identified BOX element has two T box riboswitch motifs, whereas another BOX element has open reading frames. Riboswitches can control gene expression through mRNA binding of a small target molecule and subsequent change in RNA conformation (Tucker and Breaker 2005). BOX is a nonautonomous transposable unit thought to be mobilized by IS*Stso1* (Knutsen et al. 2006). In addition, the BOX elements have been shown to be transcribed in *Str. pneumoniae* (Croucher et al. 2011). RUP has classic MITE properties with TDs and TIRs that were described before (Oggioni and Claverys 1999).

Comparison of the abundance of the three repeats between *Str. pneumoniae* clinical isolates and closely related species indicates that there was a past burst of repeat element movement in the genome of ancestors, but now, they appear dormant and their abundance is diminishing (Croucher et al. 2011). When inserted into intergenic regions, these repeats can function in gene regulation and can potentially be of benefit to the cell, but they are also found inserted into coding regions of a number of protein genes. Disruption of these genes can compromise the cell. From the evolutionary analysis of repeats, the authors conclude that streptococcal repeats are largely parasitic and may compromise the cell's ability to compete in its environment; thus, surviving species have fewer mobile elements.

Specialized Repeats in *Bacillus* sp.

Repeat sequences have been identified in the gram-positive *Bacillus* sp. (Tourasse et al. 2006), and 18 have been characterized (Kristoffersen et al. 2011). These fall into three groups whereby Group A sequences have properties of a nonautonomous transposable elements (Kristoffersen et al. 2011). This includes the repeat unit termed *bcr1* (approximately 155bp), which was extensively analyzed previously (Klevan et al. 2007). Comparisons between related *Bacillus* strains show a nonconserved genomic distribution with the repeat sequence flanked by 5 bp DRs in each case. This repeat element is transcribed. Base-pair compensatory changes are found to maintain a cruciform-like double-stranded structure at the RNA level. Figure 1*a* shows the *bcr1*-predicted secondary structure. However, a comparison of *bcr1* secondary structures between closely related *Bacillus* strains indicates that secondary structures vary in stability, and the authors suggest that *bcr1* repeats lost structural stability several times during evolution. The *bcr1* sequence may represent a special class of mobile sequences.

bcr5 is part of Group B repeats and is found associated with a gene cluster that contains a resolvase gene, as well as a transposase gene and a hypothetical protein gene (Kristoffersen et al. 2011). *bcr5* elements flank both ends of the resolvase-containing gene cluster. Although the *bcr5*-associated gene cluster does not encode an integrase, the cluster arrangement shows broad similarities to the integron clusters described above. *bcr5* does not have inverted repeats but has a predicted stable secondary structure. It has not been classified, but it does not appear to be MITE-like.

Group C elements are conserved phylogenetically in genomic locations. Some sequences may represent RNA transcripts and riboswitches as well. This work further extends the repeat element repertoire in the gram-positive bacteria.

Mycobacterial Interspersed Repetitive Units: Possible MITE-Like Sequences

A class of repeat sequences termed mycobacterial interspersed repetitive units (MIRUs) was found in several species of *Mycobacterium* (Supply et al. 1997).

The size ranges from approximately 40 to 100 bp, and these elements are found repeated approximately 40–50 times in the *Mycobacterium* genome. One of the sites containing the repeat sequence is found within an intergenic chromosomal region of *Mycobacterium tuberculosis*, between two conserved open reading frames that represent a conserved hypothetical protein and a serine/threonine phosphatase. The MIRU is transcribed as a polycistronic mRNA. Homologous regions in *Myc. leprae* do not contain the MIRU repeat sequence. MIRUs display some similarities to MITES. Comparison of empty and filled sites and the presence of tetranucleotide DRs on the 5′ and 3′ sides of filled site in *Myc. tuberculosis* suggest insertion by transposition. Although not stated as such, the MIRUs display imperfect TIRs and have internal inverted repeats that are rich in G–C bonds. MIRUs display open reading frames, and the terminal ends of the MIRU sequence overlap the adjacent genes in the polycistron mentioned above (Supply et al. 1997). Important to clinical diagnostics and epidemiological analyses, the mycobacterial interspersed repetitive sequences are currently used for *Myc. tuberculosis* genotyping for fast identification of clinical isolates (Supply et al. 2006).

REP Sequences—Multifunctional Elements

REP sequences are approximately 35 bp but range between 21 and 65 bp (Tobes and Pareja 2006). These are some of the smallest repeat sequences known. They were first found in Enterobacteriaceae species and later detected in *Pseudomonas* and *Stenotrophomonas* (Aranda-Olmedo et al. 2002; Silby et al. 2009; Nunvar et al. 2010). REP sequences are often found in high abundance with several hundred copies present in genomes either as single units or in clusters called bacterial interspersed mosaic elements (BIME) (Bachellier et al. 1994). They tend to be G+C rich and can fold into perfect or imperfect stem loops. In *Pse. syringae*, a bias for the positioning of the REP elements between convergent genes was found (Tobes and Pareja 2005). In *E. coli*, BIME clusters containing REP units have been associated with recombination. They can also affect mRNA stability (Stern et al. 1988). BIMEs form binding sites for IHF (Oppenheim et al. 1993), DNA polymerase I (Gilson et al. 1984), and DNA gyrase (Yang and Ames 1988), and it has been shown that DNA gyrase can cleave DNA in vivo in BIME regions (Espéli and Boccard 1997). Thus, REP units are intimately involved in molecular processes in the cell.

Although REP sequences do not display MITE signatures, Nunvar et al. (2010) hypothesize that REPs found in *Stenotrophomonas* sp. may be mobilized by transposase. The transposase gene termed REP-associated tyrosine transposase (RATY) was detected in *Stenotrophomonas* sp. by in silico methods (Nunvar et al. 2010). RATYS are related to the IS*200*/IS*605* family of transposases in terms of conserved amino acid motifs; however, they differ in that RATYs lack flanking stem loop sequences found in IS*200*/IS*605* (Ronning et al. 2005). Instead, several RATYs are flanked by inverted REP sequences, that is, 5′ to 3′ configuration of the REP sequence on the side of the transposase gene encoding the amino terminal end and a 3′ to 5′ REP configuration on the side encoding the carboxyl terminal. Because of the close association and conserved configuration between REPs and RATYs, this brings up the question of how *Stenotrophomonas* sp. REPs are mobilized. The authors hypothesize that RAYTs may be responsible for the proliferation of REP units, and thus, REPs may be transposable. Previously, Siguier et al. (2006) also suggested that REPs may be nonautonomous transposable elements.

Additional analyses in *Stenotrophomonas maltophilia* show that some REP sequences uniformly have a GTAG tetranucleotide sequence preceding the palindromic REP on its 5′ side (Rocco et al. 2010). These elements are termed *Ste. maltophilia* GTAG (SMAG). The SMAG REP sequences appear to alter the stability of upstream gene transcripts in *Ste. maltophilia*. The presence of one or two units of SMAG downstream of a gene has a stabilizing effect on the gene transcript yet a trimer SMAG appears to have a destabilizing effect. Thus, the SMAG sequences regulate gene expression at the post-transcriptional level in *Ste. maltophilia* but in a complex manner.

REP sequences have been implicated in the interaction with the host immune system. Synthetic oligodeoxynucleotides that mimic gram-negative bacterial REP unit sequences and their secondary structures were shown to stimulate the mammalian immune system via Toll-like receptor 9. This appears to be based on the CpG motif of REP sequences (Magnusson et al. 2007). It was hypothesized that REPs may also be involved in induction of human septic shock by pathogenic bacteria carrying REP sequences.

REP2 Repeats—Involvement in a Virulence Process

Intergenic repeat sequences that share no homologies with other known repeat sequences are found in *Neisseria* sp. (Parkhill et al. 2000). One is termed REP2, which ranges in size from 120 to 150 bp. REP2 is found repeated 26 times in intergenic regions of *N. meningitidis MC58* and 23 times *N. gonorrhoeae FA1090*. These repeats do not have TIRs and have no relationship with REP sequences. However, they have two internal inverted repeats that form predicted internal stem loops at the RNA level (fig. 1*b*). REP2 sequences appear to represent a unique class of repeats in that they contain a promoter sequence, a ribosome binding site, and an ATG initiation codon. They are often present upstream of open reading frames. In *N. meningitidis Z2491*, REP2 repeats are found immediately upstream of 14 genes that are coordinately upregulated during initial cell-to-cell contact with human cells (Morelle et al. 2003). Two of these encode the *pilC1* and *crgA* genes. PilC1 is an adhesin that mediates attachment of *N. meningitidis* to host cells. CrgA is a transcriptional regulator termed contact-regulated gene A protein (Deghmane et al. 2000). Both *pilC1* and *crgA* are induced with initial host cell contact, and both have REP2 sequences in their upstream regions. Thus, REP2 sequences participate in control of expression of genes essential for the interaction of *N. meningitidis* with human host cells (Morelle et al. 2003).

REP2 repeat sequences not only represent fusions of their translational start site with open reading frames but they also appear to contain mRNA 5′ UTR sequences. This fascinating repeat poses interesting questions concerning its origin and mechanism of proliferation in the neisserial genome. Did it originally arise from an upstream regulatory site and the 5′ UTR sequence of a protein gene?

Borrelia Sequences—Similarities to Neisserial REP2

Repeat elements in *Borrelia* chromosomes have not been reported. These chromosomes are small, for example, the *Borrelia burgdorferi* chromosome is approximately 0.9 Mb. There is tight packing of housekeeping and other genes and a paucity of intergenic space. Thus, there may be selective pressure to limit establishment of repeat elements in the *Borrelia* chromosome. However, repeats are present in *Borrelia* plasmid intergenic regions, albeit in a small copy number (Casjens et al. 2000). Sequence elements termed IR-A and IR-B that contain internal inverted repeats were found in both circular and linear plasmids (Dunn et al. 1994; Zuckert and Meyer 1996). These sequences have motifs strikingly similar to the REP2 repeat found in *N. meningitidis* (Parkhill et al. 2000; Morelle et al. 2003), that is, both the *Borrelia* IR sequences and the *Neisseria* REP2 sequence are located immediately upstream of genes and contain a promoter sequence, ribosome binding site, and an ATG start codon. Both sequences also have two internal inverted repeats close to their 5′ ends that form predicted internal stem loops 1 and 2 (fig. 1*b*). Dunn et al. (1994) originally showed the *Borrelia* stem loops at the DNA level.

Other intergenic sequences in *Borrelia* plasmids have inverted repeats identical in stem loop structure to the inverted repeats that flank termini of an IS related to IS*200*/IS*605* (Delihas 2009). The *Borrelia* IS 5′ and 3′ end flanking inverted repeats form stem loops; however, each has its own secondary structure signature. Significantly, these stem loop sequences are found associated with the 3′ ends of two types of putative lipoprotein genes and independent of transposase gene sequences. In one case (involving the IS 5′ end specific stem loop motif), the secondary structure is phylogenetically conserved at the RNA level with base-pair compensatory changes. In the other case, the IS stem loop motif associated with lipoprotein-1 genes is not conserved and the secondary structure appears to have undergone rapid evolutionary change between *Borrelia burgdorferi* strains. *Borrelia* plasmids contain many fragmented transposase gene sequences (Fraser et al. 1997). The IS*200*/IS*605* inverted repeat flanking sequences may be selectively conserved during decay of the IS element and based on findings of their evolutionary conservation or evolutionary development may form functional units when located near 3′ ends of genes.

CRISPRs—Short Palindromic Repeats Are Focal Points in a Specialized Regulatory System

CRISPRs differ from most other repeats described here in that these small sequences are part of a complex genetic arrangement. This consists of an array of palindromic DRs of approximately 28–49 bp. Linked with each repeat are variable spacer sequences that are fragments of foreign DNA (phage or plasmid DNA), or in some cases, host DNA. An array of protein genes termed CRISPR-associated (*cas*) genes are also closely associated with the palindromic repeat/spacer units. CRISPRs function as regulatory complexes. Recently, there has been great interest in the genetic and molecular characteristics of CRISPRs and for several reasons. First, the CRISPR system can function as a bacterial and archeal immune system, whereby CRISPR defends the organism from invading viral or plasmid DNA (Al-Attar et al. 2011). In addition, the mechanism of action of CRISPR systems has similarities to eukaryotic piwi-interacting RNAs (piRNA) mechanism of RNA-based immune system that inhibits mobile elements in germ line cells (Karginov and Hannon 2010; Marraffini and Sontheimer 2010a). Lastly, this genetic element offers an example of a type of Lamarckian inheritance in prokaryotes (Koonin and Wolf 2009). The CRISPR DNA complex was first found in *E. coli* (Ishino et al. 1987), although much of its characterization and functions have only been elucidated recently, approximately during the past 10 years (Barrangou et al. 2007).

Here, we provide a short description of the molecular/genetics aspects of CRISPR functions as they relate to immunity to invading or self-DNA. It is beyond the scope of this paper to describe the CRISPR complex in detail. There are numerous published reviews. We point out two recent reviews (Al-Attar et al. 2011; Terns MP and Terns RM 2011) and a perspectives paper (Makarova et al. 2011). These papers describe the history, evolution, and known mechanisms of action of the CRISPR-based defense system against virus or plasmid invasions of bacterial and archael cells.

There are basically three stages in the molecular and genetic processes of CRISPR function. During the acquisition stage, CRISPRs can capture fragments of foreign DNA from virus or plasmid sequences when challenged with the foreign DNA. A short segment (approximately 25–70 bp) of the foreign DNA, called a proto-spacer is inserted into the CRISPR locus of the host DNA between two palindromic repeat sequences. How the cell recognizes the short foreign DNA is unclear, but inserted foreign DNAs that have a small sequence (approximately a few nucleotides) adjacent to the spacer may be a recognition site (Mojica et al. 2009; Makarova et al. 2011). This small sequence is termed a proto-spacer-adjacent motif sequence (Mojica et al. 2009). Two Cas proteins may be involved in the acquisition process. Additional spacers are then added to form an array of spacer-palindromic sequence repeats. It is not known if the palindromic repeat sequences serve as Cas protein recognition sites for integration of DNA fragments into the CRISPR complex (Nam et al. 2011).

In the second stage, the CRISPR complex is transcribed and *cas* genes are transcribed and translated. In *E. coli*, the large precursor CRISPR transcript is processed by a ribonucleoprotiein complex termed Cascade (CRISPR-associated complex for antiviral defense). A Cas-specific endonuclease processes the RNA via cleavage at the base of the repeat stem loop sequence, and with additional trimming, the mature RNA is formed (Brouns et al. 2008; Gesner et al. 2011; Jore et al. 2011; Sashital et al. 2011). After processing, the Cascade complex retains RNA transcripts of foreign spacer DNA and stem loop repeat sequences and bound Cas proteins.

In the third stage, Cascade binds one strand of the target DNA via complementary base-pairing between spacer RNA and target DNA to form an RNA/DNA heteroduplex duplex. The target DNA strand is subsequently cleaved (Jore et al. 2011). Cas3 protein, which has endonuclease properties may be the major protein associated with target DNA inactivation in *E. coli*. (Brouns et al. 2008).

This molecular process that results in defense against invading DNA was shown to be present in organisms that include *Streptococcus*, *Staphylococcus*, and *E. coli* species. However, CRISPRs can display different roles in different microorganisms, and spacer DNA may consist of a fragment of a host protein gene.

In a clinical strain of *Pse. aeruginosa* lysogenized with the temperate phage DMS3, a CRISPR unit was found to be required for inhibition of biofilm formation and swarming motility (Zegans et al. 2009). One of the spacers of this unit, termed spacer 1 was found to be the determinant in inhibition (Cady and O'Toole 2011). However, spacer 1 has partial identity (approximately 84%) to phage gene *dms-42*. Thus, the correlation between this spacer sequence and inhibition of biofilm formation is puzzling, but spacer 1 was found to interact with the phage *DMS-42* gene. Another spacer in this CRISPR unit, spacer 2, was shown to carry a segment of temperate phage DMS3 DNA with 100% identity, but this does not appear to result in defense against the phage. Of interest is that the lysogenized *Pse. aeruginosa* strain that is unable to form a biofilm is a clinical isolate, as biofilm formation by *Pse. aeruginosa* is thought to be an important factor in establishment of chronic lung infections by *Pse. aeruginosa* (Palmer and Whiteley 2011).

Aklujkar and Lovley (2010) show that the capture of a fragment (proto-spacer) of the host gene *hisS* by a CRISPR complex results in inhibition of expression of host *hisS*, the histidyl-tRNA synthetase gene. Furthermore, they propose that during evolution, inhibition of expression of *hisS* by the CRISPR complex resulted in loss of ancestral genes that encode proteins containing a high percentage of histidines or have closely spaced histidines in their peptide chains. Ancestral genes that rely on histidyl-tRNA synthetase activity and were lost include those that express the subunit of an NADH dehydrogenase I complex and multiheme c-type cytochromes. Approximately 16 genes were lost during evolution of *P. carbinolicus*. It is believed that this organism survived because it retained another NADH dehydrogenase I complex, whereby a component protein does not have a cluster of histidines, and perhaps by relying on fermentation genes as well.

This is a rather far-reaching finding. The inhibition of a "self" gene activity by the CRISPR complex can be considered an autoimmune process in bacteria. This concept has been mentioned before (Marraffini and Sontheimer 2010b; Stern et al. 2010), but now has been shown experimentally by Aklujkar and Lovley (2010).

Repeats as Possible Engines for Genome Change

A comparison of genomic sequences from related species shows that repeats may be associated with high levels of intragenic recombination (Silby et al. 2009; Ogier et al. 2010; Kristoffersen et al. 2011). There is a striking lack of synteny between three closely related strains of *Pse. fluorescens* (Silby et al. 2009), and these strains vary greatly in repeat sequence abundance. For example, repeat elements R0 and R2 are highly represented in one strain (SBW25) but are absent or found in low abundance in others (strains Pf0-1 or Pf-5). In *Xenorhabdus*, ERIC-like sequences are in a chromosomal region of plasticity (termed Locus D). This region contains two ERIC (MITE) sequences, two transposase genes, and three truncated or disrupted genes. It was hypothesized that the ERIC sequence and the transposase genes play a role in plasticity of this chromosomal region (Ogier et al. 2010). In the *Bacillus cereus* group, repeat sequences *bcr4-bcr6* and their unique locations with respect to neighbor genes may be associated with genomic rearrangements (Kristoffersen et al. 2011).

In *E. coli*, the intergenic region between metabolic genes *folA* and *apaH* is highly variable (Mine et al. 2009). This is of special interest in that the toxin–antitoxin system encoded by the *ccdO157* gene complex is found between *folA* and *apaH* in *E. coli* O157:H7 EDL933. Some *E. coli* strains carry defective *ccdO157* genes or lack these genes completely in this region. Although the reason for this extensive instability is not known, an analysis of several hundred *E coli* strains shows that about 50% of the isolates contain a REP sequence in this region. REPs may in part account for the extreme plasticity of this region. Thus, evidence is accumulating to suggest that MITEs, REPs, and other repeats may play a role in genome dynamics during evolution in diverse species.

MITEs appear to play a role in evolution of individual genes. MITEs have been found inserted into gene loci of microcystin genes (*mcy*) in the cyanobacteria *Anabaena* isolated from the Baltic Sea, with the subsequent inactivation of these genes (Fewer et al. 2011). Microcystins are toxins that inhibit eukaryotic phosphatases (MacKintosh et al. 1990). MITE insertion into the *mcy* gene may provide a biological diversity in the population of the cyanobacteria. The *mcy* genes are considered ancient genes. The ability to synthesize microcystins has been repeatedly lost during evolution (Rantala et al. 2004). The *Anabaena* MITE may have been involved in this evolutionary process (Fewer et al. 2011).

IS elements recognize REP sequences as target sites for insertion. IS*1397* transposes specifically into REPs in *E.coli*, *S.enterica serovar Typhimurium*, and *Klebsiella* sp. (Wilde et al. 2001). IS*621* found in *E. coli* recognizes a 15-bp sequence in REP units and inserts into the REP sequences at its 3′ side but outside of the inverted repeat sequences. This type of insertion is found in 10 chromosomal loci (Choi et al. 2003). In addition, IS*1594*, which is present in *Anabaena* also inserts into REP-like sequences found in the *Anabaena* chromosome. Both IS*621* and IS*1594* belong to the S*110*/IS*492* family (Choi et al. 2003). Bioinformatics analyses show that REP sequences are targets for insertion of IS elements in *Pseudomonas, Neisseria*, and *Sinorhizobium* species (Tobes and Pareja 2006). Thus, the phenomenon of REPs serving as IS target sites for insertion is widespread and shows that REPs can affect plasticity.

Repeat Element Open Reading Frames, Insertion into Protein Genes

In addition to their prominent location in intergenic regions, many repeat sequences display open reading frames that are found fused in-frame with genomic open reading frames (Ogata et al. 2000; Delihas 2007; Croucher et al. 2011; Hot et al. 2011; Fewer et al. 2011). Some repeats are found fused internally into protein coding sequences (Ogata et al. 2000; Croucher et al. 2011; Hot et al. 2011). Others extend the 3′-terminal ends of protein genes (Delihas 2007; Croucher et al. 2011; Hot et al. 2011) or the 5′ ends (Croucher et al. 2011; Hot et al. 2011). An RUP insertion disrupts the coding sequence of the gene encoding a putative iron ABC transporter binding protein (Croucher et al. 2011). A repeat termed Betaproteobacterial repeat element (BRE) is present in *Bordetella* and other betaproteobacteria (Hot et al. 2011). Rather striking is the large number of protein genes (approximately 9 genes) that contain BRE inserts internally.

The possibility that repeat element fusions may create new proteins has been mentioned (Delihas 2008; Croucher et al. 2011). Of major interest is that a BOX element that potentially encodes a 42-amino acid predicted protein was found to be transcribed (Croucher et al. 2011). The detection of a translated protein product would show for the first time that a novel protein is formed by a repeat element.

Some repeats form fusions with sequences specifying protein domains such as the left-handed parallel beta helix, and others display motifs such as predicted transmembrane helices (Delihas 2007). Many of these fusions are annotated as hypothetical protein genes. It is not known if they are evolutionarily stable or transient, but some may serve as evolutionary reservoirs for new gene development (Treangen et al. 2009).

The annotation of genes whereby repeat sequences are shown to be part of an open reading frame can help define genetic loci better and/or raise questions concerning the locus. Several gene loci that contain repeat sequences have been annotated (Parkhill et al. 2000, 2001; Wei et al. 2003). But when these repeats are missed, this may raise questions concerning the locus. For example, locus *NMB0202* (Accession number NC_003112, coordinates 204159–204332) is annotated as a hypothetical 57-amino acid protein in *N. meningitidis MC58*. This sequence and three identical annotated sequences in related *N. meningitidis* strains contain a hypothetical translated 47-amino acid REP2 sequence; thus, the REP2 sequence represents approximately 82% of the open reading frame. This poses the question of whether this hypothetical gene locus is essentially an intergenic region that has a fusion of the REP2 open reading frame with a small adjacent open reading frame. REP2 sequences, in addition to having signatures at the DNA level, also display translated open reading frames.

Conclusions and Future Prospects

Small intergenic repeat sequences play an intricate role in molecular and functional aspects of the bacterial cell. Their individual signatures display a range of structure/function motifs, for example, MITE-like sequences straddle integrons in *Ent. cloacae and Acinetobacter* (Gillings et al. 2009; Poirel et al. 2009), the REP2 in *Neisseria* and the *Borrelia* IR sequences contain promoter sequences, a ribosome binding site and an ATG initiation codon followed by open reading frames (Dunn et al. 1994; Morelle et al. 2003), the Correia element in *Neisseria* and the REP cluster units (BIMEs) in *E. coli* carry an IHF binding site (Oppenheim et al. 1993; Buisine et al. 2002), and the Correia units also carry functional promoters (Siddique et al. 2011). Different repeats show a versatility in function such as regulation of expression of genes essential for interaction with human host cells (Morelle et al. 2003), serving as a recognition and cleavage site during RNA processing of the CRISPR transcript (Gesner et al. 2011), and serving as target sites for insertion of IS elements (Tobes and Pareja 2006). Neisserial intergenic mosaic elements (NIME) sequences may be involved in silent pilin gene recombination in *N. meningitidis* (Parkhill et al. 2000); these repeats are intimately associated with the pilE/S locus in a complex array of pilin genes and NIME sequences. The MITE-integron poses the question of a role in transfer of drug resistant genes.

In terms of bacterial evolution, repeat sequences are found at sites of plasticity in the bacterial genome (Mine et al. 2009; Silby et al. 2009; Ogier et al. 2010; Kristoffersen et al. 2011) and again, they can affect plasticity by serving as sites for IS integration in the genome (Tobes and Pareja 2006). In addition, mobile repeats may have influenced a cycle of active versus inactive genes during evolution (Fewer et al. 2011). As repeat elements can be detrimental when incorporated into essential genes, evolutionarily there may have been a selection against *Streptococcus* sp. carrying a large number of mobile repeats, as current populations of *Streptococcus* appear to have fewer elements than their ancestors (Croucher et al. 2011).

Did repeat sequences and associated molecular/functional signatures evolve independently in different microorganisms or were they transferred by horizontal transfer? For some repeats evidence is consistent with an independent origin. The REP2 unit in *Neisseria* and the *Borrelia* 180 IR sequences have negligible nucleotide sequence homology yet they both have similar structure/functional signatures and can be found immediately upstream of genes. MITEs in *Neisseria*, *E. coli*, and *Anabaena* have similar overall MITE features, but core sequences show no similarities in nt sequence or structure/function motifs. MITE-like sequences straddle integrons in both *Ent. cloacae* and *Acinetobacter* sp. Although their integrases are homologous, the MITE sequences show no similarities, and the internal structures of the integrons differ. This argues for an independent formation of MITE-integrons in these species, as previously proposed (Gillings et al. 2009).

How did these elements originate and how are they transferred? MITEs may have arisen by a selective conservation of IS-specific IR sequences during decay of a transposable element. Lin et al. (2011) proposed that a group of MITEs in *M. aeruginosa* originated by deletion of the IS core that encodes the transposase gene. In *Borrelia* IR IS-specific sequences may have been duplicated or were selectively conserved during decay of the IS sequence and transferred to 3′ end regions of putative lipoprotein genes (Delihas 2009). The very unusual REP2 repeat sequences may have originated from an upstream regulatory region of a gene that included the 5′ untranslated region and was subsequently duplicated and transferred to other chromosomal locations.

On mobility, MITEs can be transferred by a related transposase as exemplified by the in vivo transfer of the MITE-like sequence IMU by transposase (Poirel et al. 2009). By bioinformatics analysis, the *Nezha* MITE was shown to be recently transferred between species (Zhou et al. 2008). Inverted repeats of two MITE-like sequences in *Pse. fluorescens* are identical to the inverted repeats straddling the ends of IS elements present in the same organism (Silby et al. 2009), which hints at a transfer by the transposase. Thus evidence has accumulated to show or strongly suggest that many MITE sequences are mobilized by IS transposases. With respect to REP sequences, it has been hypothesized that the RATY may be responsible for the proliferation of REP units in the *Stenotrophomonas* chromosome (Nunvar et al. 2010).

Several repeat sequences have not been analyzed in terms of possible function, for example, ATR, REP 3-5, (Parkhill et al. 2000), and elements R0, R R2. R6, R178, and IR1_g (Silby et al. 2009). These may show additional intriguing properties.

Acknowledgment

I thank the Department of Molecular Genetics and Microbiology, State University of New York, Stony Brook, for research support.

Literature Cited

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.

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. 392:277–289.

Aranda-Olmedo I, Tobes R, Manzanera M, Ramos JL, Marqués S. 2002. Species-specific repetitive extragenic palindromic (REP) sequences in *Pseudomonas putida*. Nucleic Acids Res. 30:1826–1833.

Bachellier S, Clément JM, Hofnung M. 1999. Short palindromic repetitive DNA elements in enterobacteria: a survey. Res Microbiol. 150:627–639.

Bachellier S, Saurin W, Perrin D, Hofnung M, Gilson E. 1994. Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs). Mol Microbiol. 12:61–70.

Barrangou R, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712.

Black CG, Fyfe JA, Davies JK. 1995. A promoter associated with the neisserial repeat can be used to transcribe the uvrB gene from *Neisseria gonorrhoeae*. J Bacteriol. 177:1952–1958.

Brouns SJ, et al. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321:960–964.

Buisine N, Tang CM, Chalmers R. 2002. Transposon-like Correia elements: structure, distribution and genetic exchange between pathogenic *Neisseria* sp. FEBS Lett. 522:52–58.

Bureau TE, Wessler SR. 1992. Tourist: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283–1294.

Bureau TE, Wessler SR. 1994. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6:907–916.

Cady KC, O'Toole GA. 2011. Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. J Bacteriol. 193:3433–3445.

Casjens S, et al. 2000. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. Mol Microbiol. 35:490–516.

Chen SL, Shapiro L. 2003. Identification of long intergenic repeat sequences associated with DNA methylation sites in *Caulobacter crescentus* and other alpha-proteobacteria. J Bacteriol. 185:4997–5002.

Chen Y, Zhou F, Li G, Xu Y. 2008. A recently active miniature inverted-repeat transposable element, *Chunjie*, inserted into an operon without disturbing the operon structure in *Geobacter uraniireducens Rf4*. Genetics 179:2291–2297.

Chen Y, Zhou F, Li G, Xu Y. 2009. MUST: a system for identification of miniature inverted-repeat transposable elements and applications to *Anabaena variabilis* and *Haloquadratum walsbyi*. Gene 436:1–7.

Chinni SV, et al. 2010. Experimental identification and characterization of 97 novel npcRNA candidates in *Salmonella enterica serovar Typhi*. Nucleic Acids Res. 38:5893–5908.

Choi S, Ohta S, Ohtsubo E. 2003. A novel IS element, IS621, of the IS110/IS492 family transposes to a specific site in repetitive extragenic palindromic sequences in *Escherichia coli*. J Bacteriol. 185:4891–4900.

Correia FF, Inouye S, Inouye M. 1986. A 26-base-pair repetitive sequence specific for *Neisseria gonorrhoeae* and *Neisseria meningitidis* genomic DNA. J Bacteriol. 167:1009–1015.

Correia FF, Inouye S, Inouye M. 1988. A family of small repeated elements with some transposon-like properties in the genome of *Neisseria gonorrhoeae*. J Biol Chem. 263:12194–12198.

Croucher NJ, Vernikos GS, Parkhill J, Bentley SD. 2011. Identification, variation and transcription of pneumococcal repeat sequences. BMC Genomics 12:120.

De Gregorio E, Abrescia C, Carlomagno MS, Di Nocera PP. 2002. The abundant class of nemis repeats provides RNA substrates for ribonuclease III in *Neisseriae*. Biochim Biophys Acta. 1576:39–44.

De Gregorio E, Silvestro G, Petrillo M, Carlomagno MS, Di Nocera PP. 2005. Enterobacterial repetitive intergenic consensus sequence repeats in yersiniae: genomic organization and functional properties. J Bacteriol. 187:7945–7954.

De Gregorio E, Silvestro G, Venditti R, Carlomagno MS, Di Nocera PP. 2006. Structural organization and functional properties of miniature DNA insertion sequences in yersiniae. J Bacteriol. 188: 7876–7884.

Deghmane AE, et al. 2000. Intimate adhesion of *Neisseria meningitidis* to human epithelial cells is under the control of the crgA gene, a novel LysR-type transcriptional regulator. EMBO J. 19: 1068–1078.

Delihas N. 2007. Enterobacterial small mobile sequences carry open reading frames and are found intragenically-evolutionary implications for formation of new peptides. Gene Regul Syst Biol. 1:191–205.

Delihas N. 2008. Small mobile sequences in bacteria display diverse structure/function motifs. Mol Microbiol. 67:475–481.

Delihas N. 2009. Stem loop sequences specific to transposable element IS*605* are found linked to lipoprotein genes in *Borrelia* plasmids. PLoS One. 4:e7941. doi:10.1371/journal.pone.0007941.

Duchaud E, et al. 2003. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. Nat Biotechnol. 21:1307–1313.

Dunn JJ, et al. 1994. Complete nucleotide sequence of a circular plasmid from the Lyme disease spirochete, *Borrelia burgdorferi*. J Bacteriol. 176:2706–2717.

Espéli O, Boccard F. 1997. In vivo cleavage of Escherichia coli BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site. Mol Microbiol. 26:767–777.

Feschotte C, Zhang X, Wessler S. 2002. Miniature inverted repeat transposable elements and their relationship to established DNA transposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): ASM Press. p. 1147–1158.

Fewer DP, et al. 2011. Non-autonomous transposable elements associated with inactivation of microcystin gene clusters in strain of the genus *Anabaena* isolated from the Baltic Sea. Environ Microbiol Rep. 3:189–194.

Fraser CM, et al. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature 390:580–586.

Gesner EM, Schellenberg MJ, Garside EL, George MM, Macmillan AM. 2011. Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nat Struct Mol Biol. 18:688–692.

Gillings MR, et al. 2009. Mobilization of a Tn402-like class 1 integron with a novel cassette array via flanking miniature inverted-repeat transposable element-like structures. Appl Environ Microbiol. 75:6002–6004.

Gilson E, Clément JM, Brutlag D, Hofnung M. 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. EMBO J. 3:1417–1421.

Hancock CN, Zhang F, Wessler SR. 2010. Transposition of the Tourist-MITE mPing in yeast: an assay that retains key features of catalysis by the class 2 PIF/Harbinger superfamily. Mob DNA. 1:5. http://www.mobilednajournal.com/content/1/1/5

Higgins CF, Ames GF, Barnes WM, Clement JM, Hofnung M. 1982. A novel intercistronic regulatory element of prokaryotic operons. Nature 298:760–762.

Hot D, et al. 2011. Detection of small RNAs in *Bordetella pertussis* and identification of a novel repeated genetic element. BMC Genomics 12(1):207.

Hulton CS, Higgins CF, Sharp PM. 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. Mol Microbiol. 5:825–834.

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. 169:5429–5433.

Jacob F. 1977. Evolution and tinkering. Science. 196:1161–1166.

Jore MM, et al. 2011. Structural basis for CRISPR RNA-guided DNA recognition by cascade. Nat Struct Mol Biol. 18:529–536.

Kaneko T, et al. 2007. Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa NIES-843*. DNA Res. 14:247–256.

Karginov FV, Hannon GJ. 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell. 37:7–19.

Kikuchi K, Terauchi K, Wada M, Hirano HY. 2003. The plant MITE mPing is mobilized in another culture. Nature 421:167–170.

Klevan A, Tourasse NJ, Stabell FB, Kolstø AB, Økstad OA. 2007. Exploring the evolution of the *Bacillus cereus* group repeat element bcr1 by comparative genome analysis of closely related strains. Microbiology 153:3894–3908.

Knutsen E, Johnsborg O, Quentin Y, Claverys JP, Håvarstein LS. 2006. BOX elements modulate gene expression in *Streptococcus pneumoniae*: impact on the fine-tuning of competence development. J Bacteriol. 188:8307–8312.

Koonin EV, Wolf YI. 2009. Is evolution Darwinian or/and Lamarckian? Biol Direct. 4:42.

Kristoffersen SM, Tourasse NJ, Kolstø AB, Okstad OA. 2011. Interspersed DNA repeats bcr1-bcr18 of *Bacillus cereus* group bacteria form three distinct groups with different evolutionary and functional patterns. Mol Biol Evol. 28:963–983.

Lin S, et al. 2011. Genome-wide comparison of cyanobacterial transposable elements, potential genetic diversity indicators. Gene 473:139–149.

MacKintosh C, Beattie KA, Klumpp S, Cohen P, Codd GA. 1990. Cyanobacterial microcystin-LR is a potent and specific inhibitor of protein phosphatases 1 and 2A from both mammals and higher plants. FEBS Lett. 264:187–192.

Magnusson M, Tobes R, Sancho J, Pareja E. 2007. Cutting edge: natural DNA repetitive extragenic sequences from gram-negative pathogens strongly stimulate TLR9. J Immunol. 179:31–35.

Makarova KS, et al. 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 9:467–477.

Markham NR, Zuker M. 2005. DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Res. 33:W577–W581.

Marraffini LA, Sontheimer EJ. 2010a. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet. 11:181–190.

Marraffini LA, Sontheimer EJ. 2010b. Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 463:568–571.

Martin B, et al. 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucleic Acids Res. 20:3479–3483.

Mazzone M, et al. 2001. Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic *Neisseriae*. Gene 278:211–222.

Mine N, Guglielmini J, Wilbaux M, Van Melderen L. 2009. The decay of the chromosomally encoded ccdO157 toxin-antitoxin system in the *Escherichia coli* species. Genetics 181:1557–1566.

Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155(Pt 3):733–740.

Morelle S, Carbonnelle E, Nassif X. 2003. The REP2 repeats of the genome of *Neisseria meningitidis* are associated with genes coordinately regulated during bacterial cell interaction. J Bacteriol. 185:2618–2627.

Nam KH, Kurinov I, Ke A. 2011. Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA-binding activity. J Biol Chem. [cited 2011 Jun 21]. doi:10.1074/jbc.M111.256263.

Nunvar J, Huckova T, Licha I. 2010. Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes. BMC Genomics 11:44.

Ogata H, et al. 2000. Selfish DNA in protein-coding genes of *Rickettsia*. Science 290:347–350.

Oggioni MR, Claverys JP. 1999. Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*. Microbiology 145:2647–2653.

Ogier JC, et al. 2010. Units of plasticity in bacterial genomes: new insight from the comparative genomics of two bacteria interacting with invertebrates, *Photorhabdus* and *Xenorhabdus*. BMC Genomics 11:568. doi:10.1186/1471-2164-11-568

Økstad OA, et al. 2004. The bcr1 DNA repeat element is specific to the *Bacillus cereus* group and exhibits mobile element characteristics. J Bacteriol. 186:7714–7725.

Oppenheim AB, Rudd KE, Mendelson I, Teff D. 1993. Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli*. Mol Microbiol. 10:113–122.

Palmer KL, Whiteley M. 2011. DMS3-42: the secret to CRISPR-dependent biofilm inhibition in *Pseudomonas aeruginosa*. J Bacteriol. 193:3431–3432.

Parkhill J, et al. 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis Z2491*. Nature 404:502–506.

Parkhill J, et al. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature 413:523–527.

Poirel L, Carrër A, Pitout JD, Nordmann P. 2009. Integron mobilization unit as a source of mobility of antibiotic resistance genes. Antimicrob Agents Chemother. 53:2492–2498.

Rantala A, et al. 2004. Phylogenetic evidence for the early evolution of microcystin synthesis. Proc Natl Acad Sci U S A. 101:568–573.

Rocco F, De Gregorio E, Di Nocera PP. 2010. A giant family of short palindromic sequences in *Stenotrophomonas maltophilia*. FEMS Microbiol Lett. 308:185–192.

Ronning DR, et al. 2005. Active site sharing and subterminal hairpin recognition in a new class of DNA transposases. Mol Cell. 20:143–154.

Sashital DG, Jinek M, Doudna JA. 2011. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nat Struct Mol Biol. 18:680–687.

Sharples GJ, Lloyd RG. 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. Nucleic Acids Res. 18:6503–6508.

Siddique A, Buisine N, Chalmers R. 2011. The transposon-like Correia elements encode numerous strong promoters and provide a potential new mechanism for phase variation in the meningococcus. PLoS Genet. 7:e1001277.

Siguier P, Filée J, Chandler M. 2006. Insertion sequences in prokaryotic genomes. Curr Opin Microbiol. 9:526–531.

Silby MW, et al. 2009. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. Genome Biol. 10:R51. doi:10.1186/gb-2009-10-5-r51.

Snyder LA, Cole JA, Pallen MJ. 2009. Comparative analysis of two *Neisseria gonorrhoeae* genome sequences reveals evidence of mobilization of Correia Repeat Enclosed Elements and their role in regulation. BMC Genomics 10:70.

Snyder LA, Shafer WM, Saunders NJ. 2003. Divergence and transcriptional analysis of the division cell wall (*dcw*) gene cluster in *Neisseria spp*. Mol Microbiol. 47:431–442.

Stern MJ, Ames GF, Smith NH, Robinson EC, Higgins CF. 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell 37:1015–1026.

Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genet. 26:335–340.

Stern MJ, Prossnitz E, Ames GF. 1988. Role of the intercistronic region in post-transcriptional control of gene expression in the histidine transport operon of *Salmonella typhimurium*: involvement of REP sequences. Mol Microbiol. 2:141–152.

Supply P, et al. 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol. 44:4498–4510.

Supply P, Magdalena J, Himpens S, Locht C. 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol Microbiol. 26:991–1003.

Terns MP, Terns RM. 2011. CRISPR-based adaptive immune systems. Curr Opin Microbiol. 14:321–327.

Tobes R, Pareja E. 2005. Repetitive extragenic palindromic sequences in the *Pseudomonas syringae pv. tomato* DC3000 genome: extragenic signals for genome reannotation. Res Microbiol. 156:424–433.

Tobes R, Pareja E. 2006. Bacterial repetitive extragenic palindromic sequences are DNA targets for insertion sequence elements. BMC Genomics 7:62. doi:10.1186/1471-2164-7-62.

Tourasse NJ, Helgason E, Økstad OA, Hegna IK, Kolstø AB. 2006. The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. J Appl Microbiol. 101:579–593.

Treangen TJ, Abraham AL, Touchon M, Rocha EP. 2009. Genesis, effects and fates of repeats in prokaryotic genomes. FEMS Microbiol Rev. 33:539–571.

Tucker BJ, Breaker RR. 2005. Riboswitches as versatile gene control elements. Curr Opin Struct Biol. 15(3):342–348.

Wei J, et al. 2003. Complete genome sequence and comparative genomics of *Shigella flexneri serotype 2a strain 2457T*. Infect Immun. 71:2775–2786.

Wilde C, Bachellier S, Hofnung M, Clément JM. 2001. Transposition of IS1397 in the family Enterobacteriaceae and first characterization of ISKpn1, a new insertion sequence associated with *Klebsiella pneumoniae* palindromic units. J Bacteriol. 183:4395–4404.

Wilson LA, Sharp PM. 2006. Enterobacterial repetitive intergenic consensus (ERIC) sequences in *Escherichia coli*: evolution and implications for ERIC-PCR. Mol Biol Evol. 23:1156–1168.

Wolk CP, Lechno-Yossef S, Jäger KM. 2010. The insertion sequences of *Anabaena* sp. *strain PCC 7120* and their effect on its orfs. J Bacteriol. 192:5289–5303.

Yang G, Nagel DH, Feschotte C, Hancock CN, Wessler SR. 2009. Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. Science 325:1391–1394.

Yang Y, Ames GF. 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. Proc Natl Acad Sci U S A. 85:8850–8854.

Zegans ME, et al. 2009. Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. J Bacteriol. 191:210–219.

Zhou F, Tran T, Xu Y. 2008. *Nezha*, a novel active miniature inverted-repeat transposable element in cyanobacteria. Biochem Biophys Res Commun. 365:790–794.

Zuckert WR, Meyer J. 1996. Circular and linear plasmids of Lyme disease spirochetes have extensive homology: characterization of a repeated DNA element. J Bacteriol. 178:2287–2298.

Associate editor: Bill Martin