ARTICLE IN PRESS

S0923-2508(02)01344-X/REV AID:1344 Vol.●●●(●●●)
ELSGMLTM(RESMIC):m5 2002/06/28 Prn:12/07/2002; 11:04

RESMIC1344

P.1 (1-7)
by:CG p. 1

ELSEVIER

Research in Microbiology ●●● (●●●●) ●●●–●●●

Research in Microbiology

www.elsevier.com/locate/resmic

Repetitive DNA, genome system architecture and genome reorganization

James A. Shapiro

Department of Biochemistry and Molecular Biology, University of Chicago, 920 E, 58th street, Chicago, IL 60637, USA

Received 10 April 2002; accepted 11 June 2002

Abstract

Repetitive DNA elements are major organizational components of the genome inolved in replication, in transmission to daughter cells, and controlling expression of genomic coding sequences. Repetitive elements format the genome system architecture characteristic of each taxonomic group. Appreciating the functional significance of repetitive DNA provides new concepts of genome organization and genome reorganization in evolution. © 2002 Éditions scientifiques et médicales Elsevier SAS. All rights reserved.

Keywords: DNA repeats; Mobile genetic elements; Genetic regulation; Genome maintenance; Evolution; Adaptive mutation; Non-random genetic change

1. Pioneering studies of repetitive DNA in bacteria

Maurice Hofnung's science was characterized by penetrating thoughtfulness and the courage to pursue unfashionable topics. Maurice was one of the first to introduce computational genomics to the Institut Pasteur, and his laboratory discovered the first class of complex repetitive DNA elements in bacterial genomes, the BIMEs (bacterial interspersed mosaic elements) [8,9]. Recently, I had the privilege of co-editing a special issue of Research in Microbiology with Maurice on microbial DNA repeats [10].

When Maurice began his work on repetitive DNA, this was far from a fashionable subject. Orgel and Crick had recently coined the term "junk DNA" to describe the excess, supposedly non-coding DNA that did not directly determine the primary sequences of RNA and polypeptide chains [22]. Unfortunately, this term has stuck in the minds of many biologists and geneticists, despite the fact that it was based purely on ignorance and failed to take account of an existing literature that documented many examples of specificity and function for repetitive DNA sequences. Today, this attitude is changing due to the accumulation of new information about repetitive DNA. The most important event was the publication last year of the draft human genome, showing that less than 5% comprises protein-coding exons and well over 60% is highly repetitive (43% in dispersed mobile genetic elements, or MGEs, plus 18% in unsequenced heterochromatin regions composed largely of tandem repeat arrays; see Fig. 1) [11].

In this article, I will write in defense of repetitive DNA and attempt to explain why it is an essential component of the genome. In fact, thinking about the role of repeated sequence elements leads us into a 21st Century view of genome function and opens up new ways of thinking about the evolution of genomes as complex information systems (Table 1).

2. Functional roles of repetitive elements

Our understanding of the roles played by repetitive elements extends back to two seminal episodes in the history of analyzing genome function: The elaboration of the operon model for control of transcription [12] and the recognition that distributed repeat sequences can form the physical basis for integrated genomic networks [3]. Fig. 2 summarizes the history of how we have successively conceptualized the *lac* operon as it was deconstructed from a single point on a genetic map into an interactive system of regulatory and protein-coding components.

The key advance in our thinking was the identification of the operator (O) as a cis-acting site where the repressor recognizes the DNA, quite a different entity from the classical notion of a "gene" encoding a product related to a specific phenotype. Today, our understanding of how the

E-mail address: jsha@midway.uchicago.edu (J.A. Shapiro).

PII: S0923-2508(02)01344-X

Fig. 1. Dispersed and tandem arrays of repetitive DNA elements.

Table 1
A summary of the overall argument

- The genome has multiple functions and uses multiple sequence codes to carry them out
- Genomic coding involves multicomponent systems, not units
- Repetitive sequences provide common signals for interaction with proteins and RNAs at distinct genomic locations, thereby integrating multiple loci into genome-wide systems
- Functional and computational formatting by repetitive DNA elements defines a genome system architecture for each taxon
- The systemic nature of genomes implies that functional changes occur chiefly by rearrangement of modular components
- Cellular natural genetic engineering functions carry out regulated, non-random DNA rearrangements to generate new functions and new system architectures
- In the 21st Century, systems engineering will become our chief metaphor for genome reorganization in evolution

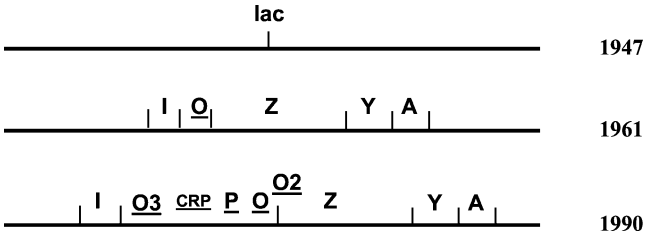

Fig. 2. Steps in the historical deconstruction of the *lac* operon from a point on a genetic map in 1947 to our current view of an integrated system of regulatory and coding sequences. Protein binding sites are underlined. See Refs. [12,30–32] for details.

genome functions is based largely on the notion that there are many such cis-acting sites carrying codes for genome interaction with other cellular components, such as the DNA replication, chromosome segregation, and transcription complexes.

The *lac* operon illustrates in a simple and direct way how each genetic locus is organized to facilitate cellular computation about genome functioning, in this case making the decision when to transcribe *lacZYA*. By virtue of a network of cell-wide connections between DNA sites and cellular activities for transporting and metabolizing sugars and ATP, the *Escherichia coli* cell is able to discriminate between glucose and lactose and compute the following algorithm: "IF lactose present AND glucose not present AND cell can synthesize active LacZ and LacY, THEN transcribe *lacZYA* from *lacP*" [31]. As Britten and Davidson recognized [3], multiple copies of cis-acting sites could create control networks leading to the integration of many genetic loci into coordinately functioning systems. The iteration of the CRP binding site for the cAMP receptor protein is a good example. Since *E. coli* cells use the level of intracellular cAMP as a molecular indicator of the availability of glucose in the environment, genetic loci containing CRP are integrated into a sophisticated regulon capable of responding to changes in carbohydrate metabolism. Eukaryotic development provides even more elaborate illustrations of combinatorial complexity and computational sophistication [31,32].

In addition to MGEs and computationally/functionally organized protein binding sites, repetitive DNA comes in a wide variety of sizes and arrangements. Some of these are similar to the dispersed repeats discovered in Maurice's laboratory and described by other articles in this issue. A major class of repeats consist of simple sequence repeats (SSRs) arrayed in tandem. SSRs and other tandem arrays include basic units that range in size from one or two base pairs (homopolymeric tracts and dinucleotide repeats) up to the several hundred base pairs that characterize the tandem repeats surrounding the centromeres of most eukaryotic chromosomes [32]. These highly regular sequence structures generally assume a different conformation from the less regular regions of the genome and have profound effects on transcription and other aspects of genome function. In eukaryotes, regions rich in tandemly repeated DNA elements form heterochromatin [32].

Living cells use repetitive DNA sequences in various ways to affect the expression of coding sequences. In bacteria, many organisms with reduced genome size use recombination and changes in the size of repeat arrays to alter the nature and level of expressed proteins (Table 2).

In eukaryotes, repeated DNA sequences have similar effects on protein synthesis. The effects of SSR expansion and contraction are to "tune" the level of expression [31,32]. Longer repeat arrays inhibit expression, while contraction of the array relieves inhibition. This effect is becoming more widely known through certain inherited human disease states which result from loss of function when repeat arrays expand.

The consequence of placing many genetic loci near repeat-rich heterochromatic regions is to shut off expression during development, a phenomenon known as "position effect variegation" [31,32]. The position effect literature was long considered to be a curiosity of little relevance to the mainstream of genetics, but nowadays position effect is seen as one example of epigenetic control of genome expression.

It is clear that repetitive DNA elements play a major role in the control of how RNA and protein coding information

Table 2
Some functions of bacterial repeats in regulation of protein synthesis

CRP sites	Coordinate regulation in response to glucose metabolism
Promoters	Coordinate regulation in response to sigma factors
Dam methylation sites	Regulation of promoter activity (Tn10), fimbrial phase variation (*E. coli; Salmonella*) [2,21]
Homopolymer tracts	Tuning of promoter activity (*N. meningitidis*) [26,33]; phase variation of surface proteins, polysaccharide biosynthetic enzymes, restriction and modification proteins in *Campylobacter jejuni* [15], *N. meningitides* [23, 24,27]
Tandem pentamers	Opacity protein phase variation by array size in coding sequence (*Neisseria*) [13]
Tandem heptamers	Adhesin phase variation by array size in promoter (*H. influenzae*) [5]
DHS (200 bp), 17–18 bp repeats	Vmp antigenic variation by expression site switching (*Borrelia*) [1,34]
NIMEs (dRS, RS), Sma/Cla repeats	Pilin antigenic and phase variation by gene conversion (*Neisseria*) [23,27]
vis (35 bp)	Vsp phase variation by site-specific inversion (*M. bovis*) [16]

Table 3
Some genomic functions of repeat elements in bacteria

DNA uptake sequences	Permit intrageneric transformation (*Neisseria, Haemophilus*) [10,23]
chi and chi-like sequences	Initiation of homologous recombination for double-strand break repair, integration of exogenous sequences, rearrangements [10]
Tandem repeats	Replication origins [6,18]
Dam methylation sites	Methyl-directed mismatch repair; segregation of newly replicated oriC sequences [18]; inhibit transposase action (Tn10)
Telomeres	Maintenance of linear replicons (*Borrelia*) [7,35]
IS elements	Integrate laterally transferred DNA; mobilize resistance, catabolic and other determinants to new locations [4,10]
59 bp elements, VCR elements	Insert and remove cassettes from integrons in plasmids, transposons, build multicistronic operons [10]

is read from genome sequences. But coding is only one of many functions the genome fulfills in the information economy of the cell. One of the best metaphors for the role the genome plays is to consider it the long-term information repository of the cell and to think of DNA as a data storage medium that must be dynamically accessed, replicated, proofread, repaired, packaged, transmitted, and reprogrammed when necessary. These processes each employ their own distinctive genetic codes, comprised of signals that are generally present many times within the genome. As in electronic information systems, the various files have to be tagged with content-independent identifiers for access, error correction, accurate data transmission and for storing new information and programs. In bacterial genomes, which are often cited as being free of repetitive DNA, as well as in eukaryotes, these systemic aspects of functioning involve repetitive sequences (Table 3).

3. Taxonomic specificity of repetitive DNA, genome system architecture and their significance for evolution

The aspects of genome organization sketched out above and elsewhere [31,32] involve two essential features. The first one is that all genomic elements, down to the level of the individual nucleotide pair, constitute multicomponent *systems*. This generalization applies to protein coding sequences (systems of triplet codons arranged into higher order regions encoding evolutionarily mobile domains), cis-acting regulatory sites (systems of nucleotides and organized motifs), genetic loci (systems of regulatory and coding regions), chromosome domains (systems of genetic loci and variously formatted chromatin), whole chromosomes, and dispersed multilocus systems throughout the genome. The second essential feature is that all of this modular, hierarchical organization is formatted by repetitive DNA elements in a way that was predicted by Britten and Davidson [3] but which is far more involved than they could have anticipated in the 1960s.

Another important fact about repetitive DNA is that it is the most highly variable, and consequently the most taxonomically specific, component of the genome. Species that may be highly related in their protein coding DNA often differ markedly in their repetitive DNA content. For example, each order of mammals shares largely the same set of proteins, but they contain quite distinct collections of tandemly arrayed centromeric repeats and dispersed reverse-transcribed SINEs (short interspersed nucleotide elements) [31,32]. Thus, the easiest way to identify a mammalian cell culture is by examining its content of centromeric satellite DNA or SINE elements. The specificity of repetitive DNA is so exquisite that examination of microsatellite SSR repeats forms the basis of forensic DNA analysis for identifying individuals. It is not by accident that Maurice and Agnes Ullman took out a patent for using DNA repeats as a diagnostic tool for identifying bacterial cultures.

Given the many genomic roles played by repeat elements and considering their taxonomic specificity, it is not hard to see that changes in repetitive DNA can be linked to the formation of quite distinct groups of organisms. In the case of the centromeric DNA repeats, such a connection between phylogenetic divergence and alteration in repeat DNA content is obvious. Similarly, different species and genera may

Table 4
Implications of genome system architecture for evolutionary change

- Novelty arises by rearrangement of modular components (Lego-like)
- Important effects of changes in repetitive "non-coding" DNA
- Systemic changes in genomes (reformatting, creation of new multi-locus systems)

share virtually all their proteins but differ markedly in the way those proteins are expressed during development or in response to outside cues. These differences often result from altered regulatory configurations, either novel combinations of transcriptional control signals or differences in chromatin formatting. In both cases, we have seen that such adaptively significant changes can result from redistribution of repetitive elements.

From the foregoing, it may be argued that repetitive DNA formatting of essential genome functions is a major aspect of defining a *genome system architecture* characteristic of each taxon. The idea of a system architecture self-consciously recalls the differences between system architectures in computer operating systems. The various architectures generally accomplish the same tasks, but they organize and control them differently. In the same way, different cells and organisms can use distinct architectures to accomplish parallel goals. For example, bacteria with larger genomes ($\geqslant 4$ MB) tend to use regulatory proteins to control the variation in protein expression, while bacteria with smaller genomes ($\leqslant 2$ MB) have relatively fewer regulatory proteins and often use expansion and contraction of tandem repeat arrays for the same purpose (Table 2).

Looking at the growing database of whole genome sequences, it is apparent that some of the most basic genetic changes in evolution occur by reassortment of component genomic modules rather than by the accumulation of large numbers of localized changes in base sequence. Differences in repetitive DNA content and regulatory regions have already been mentioned. A great deal of attention is now being focussed on segmental duplications in genomes which range in length from a few thousand to many millions of base pairs [32]. Even at the level of protein evolution, the dominant processes appear to be domain swapping and domain accretion to generate molecules with novel functions and specificities [11]. This kind of Lego-like process is exactly what the concept of genome system architecture predicts (Table 4). How does it fit with other lessons from molecular genetics?

4. Natural genetic engineering – cellular control of genome reorganization

One of the major discoveries of molecular genetics has been the universality of cellular mechanisms for repairing, mutating and rearranging DNA. The series of discoveries that followed from analysis of induced and spontaneous mutagenesis extended McClintock's pioneering observations on chromosome healing and transposable elements to an extent that could never have been predicted [19]. All sequenced organisms contain biochemical systems for repairing and recombining their genomes, and it is only a small minority of highly specialized bacterial parasites that lack active MGEs. Thus, we can say that all cells have the capacity for natural genetic engineering, and the full potential of these processes includes exactly the modular rearrangement functions needed for rapid evolution of genomic systems, subsystems, and overall system architectures (Table 5).

Examination of sequenced genomes provides abundant evidence for the activity of natural genetic engineering functions in evolutionary history. There are multiple drug resistance determinants (plasmids, transposons, integrons) and pathogenicity islands in prokaryotes, regulatory regions, gene family amplifications, segmental duplications, and even the appearance of new exons [20] in eukaryotes, all resulting from the action of MGEs or other DNA rearrangement activities (summarized in Refs. [31,32]). In prokaryotes, the rearrangements typically involve DNA-based elements, while retroposon-based functions appear to be more common in mammals and other higher eukaryotes. The fact that the human and other genomes contain dispersed MGEs as a very high percentage of their total DNA content means that these genomes were constructed to a very large extent by bursts of transposition and retrotransposition events [11]. The taxonomic specificities of the MGEs and other DNA repeats means that such bursts have occurred repeatedly in evolution.

The fact that much (probably the vast majority) of significant evolutionary change in genomes results from the action of cellular biochemical complexes has profound implications for understanding how organisms create genomic novelty. Instead of change due to stochastic, random events and replication errors (all of which are subject to proofreading and repair), we now see DNA reorganization as a cell biological process, in which the synthesis and activity of the responsible biochemical functions can be highly regulated. Thus, change can occur episodically, when it is most needed, by response to challenge and stress. An example is the adaptive mutation phenomenon, first described in a bacterial system involving genetic fusions mediated by the transposable bacteriophage Mu (Fig. 3) [29,30]. From studies of Mu-mediated fusions and other adaptive mutation systems, it is becoming clear that cellular control functions such as RpoS sigma factor, ClpXP and Lon proteases, CRP activator protein, and the SOS regulon respond to oxidative starvation conditions and activate various natural genetic engineering activities, including MGEs and mutator polymerases, to produce a hypermutable state [14,25]. Analogous cases of stress- and hybridization-induced activation of natural genetic engineering functions in vegetatively and sexually reproducing eukaryotes are well documented and have been summarized elsewhere [31].

Not only does natural genetic engineering introduce temporal specificity into the process of genome change. The

Table 5
Some natural genetic engineering capabilities (see Ref. [32] for detailed citations)

DNA reorganization functions	DNA rearrangements carried out
Homologous recombination systems	Reciprocal exchange (homologous crossing-over); amplification or reduction of tandem arrays (unequal crossing-over); duplication, deletion, inversion or transposition of segments flanked by dispersed repeats; gene conversion
Site-specific recombination	Insertion, deletion or inversion of DNA carrying specific sites; serial events to build operons, tandem arrays
Site-specific DNA cleavage functions	Direct localized gene conversion by homologous recombination (mating type interconversion in *S. cerevisiae*); create substrates for gene fusions by NHEJ (VDJ recombination in the immune system)
Non-homologous end-joining (NHEJ)	Precise and imprecise joining of broken DNA ends; create genetic fusions; facilitate localized hypermutation
Mutator polymerases	Localized hypermutation
DNA transposons	Insertion, excision; carry signals for transcriptional control, RNA splicing and DNA bending; non-homologous rearrangements of adjacent DNA sequences (deletion, inversion or mobilization to new genomic locations); amplifications
Retroviruses and other terminally repeated retrotransposons	Insertion and amplification; carry signals for transcriptional control, RNA splicing and chromatin formatting; mobilization of sequences acquired from other cellular RNAs
Retrotransposons without terminal repeats	Insertion; amplification; carry signals for transcription and RNA splicing; reverse transcription of cellular RNAs; insertion of the cDNA copies; amplification and dispersal of intron-free coding sequences; mobilization of adjacent DNA to new locations (e.g., exon shuffling)
Terminal transferases	Extend DNA ends for NHEJ; create new DNA sequences in the genome
Telomerases	Extend DNA ends for replication

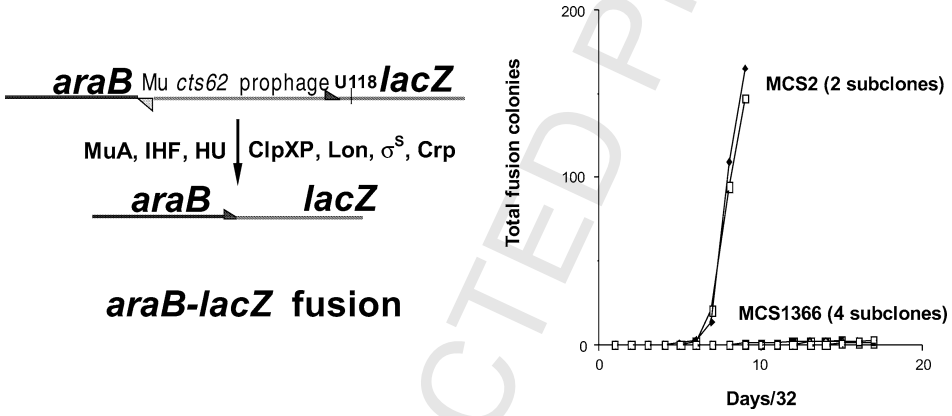

Fig. 3. Adaptive mutation in the Mu-mediated *araB-lacZ* fusion system developed by Casadaban [29]. The left panel summarizes the process of fusion formation, and the right panel illustrates the kinetics of appearance of fusion colonies on selection plates for strain MCS2 (with MuA transposase) but not for strain MCS1366 (lacking MuA transposase). The DNA rearrangements creating fusions require the MuA, IHF and HU proteins [30]. Activation of MuA expression by aerobic starvation on selective medium and subsequent fusion formation take several days and require the ClpXP and Lon proteases, the RpoS sigma factor, and Crp protein [14]. The absence of fusion colonies in the first five days of incubation demonstrates that fusions do not occur during normal growth of MCS2 cultures but must be triggered by aerobic starvation under selective conditions [17].

process is inherently non-random. Even with targeting, the movement of a defined segment of DNA, such as a proretrovirus containing complex transcriptional and post-transcriptional control sites, is far from a random event. Moreover, we know that natural genetic engineering processes can be specifically activated and targeted to particular genomic sites for a defined purpose because our lives depend upon it—the DNA rearrangements in the lymphocytes which produce antigen recognition molecules are controlled in just this way [32]. Targeting in immune system rearrangements appears to depend upon two factors: The presence of specific recognition signals for the RAG1,2-transposase and a coupling between transcription and the formation of double-strand breaks in DNA regions which will be joined together by NHEJ functions in novel combinations. Targeting of somatic hypermutation to particular regions of immunoglobulin coding sequences also appears to involve a coupling with transcription. In an analogous fashion, there is accumulating evidence that various MGEs (homing introns, yeast Ty elements, *Drosophila* P factors) use site-specific endonucleases, transcriptional control molecules, and chromatin formatting to target their non-random alterations of genomic information (summarized in Ref. [32]).

The activation and targeting of natural genetic engineering functions invalidates the assumption that each genetic change is a rare, unique event independent of every other change. For example, activation of a specific class of MGEs, such as P factors in *Drosophila* hybrid dysgenesis [31], leads

Table 6
21st Century perspective: Evolution as systems engineering

- Major evolutionary change by rearrangement of pre-existing modules:
  - following duplication
  - in the "facultative" R & D sector of the genome (M. Golubovsky, A. Katzenellenboigen)
  - functional significance of changing repetitive DNA
- Large-scale genome reorganization by activation of natural genetic engineering systems in response to major challenges—i.e., rapid, episodic changes throughout the genome during periods of crisis
- Targeting of DNA changes to particular regions of the genome, thereby enhancing the probability of generating useful new multi-locus systems
- Natural selection eliminates misfits after episodes of genome reorganization
- Fine-tuning of survivors carrying novel genomic systems by micro-evolution

to a temporally coordinated series of mechanistically similar mutations. Such mutations distribute a well-defined set of DNA signals to different locations in the genome. When synchronous mutational events are combined with targeting to regions of the genome that share transcriptional control signals, the mechanistic basis exists for functionally coordinated changes at diverse locations throughout the genome. In this way, the non-random behavior of natural genetic engineering functions provides a way to begin thinking about one of the major problems in evolutionary theory: The rapid invention of complex adaptive systems involving the products of multiple genetic loci.

5. Conclusion: A 21st Century view of evolution

It is clear from the foregoing discussion that experience with repetitive DNA and mobile genetic elements leads to some fundamentally new ways of thinking about basic issues in genome function, genome organization and genome evolution. I expect that the 21st Century will adopt a very different perspective on the evolutionary process compared to the dominant neo-Darwinian Modern Synthesis formulated at the middle of the 20th Century. The basic intellectual metaphor will be systems engineering, and the microevolutionary processes now emphasized will be relegated to a secondary role of fine tuning once major adaptive innovations have been constructed by natural genetic engineering (Table 6).

For both personal and scientific reasons, I will always associate the memory of Maurice Hofnung with two fundamentally important aspects of genomes that were completely unanticipated when we began our adventures in molecular genetics: Natural genetic engineering and repetitive DNA. The personal reason comes from the deep pleasure and camaraderie that Maurice and I shared as lab mates during my first stay at the Institut Pasteur in 1967–68. That was a very productive period for me; I was able to show that unusual mutations in the *E. coli gal* operon resulted from the insertion of DNA segments that came to be known as IS elements, the first class of MGEs to be demonstrated by molecular techniques [4,28]. The scientific reason comes from Maurice's pioneering work in bacterial genomics and his ability to begin bringing intellectual order to the unappreciated topic of repetitive DNA elements in the *E. coli* genome. I learned a tremendous amount about the French mode of scientific reasoning and the Cartesian tradition from Maurice. In gratitude for those lessons and for decades of unfailing friendship, I feel privileged to contribute to this Symposium in memory of a truly creative scientist.

References

[1] A.G. Barbour, C.J. Carter, C.D. Sohaskey, Surface protein variation by expression site switching in the relapsing fever agent *Borrelia hermsii*, Infect. Immun. 68 (2000) 7114–7121.
[2] I.C. Blomfield, The regulation of pap type 1 fimbriation in *Escherichia coli*, Adv. Microb. Physiol. 45 (2001) 1–49.
[3] R.J. Britten, E.H. Davidson, Gene regulation for higher cells: A theory, Science 165 (1969) 349–357.
[4] A.I. Bukhari, J.A. Shapiro, S.L. Adhya (Eds.), DNA Insertion Elements Plasmids and Episomes, Cold Spring Harbor Laboratory, 1977.
[5] S. Dawid, S.J. Barenkamp, W.J. StGeme, Variation in expression of the *Haemophilus influenzae* HMW adhesins: A prokaryotic system reminiscent of eukaryotes, Proc. Natl. Acad. Sci. USA 96 (1999) 1077–1082.
[6] G. del Solar, R. Giraldo, M.J. Ruiz-Echevarría, M. Espinosa, R. Díaz-Orejas, Replication and control of circular bacterial plasmids, Microbiol. Mol. Biol. Rev. 62 (1998) 434–464.
[7] C.M. Fraser et al., Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*, Nature 390 (1997) 580–586.
[8] E. Gilson, J.-M. Clément, D. Brutlag, M. Hofnung, A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*, EMBO J. 3 (1984) 1417–1421.
[9] E. Gilson, W. Saurin, D. Perrin, S. Bachellier, M. Hofnung, Palindromic units are part of a new bacterial interspersed mosaic element (BIME), Nucleic Acids Res. 19 (1991) 1375–1382.
[10] M. Hofnung, J.A. Shapiro (Eds.), Research in Microbiology, special double issue on repetitive DNA sequences in microbes 150 (9–10) (1999).
[11] International Human Genome Consortium, Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.
[12] F. Jacob, J. Monod, Genetic regulatory mechanisms in the synthesis of proteins, J. Mol. Biol. 3 (1961) 318–356.
[13] A.B. Jonsson, G. Nyberg, S. Normark, Phase variation of gonococcal pili by frameshift mutation in pilC, a novel gene for pilus assembly, EMBO J. 10 (1991) 477–488.
[14] S. Lamrani, C. Ranquet, M.-J. Gama, H. Nakai, J.A. Shapiro, A. Toussaint, G. Maenhaut-Michel, Starvation-induced Mucts62-mediated coding sequence fusion: Roles for ClpXP, Lon, RpoS and Crp, Molec. Microbiol. 32 (1999) 327–343.
[15] D. Lintona, A.V. Karlysheva, B.W. Wren, Deciphering *Campylobacter jejuni* cell surface interactions from the genome sequence, Curr. Opin. Microbiol. 4 (2001) 35–40.

[16] I. Lysnyansky, Y. Ron, D. Yogev, Juxtaposition of an active promoter to vsp genes via site-specific dna inversions generates antigenic variation in *Mycoplasma bovis*, J. Bacteriol. 183 (2001) 5698–5708.

[17] G. Maenhaut-Michel, J.A. Shapiro, The roles of starvation and selective substrates in the emergence of *araB-lacZ* fusion clones, EMBO J. 13 (1994) 5229–5239.

[18] G.T. Marczynski, L. Shapiro, Bacterial chromosome origins of replication, Curr. Opin. Gen. Dev. 3 (1993) 775–782.

[19] B. McClintock, Discovery and Characterization of Transposable Elements: The Collected Papers of Barbara McClintock, Garland, New York, NY, 1987.

[20] A. Nekrutenko, W.-H. Li, Transposable elements are found in a large number of human protein coding regions, Trends Genet. 17 (2001) 619–625.

[21] B. Nicholson, D. Low, DNA methylation-dependent regulation of pef expression in *Salmonella typhimurium*, Mol. Microbiol. 35 (2000) 728–742.

[22] L.E. Orgel, F.H. Crick, Selfish DNA: The ultimate parasite, Nature 284 (1980) 604–607.

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

[24] A.R. Richardson, I. Stojiljkovic, HmbR a hemoglobin-binding outer membrane protein of *Neisseria meningitidis* undergoes phase variation, J. Bacteriol. 181 (1999) 2067–2074.

[25] S.M. Rosenberg, Evolving responsively: Adaptive mutation, Nat. Rev. Genet. 2 (2001) 504–515.

[26] J. Sarkari, N. Pandit, E.R. Moxon, M. Achtman, Variable expression of the Opc outer membrane protein in *Neisseria meningitidis* is caused by size variation of a promoter containing poly-cytidine, Mol. Microbiol. 13 (1994) 207–217.

[27] N.J. Saunders, A.C. Jeffries, J.F. Peden, D.W. Hood, H. Tettelin, R. Rappuoli, E.R. Moxon, Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58, Mol. Microbiol. 37 (2000) 207–215.

[28] J.A. Shapiro, Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*, J. Mol. Biol. 40 (1969) 93–105.

[29] J.A. Shapiro, Observations on the formation of clones containing *araB-lacZ* cistron fusions, Molec. Gen. Genet. 194 (1984) 79–90.

[30] J.A. Shapiro, Genome organization, natural genetic engineering, and adaptive mutation, Trends Genet. 13 (1997) 98–104.

[31] J.A. Shapiro, A 21st Century view of evolution, J. Biol. Phys. 28 (1997) 1–20; http://shapiro.bsd.uchicago.edu/21st_Cent_View_Evol.html.

[32] J.A. Shapiro, Genome organization and reorganization in evolution: Formatting for computation and function, Ann. NY Acad. Sci. (2002) in press; http://shapiro.bsd.uchicago.edu/contextgenome.html.

[33] A. van der Ende et al., Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the −10 and −35 regions of the promoter, J. Bacteriol. 177 (1995) 2475–2480.

[34] G. Wang, A.P. van Dam, J. Dankert, Analysis of a VMP-like sequence (vls) locus in *Borrelia garinii* and Vls homologues among four *Borrelia burgdorferi* sensu lato species, FEMS Microbiol. Lett. 199 (1995) 39–45.

[35] W.R. Zuckert, J. Meyer, Circular and linear plasmids of Lyme disease spirochetes share extensive homology: Characterization of a repeated DNA element, J. Bacteriol. 178 (1996) 2287–2298.