

## Short-Sequence DNA Repeats in Prokaryotic Genomes

ALEX VAN BELKUM,<sup>1\*</sup> STEWART SCHERER,<sup>†</sup> LOEK VAN ALPHEN,<sup>2‡</sup> AND HENRI VERBRUGH<sup>1</sup>

*Department of Medical Microbiology & Infectious Diseases, Erasmus Medical Center Rotterdam, 3015 GD Rotterdam,<sup>1</sup>  
and Department of Medical Microbiology, Academic Medical Center, 1105 AZ Amsterdam,<sup>2</sup> The Netherlands*

INTRODUCTION .....	275
MOLECULAR ANALYSIS OF SSRs .....	276
WHOLE GENOME SEQUENCES AS A SOURCE OF SSRs .....	278
STRUCTURAL FEATURES OF SSRs .....	278
DISPERSED REPEATS IN PROKARYOTIC DNA .....	279
CONTIGUOUS REPEATS IN PROKARYOTIC DNA .....	280
SPECIFIC SSRs IN DNA FROM VARIOUS BACTERIAL SPECIES .....	280
<i>Haemophilus influenzae</i> .....	280
<i>Neisseriae</i> .....	282
<i>Mycobacteria</i> .....	283
<i>Staphylococcus aureus</i> .....	283
<i>Streptococci</i> .....	284
<i>Enterococcus faecalis</i> .....	284
<i>Mycoplasmas</i> .....	284
<i>Escherichia coli</i> .....	285
Other Bacterial Species .....	285
SSRs IN PATHOGENIC PROTOZOA AND LOWER EUKARYOTES .....	285
SSRs AND MICROBIAL EVOLUTION .....	286
SSRs AND MICROBIAL PATHOGENESIS .....	287
SSRs IN MOLECULAR EPIDEMIOLOGY .....	287
CONCLUDING REMARKS AND PROSPECTS .....	288
ACKNOWLEDGMENTS .....	288
REFERENCES .....	288

### INTRODUCTION

Repetitive DNA, which occurs in large quantities in eukaryotic cells, has been increasingly identified in prokaryotes. In eukaryotic genomes, this repetitive DNA is infrequently associated with coding regions and consequently is located primarily in extragenic regions (33). Repetitive DNA consists of simple homopolymeric tracts of a single nucleotide type [poly(A), poly(G), poly(T), or poly(C)] or of large or small numbers of several multimeric classes of repeats. These multimeric repeats are built from identical units (homogeneous repeats), mixed units (heterogeneous repeats), or degenerate repeat sequence motifs (Fig. 1 shows a schematic overview). Short tandemly repeated sequences occur in several to thousands of copies dispersed through the genome of many if not all higher eukaryotes (90). These sequence elements showed hypervariability among individual persons, and genetic mapping can be used to prepare DNA fingerprints that are specific for an individual (91). Although these sequences were initially defined as mini- and microsatellite DNA consisting of short sequence repeats (SSRs) or short tandem repeats (STRs) (90), Nakamura et al. introduced a more accurate terminology (130, 131). Some of the repeats, especially those representing a single locus and showing inter-individual length variability, were designated

“variable number of tandem repeat” (VNTR) loci. VNTRs and SSRs are now well-established molecular targets for pedigree analysis (92). For the sake of clarity, we will use the abbreviation SSR as the acronym for most of the DNA repeats found in prokaryotes. This is analogous to eukaryotic nomenclature. However, some of the repeats could also be classified as genuine VNTRs in the sense that repeat number variation is associated with a single genetic locus.

Variability observed in VNTRs is thought to be caused by slipped-strand mispairing (SSM), which may occur in combination with inadequate DNA mismatch repair pathways (178). The peculiar tertiary structure of repetitive DNA allows mismatching of neighboring repeats, and, depending on the strand orientation, repeats can be inserted or deleted during DNA polymerase-mediated DNA duplication (24, 28, 70) (Fig. 2 shows a schematic presentation). Molecular studies in *Saccharomyces cerevisiae* have shown that poly(G-T) tracts are extremely unstable, with length alterations occurring at a frequency of at least  $10^{-4}$  event per cellular division (73). The instability of these moieties can be increased further by mutations in DNA mismatch repair genes. This strengthens the notion that repeat instability is associated with polymerase slippage during replication. Alternatively, variation in repeat numbers and sequence degeneracy can be explained by DNA recombination between multiple loci consisting of homologous repeat motifs. Furthermore, there is experimental evidence that the regions bordering the SSR loci are susceptible to more frequently occurring mutagenic events (136).

In eukaryotic cells, SSRs are associated with regulatory functions. In cells such as fibroblasts, repeat sequence variation has been proposed as a mechanism determining the limiting divi-

\* Corresponding author. Mailing address: Erasmus Medical Center Rotterdam, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands. Phone: 31-10-4635813. Fax: 31-10-4633875. E-mail: vanbelkum@bacl.azr.nl.

† Present address: Acacia BioSciences Inc., Richmond, CA 94806.

‡ Present address: National Institute of Public Health and the Environment, LVM RIVM, 3720 BA Bilthoven, The Netherlands.

A.		
LENGTH OF REPEAT UNIT (BASEPAIRS)	REPEAT	ANNOTATION
1	5' -AAAAAAAAAAAAAAAA-3'	5' - (A) <sub>15</sub> -3'
2	5' -ATATATATATATATAT-3'	5' - (AT) <sub>8</sub> -3'
3	5' -GCCGCCGCCGCCGCCGCCGCC-3'	5' - (GCC) <sub>8</sub> -3'
4	5' -GATCGATCGATCGATCGATCGATC-3'	5' - (GATC) <sub>6</sub> -3'
5	5' -GCTCCGCTCCGCTCCGCTCCGCTCC-3'	5' - (GCTCC) <sub>5</sub> -3'
6	5' -AAAATTAAAATTAAAATTAAAATT-3'	5' - (AAAATT) <sub>4</sub> -3'

B.	
5' -	----AAC AAC AAC GTCTC GTCTC AT AT AT AT AT AT ----3'

C.	
SEQUENCE IDENTIFICATION	
A	5' - GATCGATCAA GATCGATCAA GATCGATCAA -3'
B	5' - --A----- -----A-- -----G----- -3'
C	5' - ---G----- -----T----- -3'
D	5' - ----- CTA----- -----TT -3'

FIG. 1. Schematic survey of SSRs. (A) Examples of homogeneous simple sequence motifs consisting of repeat units varying from 1 (homopolymeric tract) to 6 nucleotides in length. (B) Example of a composite, heterogeneous repeat built from three 3-nucleotide units, two 5-nucleotide units, and seven 2-nucleotide motifs. (C) Comparative analysis of four different repeats built from three 10-nucleotide units showing degeneracy among units. Identity of the nucleotide sequences B through D with the consensus given in sequence A is indicated by dashes.

sion potential (132). Assuming that a single copy of a repeat unit can be deleted with surgical precision per cell cycle, certain SSRs may serve as precise counters of the number of cell doublings. Most human SSRs, however, have an as yet unknown function. For a limited number of SSRs, especially those in the trinucleotide (triplet) repeat class, involvement in human genetic disease has been demonstrated (20). Consequently, repeat variation can identify a predisposition to inheritable diseases in some instances. Moreover, it was demonstrated that SSM at repeats is associated with the development of cancer (101, 154). The (CA)<sub>n</sub> SSRs in DNA from approximately 30% of colorectal tumors differed significantly in length from those encountered in healthy cells from the same individual (184). SSR variation was correlated with patient survival and could be used as a marker for disease severity (1). Neurodegenerative disorders have been linked to repeat expansion as well. In studies of Huntington's disease, elegant experiments with transgenic mice harboring various lengths of a (CAG)<sub>n</sub> SSR in the 5' end of the Huntington's disease gene indicated that CAG variability is causally related to the development of a progressive neurological phenotype (111). Genetic heterogeneity based on repeat variation was even found in vertebrate organisms that seem to lack effective mechanisms for genetic recombination, emphasizing the relative importance of SSM (188). Changes in the number of repeats in a given genetic locus apparently are an important source of DNA variability in clonal lines of species such as *Poecilia formosa* or *Rivulus marmoratus*. The examples mentioned above suggest that many more links between SSRs and disease may be identified in the near future. A very recent observation may even allow the theoretical identification of hitherto unknown disease-associated repeats. It was shown that microsatellites which are

not associated with so-called retrotransposons (160) or which are located in genes (129) may be the disease-associated ones.

Since the abundance of repetitive DNA in eukaryotic genomes is largely unexplained (see references 129 and 209 for reviews), the identification and functional analysis of repetitive DNA in prokaryotes is a timely subject of experimental studies. The presence of prokaryotic SSRs is well documented, and some SSRs show extensive length polymorphisms. SSM has been documented as an important prerequisite for bacterial phase variation and adaptation. In this review, we summarize which SSRs occur in prokaryotes and how genes are affected by variation in these moieties. We first describe the techniques used to characterize the SSRs in prokaryotes and how repeat variability can be monitored. SSRs identified by sequence analysis of distinct genes or gene clusters are described, and the SSRs that are detectable after whole-genome analysis are presented. Basic mechanisms involved in variation of the number of repeat motifs are discussed. Although repetitive DNA structures are named differently in various publications, they will be identified as SSRs throughout this review. Since in microbial genetics most is known about repeats that are relatively short in unit length and contiguous in nature, this class of repeats will be discussed most extensively. The focus will be on the medically relevant microorganisms, since these have been studied in most detail.

#### MOLECULAR ANALYSIS OF SSRs

Repetitive DNA has characteristic physical features due to its specific nucleotide composition. The detection of the first eukaryotic repetitive DNA moieties was the immediate consequence of their aberrant density. When subjected to density

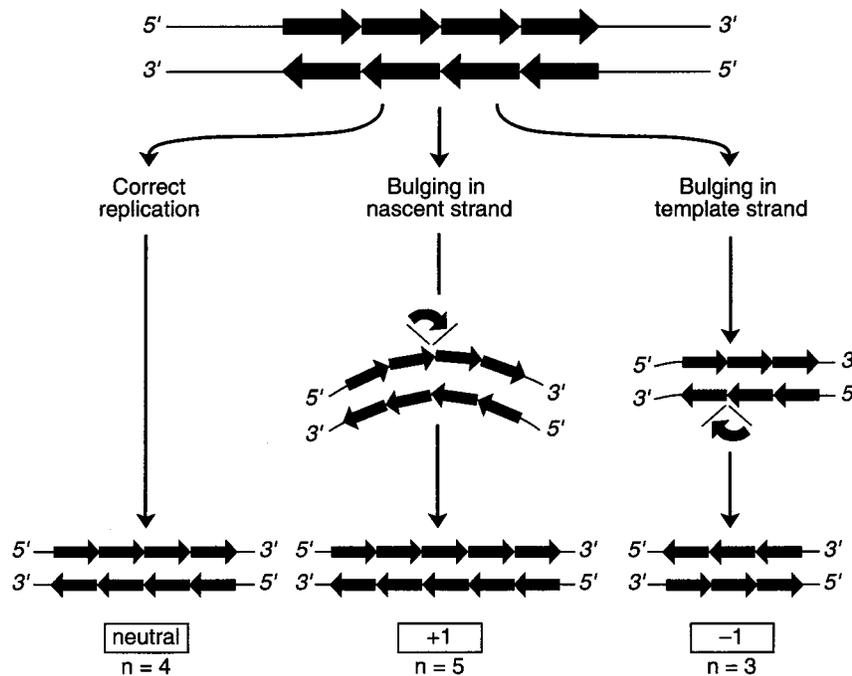


FIG. 2. Schematic representation of the mechanism of SSM during replication, which results in shortening or lengthening of SSRs. Individual repeat units are identified by arrows; bulging is the presence of non-base-pair base residues interrupting a regular 2-strand DNA helix. Bulging in the nascent strand leads to a larger number of repeat units; bulging in the template strand results in a smaller number of units. During replication, bulges can occur in both strands, and the effect of insertion or deletion can be neutralized by occurrence of the adverse event. The number of repeat units can decrease or increase by multiple repeats once multiple bulging in one strand has occurred.

gradient centrifugation, repetitive DNA lagged behind the bulk of DNA and presented as satellite fractions due to differences in thermodynamic stability and reassociation kinetics (14) (Fig. 3A). Whereas the variability in repetitive DNA domains was initially detected by relatively cumbersome Southern hybridization techniques with DNA probes recognizing the repeat consensus motif (Fig. 3B), the emergence of PCR technology enabled a more straightforward DNA amplification-mediated approach (89). In this method, PCR primers bordering the SSR region are constructed and polymorphism in repeat unit number is documented by simple electrophoretic techniques once DNA amplification has been performed (Fig. 3C). Regions bordering the repeats are generally sufficiently well-conserved targets for PCR-mediated amplification. Consequently, repeat degeneracy can be analyzed by direct sequencing (Fig. 1). Moreover, border sequence conservation is sometimes even observed among different species, allowing a broad-spectrum analysis of the nature of the species and sub-specific genetic polymorphisms (164).

In several instances, repeat locus multiplication has been demonstrated. This results in multiple bands in the amplification assays. Genuine locus multiplication was proven by Southern hybridization studies and DNA sequencing. The presence of SSRs sharing completely identical neighboring sequences can be explained by full gene duplications. Based on sequence comparisons of SSR-neighboring regions, it has been suggested that recombination between repeat loci does occur once multiple sites are available in a single genome (196). Whether intergenomic recombination takes place is unknown. Polymerase slippage generally results in DNA ladder-type patterns (70, 82), giving rise to differences in small numbers of a single repeat unit. Multiple bands in PCR amplifications can also be caused by erroneous amplification processes. These can be

largely circumvented by using thermolabile DNA polymerases, which are more reliable copiers of reiterated DNA motifs (82). Since the SSR typing assays with thermolabile polymerases are more complicated to perform, the use of such polymerases is discouraged. Multiplex PCRs allow the simultaneous analysis of more than a single repeat locus (22, 196), which has the advantage that multiple SSRs can be analyzed in a single assay. PCR tests based on amplification with five different primer pairs have been successfully performed (197a). The propensity toward expansion or reduction of the number of repeat units at a given locus through slipped-strand mispairing events is emphasized by epidemiological studies. Even genetically homologous strains may show differences in SSR size, even though these may be small compared to those determined for more distantly related strains (104, 199).

Cloning of repetitive DNA for further characterization can be done by combining different molecular procedures. Due to aberrant density of repetitive DNA, simple procedures for purification of the major eukaryotic repeat species have been developed (Fig. 3A). Moreover, since repetitive DNA present in high copy numbers reassociates differently from other DNA moieties, DNA containing repeats can be enriched for by subtractive hybridization. Usually, repeats can also be identified by specific DNA oligonucleotides built from a limited number of the key sequence elements. This allows straightforward identification, and the cloning process can be monitored with the help of this type of specific hybridization. Random amplification of polymorphic DNA (213, 215) allows screening for genetic variation without prior knowledge of DNA sequence information. Consequently, the size-variable SSRs are relatively frequently highlighted by this specific PCR application, which generates DNA fragments highly characteristic of the template DNA used. When these DNA elements are hybrid-

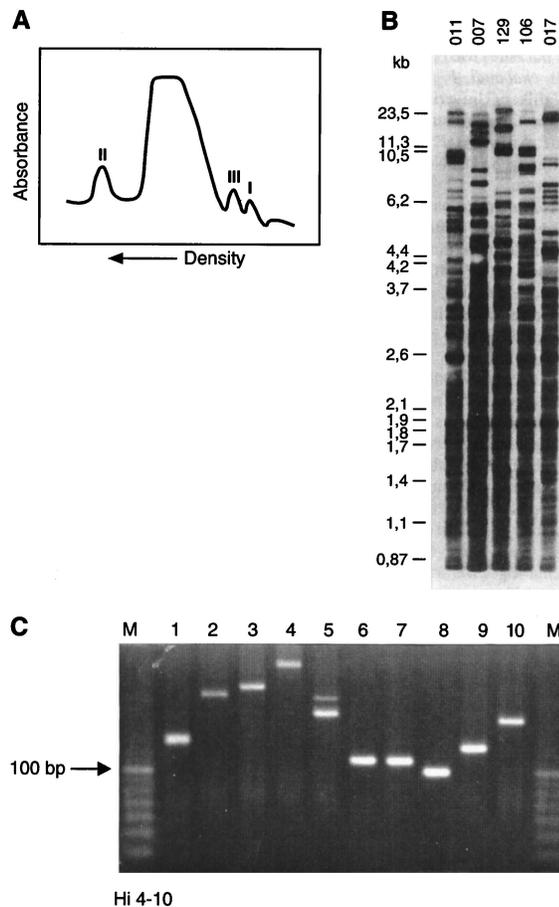


FIG. 3. Molecular identification of SSR-type DNA. (A) Microdensitometer tracing of human leukocyte native DNA centrifuged to equilibrium in density gradients with an analytical ultracentrifuge. The satellite peaks representing repetitive DNA fractions displaying aberrant densities are indicated (I to III). (B) DNA fingerprints of human individuals generated by probing with repetitive DNA. DNA from five individuals (identified by numerals above the lanes) was digested with a restriction enzyme, the fragments were separated by electrophoresis, and after blotting, the resulting Southern blot was probed with a synthetic oligonucleotide SSR consisting of 10 units of a TTAGG motif. The autoradiograph shows that this SSR is widely dispersed throughout the human genome and clearly depicts the hypervariability in the observed banding patterns. (C) PCR amplification of SSR regions. A specific SSR was identified in the genome of *H. influenzae*, and primers bordering the repetitive motif were synthesized. When DNA from bacterial strains 1 to 10 was used as template, various amplicons were generated, most of them showing clear differences in length (related to the number of repeat units present). Lanes M contain 10-bp molecular size markers.

ized to repetitive DNA probes, potential SSRs can be identified (25, 148). Subsequent and straightforward cloning and sequencing of these amplicons allows molecular verification of the nature of the SSR. Another procedure, directed amplification of minisatellite-region DNA, capitalizes on the use of SSR core sequences as primers during PCR (71). In the present high-resolution DNA-sequencing era, SSRs can be identified in sequence databases with the help of straightforward software (see below). In summary, repetitive DNA has specific biophysical characteristics and can be identified by means of hybridization or DNA amplification. Searching for and direct cloning of SSRs is currently facilitated by the availability of increasing amounts of primary DNA sequence information in general.

## WHOLE GENOME SEQUENCES AS A SOURCE OF SSRs

In recent years, the genomes of several eubacterial species have been sequenced completely (48, 55, 81). The availability of this type of data for *Haemophilus influenzae*, *Mycoplasma genitalium*, and *M. pneumoniae* has enabled detailed analysis of these genomic molecules for the presence of repetitive DNA with SSR-like characteristics. SSRs highlighted on previous occasions (94) are identified within the sequence databases, but numerous additional SSR-like domains have been identified (see Table 2). Variability of repeats consisting of 2 to 6 nucleotide units is observed in various genome regions outside the lipopolysaccharide (LPS)- and fimbrium-related genes of several other bacteria.

In the genome of the methanogenic archaeon *Methanococcus jannaschii* (17), only dinucleotide SSR candidates were detected (data not shown). Of these 35 loci, 2 were located on one of the two extrachromosomal plasmids. All the SSRs were shorter than 7 dinucleotide units. This indicates the nonuniversal character of the SSR-type domains. Apparently, microorganisms exist that essentially do not require SSRs. Whether this is a reflection of a lack of the need of this particular archaeon to adapt to variations in the environment is unclear, but it should be noted that *M. jannaschii* lives in a spectacular but relatively constant niche. The recent publication of the whole genome sequence for *Escherichia coli* (8) revealed that in this gram-negative bacterial species, SSR-type regions harboring repeat units of less than 8 nucleotides are also very rare. Whether this is simply because the laboratory strain K-12 has lost many pathogenicity factors due to progressive laboratory propagation needs to be established in future studies.

The single eukaryotic organism for which the entire genome sequence is now available is *S. cerevisiae* (64). In 12,068 kbp of DNA, 5,885 potential protein-encoding genes were identified apart from the numerous genes for rRNAs, small nuclear RNA molecules, and tRNAs. Repetitive DNA was encountered in large quantities. Overall, 1,320 kbp of repeats larger than 2 kbp covered 11% of the entire genome. Searches for small repeats identified numerous examples of potential SSRs (Table 1). The search was specifically aimed at short repeats, 2 to 8 nucleotides in unit length, with the use of recently developed software that is available through the World Wide Web (196). The number of nucleotides involved was 22.6 kbp (0.2% of the overall genome). It would be interesting to study whether these loci are functionally involved in virulence gene regulation similar to that described below for prokaryotes. Studies like these are important in the light of the potential pathogenicity of the organism, for instance in recurrent vaginitis (171, 220). Since it is certain that further whole genome sequences will become publicly available in the future, additional SSRs will be identified in huge numbers. This will turn out to be highly relevant to the increase of our knowledge on microbial diversity, taxonomy and pathogenesis.

## STRUCTURAL FEATURES OF SSRs

Grossly, bacterial SSR-type DNA can be divided into four main categories. First, dispersed repeat motifs that generally do not occur in tandem have been identified. Although these repeats occur throughout genomes of a multitude of microorganisms, they are sometimes organized in tandem as well. A second class is formed by the homopolymeric tracts. Multimers of one of the four nucleotides are peculiar sequence elements that are frequently encountered in the genome of *S. cerevisiae*, for instance. These homogeneous stretches can amount to as

TABLE 1. Perfect small-unit SSRs in the *S. cerevisiae* genome<sup>a</sup>

No. of units	No. of SSRs of unit length:							
	1	2	3	4	5	6	7	8
3	— <sup>b</sup>	—	—	—	17	105	8	7
4	—	—	—	—	20	37	2	1
5	—	341	—	6	7	9	0	1
6	—	140	104	5	1	5	0	0
7	—	69	49	0	1	1	0	0
8	—	52	43	2	0	1	0	0
9	—	35	26	0	0	0	0	0
10	756	35	20	0	0	0	0	0
11	483	37	5	0	0	0	0	0
12	345	15	6	1	0	0	0	0
13	245	20	4	0	0	0	0	0
14	132	5	3	0	0	0	0	0
15	91	8	2	0	0	0	0	0
16	67	6	1	0	0	0	0	0
17	51	7	1	0	0	0	0	0
18	28	4	0	0	0	0	0	0
19	32	3	0	0	0	0	0	0
20	24	3	1	0	0	0	0	0

<sup>a</sup> This table summarizes the occurrence of SSRs in the *S. cerevisiae* chromosomes. Numbers indicate the number of SSRs being present whereby the unit length defines the size of the repeat motif (1 to 8 nucleotides, horizontal scheme) and the unit number defines the number of repeat units present. No imperfections in the repeat unit were allowed for; repeat regions that are mentioned separately in the table were sometimes physically coupled (neighboring stretches of DNA). The maximum unit number was arbitrarily chosen to be 20. In the mononucleotide SSR class, the largest individual repeat contained 42 nucleotides. Incidentally, additional 2- and 3-nucleotide SSRs were encountered (results not shown).

<sup>b</sup> —, not searched for because of expectations of an extremely large number; the minimum detectable number as used in the computer search program equals those targets not searched for.

much as 42 nucleotides (Table 1). Third, short-motif SSRs are identified. With repeat units differing from 2 to 6 bases, it is this class of repeats that is most liable to unit number variation at a given locus. Particularly, when these short-motif repeats are located within genes and are not 3 or 6 nucleotides long, they can drastically affect the coding potential of a given transcript. Fourth, repeats harboring more than 8 nucleotides per unit, form a separate category. It is surprising that in the *S. cerevisiae* genome, SSRs with 7 or 8 bases per unit are sparse. However, the software scores only perfect repeats, neglecting SSRs built from degenerate motifs. In addition, the longer the repeat unit, the greater the chance that point mutations will be introduced.

For the SSR species that are involved in human disease, it has been observed that five of the six currently documented repeat motifs allow the generation of hairpin-type structures (122). The presence of these hairpin structures during replication, transcription, or translation may interfere with faithful copying of the SSRs or other sequence-based protein-DNA/RNA interactions. The base composition of the repeats is also important. The higher the percentage of adenosine and thymidine, the more unstable the DNA helix will be (125). Theoretically, lower melting profiles may increase the possibilities of SSM, but well-documented studies are not yet available. The effect of a lower fidelity of DNA repair for the frequency of SSM was mentioned previously (52, 178). Not only the repeat motifs themselves but also neighboring structures have an impact on the frequency with which SSM takes place (155). Depending on the presence or absence of stable hairpin structures in the vicinity of the repeats, deletion of repeat units may be enhanced more than 50-fold. SSRs within regions of repetitive bacterial DNA have been studied in model systems such as

infection of *E. coli* by recombinant M13 phages (103). When artificial poly(C-A) tracts were cloned into the phage genome, easily detectable expansion or deletion within these tracts was seen during bacteriophage multiplication.

Several classes of microbial SSRs exist. Each class displays individual structural features, and sometimes even the SSR surrounding DNA regions is involved in maintaining the physical integrity of the repeat regions. The SSRs that will be discussed in this review are either the short ones (repeat units of 1 to 6 nucleotides) or repeats that are much longer (repeat units of >15 nucleotides). Repeats with intermediately sized unit lengths are only rarely encountered. It is interesting that the shorter unit repeats, in particular, are involved in regulatory processes that are affected by SSM. Among the longer repeats, a larger degree of sequence heterogeneity is observed. This heterogeneity is thought to be indicative of more frequent recombination. Analyses of the precise function of the repeat locus are often missing. It is regularly assumed that these repeats encode protein sequences spanning membranes or cell walls. Therefore, they play a physical more than a regulatory role. These longer repeats are candidate regions for determining phylogenetic relatedness between species or strains. Both repeat number and sequence heterogeneity among the repeats are valuable parameters, but a precise relationship between phylogeny and repeat variability has not yet been defined.

#### DISPERSED REPEATS IN PROKARYOTIC DNA

The number of reports on dispersed prokaryotic repetitive sequence motifs is even larger than for contiguous repeats. Interspersed repetitive DNA has been identified in numerous bacterial species. This topic has been reviewed by Lupski and Weinstock (106). Most of these elements are shorter than 200 bp, noncoding, intercritically located, and distributed evenly in genomic molecules. Characteristic prokaryotic repeats such as the enterobacterial repetitive intergenic consensus (ERIC) sequences and the repetitive extragenic palindromic (REP) sequence motif have been found in numerous different enterobacterial species as well (207). The REP sequence element was discovered in 1982 (77), whereas the ERIC motifs, initially named intergenic repeat units, were described in detail in 1990 and 1991 (61, 85, 167). The possible functions of these repeats are still enigmatic, although recent reports describe the association between a unique class of REP elements and integration host factor sequences. This has clear implications for overall genome structure and DNA topology (9, 135). Although the dispersed repeats were initially deemed to be rather species specific (63), analogous sequences have also been detected in other microbial species including the cyanobacteria (5) and soil bacteria such as *Rhizobium meliloti* (35). Whether these analogs really represent canonical ERIC/REP motifs still awaits verification by DNA sequencing.

In the early 1990s, another dispersed-repeat motif was identified in *Streptococcus pneumoniae* (113). This so-called BOX repeat appeared to be unrelated to the repeat types mentioned above. The BOX repeat forms stable secondary structures and appears to be transcribed in some instances. Most of the BOX sequences were encountered in close proximity to genes, suggesting their potential role as a regulatory element controlling coordinate virulence- or competence-related gene expression. Since the discovery of the pneumococcal BOX repeat, these DNA elements have been used for molecular typing purposes as well (42, 100, 194). All of these prokaryotic moieties represent repetitive DNA that is not organized in tandem repeats but is scattered throughout the entire genome of microorganisms. An excellent review written by Versalovic et al. (206)

elegantly discusses both the theoretical and technical aspects of repetitive sequence-based PCR in which REP, ERIC, or BOX motifs are used to delineate strain relatedness. It is demonstrated that due to the specificity of the repeat PCRs, crude cell lysates can be used as amplification templates. Moreover, even material directly derived from infections can be used immediately without a bacterial cultivation step. Theoretically, this enables bacterial detection and typing in a single PCR test. Amplicons derived from repeat PCRs show a very high degree of variability, but some of the fragments have been successfully used as species-specific DNA probes (60).

ERIC, REP, and BOX motifs constitute different classes of prokaryotic dispersed repeat elements. Although little is known about their function, their ubiquity suggests that they are involved in important aspects of microbial life.

### CONTIGUOUS REPEATS IN PROKARYOTIC DNA

As an obvious example of contiguous repeats in genomic DNA, highly repetitive regions containing tRNA genes were detected in bacterial genomes (63). However, Field and Wills recently performed the first systematic, computerized search for the presence of longer arrays of short repeat units in the genomes of simple organisms (44). A total of 78 putative sites were discovered in the GenBank database, but not all of these were derived from eubacterial DNA sequences. Several "eubacterial loci" were detected in the genome of *H. influenzae*. A (CTA)<sub>9</sub> domain was discovered in the genome of *Neisseria meningitidis*, whereas (CTA)<sub>11</sub> and (CTT)<sub>21</sub> were discovered in sequences derived from *Mycoplasma genitalium* and *Mycobacterium leprae*, respectively. It was suggested that these bacterial loci may be as prone to genetic variability as those encountered in eukaryotes. Experimental proof is lacking, however. Based on the observations made by Field and Wills (44), additional searches for bacterial SSRs are required. The contents of the GenBank sequence depository still show exponential growth, and novel SSR loci may present useful markers for identification and tracking of bacterial strains. Moreover, since examples indicating the involvement of repeat variability in environmental adaptation have been presented previously, additional repeat loci might be of help in the identification of novel virulence factors (126). In this respect, it is interesting that probing the genome of different strains of *Helicobacter pylori* with short, microsatellite-type DNA probes was highly informative, with clear and epidemiologically informative DNA polymorphisms being documented (112).

In several instances, DNA repeats were detected in plasmids as well. An important example of this type of SSRs is the so-called iteron (see reference 45 for a review). Iterons are present in multiple copies at the plasmid origin of replication and specifically bind copies of the initiator protein RepA. A replication initiation complex is obtained upon saturation of all iteron moieties with RepA (12). Both plasmid and genomic DNAs of prokaryotes contain SSRs with various functions.

### SPECIFIC SSRs IN DNA FROM VARIOUS BACTERIAL SPECIES

Various specific examples of SSRs involved in adaptive behavior of different bacterial species will be discussed below. Tables 2 and 3 summarize species, SSRs, and associated functions.

#### *Haemophilus influenzae*

The genome of *H. influenzae* has been extensively studied with respect to SSR composition, including the functional as-

TABLE 2. Occurrence of contiguous perfect small-unit SSRs in the genome of *M. genitalium* and *M. pneumoniae*

Unit length	Repeat position <sup>a</sup>	Unit sequence	No. of units	Sequence homology
<i>M. genitalium</i>				
3	86051–86066	CTT	5	None
3	169475–169523	TAG	16	None
3	224534–224555	TAG	7	Attachment protein MgPa
3	227130–227163	TAG	11	114-kDa MgPa
3	349735–349750	TCT	5	None
3	351452–351482	TAG	10	None
3	425824–425857	TGT	11	Unidentified protein
3	429308–429335	TAG	9	None
3	429967–423015	CTT	16	None
5	212939–212954	CAAAA	3	None
6	384465–384489	TATTAC	4	None
<i>M. pneumoniae</i>				
2	60190–60210	AG/AC <sup>b</sup>	5/5	Unidentified protein
2	65147–65169	CT	11	None
3	23651–23669	TAG	6	Adhesin P1
6	34932–34950	AACCCC	3	Adhesin P1
6	775707–775725	AACCCC	3	Adhesin P1 precursor

<sup>a</sup> Repeat positions are based on nucleotide sequence information described in reference 55. Nucleotide repeats consisting of monomers [poly(A), poly(C), poly(G), or poly(T) tracts] were not searched for; 2-, 4-, 7-, and 8-unit repeats were not encountered. Repeats were identified by a computer program written by Stewart Scherer, University of California Los Angeles.

<sup>b</sup> The two sequence motifs were found to be adjacent; both motifs occurred five times and were exactly neighboring.

pects of repeat variability. The first genetic locus involved, *lic1*, was described in 1989 (212). With the use of monoclonal antibodies, different patterns of LPS expression were observed and the molecular switch leading to this phenotypic variability appeared to be dependent on the translational capacity of the *lic1* mRNAs. Within the gene, variable numbers of the tetranucleotide unit 5'-CAAT-3' were detected. Variation in the overall number of units moves one of three potential ATG codons in or out of the protein synthesis reading frame. This directly affects protein synthesis and the primary amino acid sequence. Subsequently, additional regions in the *H. influenzae* genome containing CAAT repeats were identified (211). Similar switching signals were identified in the *lic2* and *lic3* genes, which are also involved in LPS biosynthesis (152, 153, 211). For the *lic2* gene, SSR polymorphisms were detected by PCR amplification of the region involved (78) and homologous repeat sequences were identified in *Neisseria* spp. and *Moraxella catarrhalis* (142). Also, in *Haemophilus somnus*, variation in the CAAT copy number for expression of a gene encoding a protein involved in the synthesis of lipooligosaccharides appeared essential for variation (86). Inactivation of the *mrp* gene, which is involved in LPS biosynthesis as well, by transposon mutagenesis results in changes in the number of repeats present in *lic2* (79). This suggests an intricate, close relationship between the various forms of regulation of expression of the different enzymes involved in the biosynthesis of LPS.

A CCAA repeat was found in a gene encoding a 120-kDa heme-repressible hemoglobin-binding outer membrane pro-

TABLE 3. Survey of eubacterial SSR-containing genes with a known function

Species	Repeat motif	Gene	Level of regulation		Gene function	Reference(s)
			Transcription	Translation/protein		
<i>H. influenzae</i>	CAAT	<i>lic1-lic3</i>	–	+	LPS biosynthesis <sup>a</sup>	83, 211, 212
	GCAA	<i>yadA</i>	–	+	Adhesin ( <i>Yersinia</i> homolog)	83
	GACA	<i>lgtC</i>	–	+	Glycosyltransferase	83
	TTGG	ND <sup>f</sup>	–	+	Iron binding proteins	83
	AGTC	ND	–	+	Restriction modification, methyltransferase	83
	TTTA	ND	–	+	Unknown <i>Bacillus</i> homolog	83
	TA	<i>hifA/B</i>	+	–	Synthesis of fimbriae	199, 200
<i>N. meningitidis</i>	G	<i>lsi2</i>	+	–	LPS biosynthesis	18
	CTCTT	<i>opa</i>	–	+	Opacity surface proteins	119
	A	<i>opa</i>	+	–	Opacity surface proteins	119
	G	<i>porA</i>	+	–	OMP	198
<i>S. aureus</i> <sup>b</sup>	93 bp	<i>fnb</i>	–	+	Fibronectin binding protein	139
	561 bp	<i>cna</i>	–	+	Collagen adhesin	140
	81 bp	<i>coa</i>	–	+	Coagulase	65, 165
	GAAGAXXXAAXAAXCCTXGXAAA	<i>spa</i>	–	+	Protein A	27
	GAXTCXGAXTCXGAXAGX	<i>clf</i>	–	+	Clumping factor, fibrinogen receptor	116, 117
<i>Streptococcus</i> spp.	60 bp	<i>pspA</i>	–	+	Pneumococcal surface protein <sup>f</sup>	219
	69 bp	<i>emm</i>	–	+	Phagocytosis resistance M protein	6
	246 bp	$\alpha$ C	–	+	$\alpha$ C protein (antibody-mediated killing)	67
<i>E. faecalis</i>	TAGTARR	<i>rep1</i> and <i>rep2</i>	+	–	Iteron, regulates plasmid replication and transfer	72
<i>M. hyorhinus</i> <sup>d</sup>	36 and 39 bp	<i>vlp</i>	–	+	Variant membrane lipoprotein	218
	A	<i>vlp</i>	+	–	Variant membrane lipoprotein	218
<i>M. bovis</i>	24 bp	<i>vspA</i>	–	+	Membrane surface lipoprotein	107
<i>M. fermentans</i>	A	<i>P78</i>	–	+	Lipoprotein in ABC transporter complex	183
<i>U. urealyticum</i>	GGTAAAGAACAACCAGCA	<i>MB</i>	–	+	Serology-specific MB antigen	222
<i>B. anthracis</i>	CAATATCAACAA	<i>vvrA</i>	–	+	Microfilarial sheath protein homolog	87
<i>L. monocytogenes</i>	66 bp	<i>prfA</i>	–	+	Leucine-rich internalin like protein	37
<i>E. coli</i>	A and C <sup>c</sup>	<i>lac</i>	+	–	$\beta$ -Galactosidase	52, 156
<i>A. marginale</i>	87 bp	<i>mvp1<math>\alpha</math></i>	–	+	Major surface protein	2

<sup>a</sup> Homologues for these sequences have been described for *Moraxella catharralis* and the human pathogenic *Neisseria* species.

<sup>b</sup> The genes described for this species all produce MSCRAMMs.

<sup>c</sup> Homologous structures were identified in diverse streptococci and *Clostridium difficile*.

<sup>d</sup> See Table 2 for the analysis of *M. genitalium* and *M. pneumoniae* on the basis of the whole genome sequences as available.

<sup>e</sup> Homopolymeric tracts.

<sup>f</sup> ND, not defined.

tein of *H. influenzae* (93). Such a repeat was not detected in homologous genes from other microorganisms. It was suggested that the regulation of this particular gene resembled that of the *lic* loci. Another example of an *H. influenzae* SSR locus is found between the genes *hifA* and *hifB*, encoding fimbrial subunit proteins (199). Reversible phase variation is due to changes in the number of TA repeats, which space the –35 and –10 boxes of the dual promoter controlling *hifA* and *hifB*. This has a clear impact on transcriptional efficacy (Fig. 4).

Interestingly, detailed genetic analysis of the entire fimbrial gene cluster also revealed the presence of a number of tandemly repeated REP units with an as yet unknown function (200).

Upon determination of the full genome sequence for *H. influenzae* Rd (48), algorithms capable of identifying short repeat motifs were developed (96, 196) and used to screen for SSRs. Various potential di- to hexanucleotide SSRs were identified. Repeats with units of 7 or 8 nucleotides were not en-

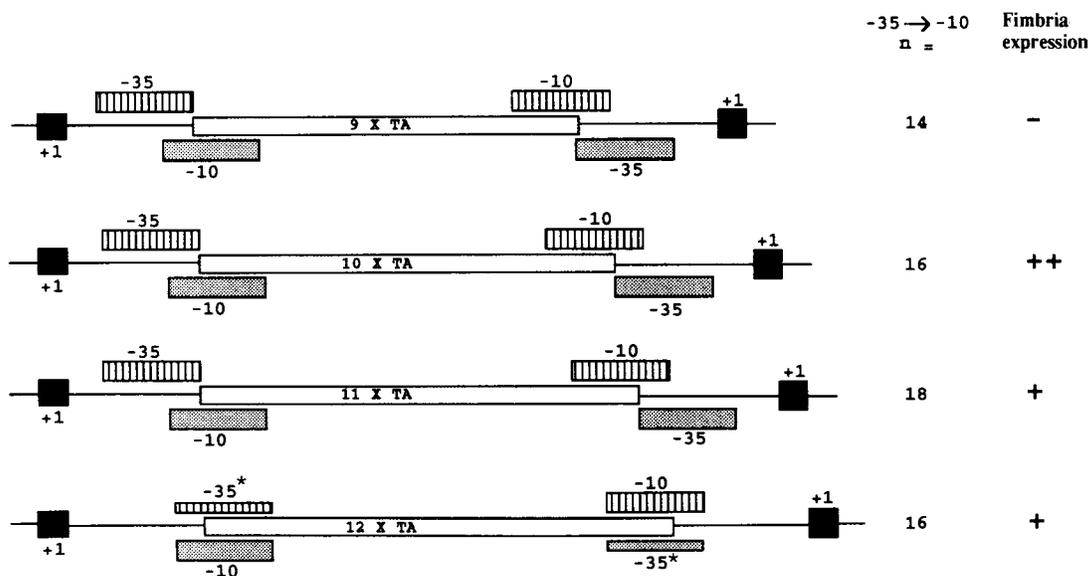


FIG. 4. Hypothetical model for the mechanism of fimbrial phase variation in *H. influenzae*. Transcription of the two divergently oriented genes *hifA* and *hifB* is controlled by a variable sequence of reiterated TA units (long open box) in their combined promoter region. In nine units in this SSR, the putative  $-10$  and  $-35$  promoter sequences for both *hifA* (hatched boxes) and *hifB* (short stippled boxes) are separated by 14 bases, which does not allow transcription of either gene. With 10 or 11 TA units, the  $-10$  and  $-35$  motifs are separated by 16 or 18 bases, respectively, allowing transcription of both genes, resulting in the expression of fimbriae. A spacing of 16 bases results in the highest level of fimbriation. With 12 TA units, *hifA* and *hifB* transcription can take place with alternative  $-35$  promoter sequences (narrow boxes), which are separated by 16 bases from the corresponding  $-10$  sequences. Solid boxes represent the  $+1$  transcriptional start point. Reprinted from reference 199 with permission of the publisher.

countered. For all of the 3- to 6-nucleotide repeats in the *H. influenzae* chromosome, PCR tests capable of detecting allelic polymorphisms were designed (196). Of 18 potential SSRs, 14 were indeed highly polymorphic when different strains were screened. For some of the poly-tetranucleotide motifs, for instance, the number of repeats encountered per locus varied from 4 to over 40. The length of the SSRs was a stable genetic marker for separate colonies derived from a single clinical specimen or strains passaged for several weeks on chocolate agar plates. When several strains isolated from different patients during an outbreak of lung disease caused by *H. influenzae* were analyzed (191), increased but limited variation was encountered in the 4-nucleotide unit SSR sites (197). It appeared that the 3-, 5-, or 6-nucleotide SSRs were more stable in nature and consequently more suited for epidemiological studies. One of the two 5-nucleotide SSRs, however, proved to be hypervariable when strains isolated during another epidemic were investigated (196). This may be indicative of the selection imposed by individual patients. Closely related populations already show some heterogeneity, which becomes more evident among more distantly related strains. The different repeats appear to differ with respect to intrinsic (in)stability as well. The 4-nucleotide SSRs may be used for high-frequency adaptation, whereas the others could be more stable and useful for epidemiological mapping studies. All 4-nucleotide SSRs appeared to be involved in bacterial virulence, and molecular knockout of one of these genes resulted in attenuated virulence (83). Several old and new candidate virulence genes could be identified by straightforward localization of the repetitive domains (83, 196). These genes varied from the well-known LPS biosynthesis genes and the adhesin- and glycosyl-transferase-encoding genes to several iron binding protein genes.

### Neisseriae

The genome of the two pathogenic *Neisseria* species *N. gonorrhoeae* and *N. meningitidis* was found to contain a 26-meric directly repeated DNA element at high copy number (32). This sequence motif was named Ngrep, and specific targeting of this element by PCR is used for the detection of genetic heterogeneity among strains of both *Neisseria* species (204). The ability to perform these genotyping tests on crude cellular extracts without the need to purify template DNA means that this technique will be able to be used with primary clinical materials (146). The epidemiological validity of the Ngrep PCR still needs to be evaluated, as has been done before in comparative studies for some of the other neisserial typing strategies (143, 181).

Comparison of the *N. gonorrhoeae* genome with that of the meningococcus showed that the major genetic differences between the species are confined to three loci (36, 186). One of the three loci contains capsular genes, which seems obvious: this finding appears to be in agreement with the species specificity that is generally encountered when surface-exposed gene products are analyzed. At the surface of *Neisseria* cells, pili, opacity proteins, and LPS are the main exposed structures that are also subject to phase variation. Nonpiliated variants can be observed (59, 69, 166).

LPS phase variation in *N. gonorrhoeae* functions as an adaptation mechanism enabling the bacteria to escape from the immune system and translocate across various physical barriers (201). The variation occurs spontaneously within strains, and various forms of phase variation have been documented. The nature and type of LPS variability of *N. gonorrhoeae* and *H. influenzae* are similar (19, 208). Whether gonococcal LPS variation occurs through the same mechanisms as has been described for *H. influenzae* seems quite likely but has not yet been documented in detail. A recent study by Burch et al. (18)

emphasized the importance of translational or transcriptional frameshifting in a polyguanine SSR in the *lsi2* gene of *N. gonorrhoeae*. The variation allows expression of multiple or single lipooligosaccharide structures.

Variation in the opacity (*Opa*) proteins occurs by *recA*-independent rearrangements in the so-called coding repeat sequence. In this region, shifting of the translational reading frame occurs because of changes in repetitive DNA numbers (127, 176, 177). These changes occur independently in any of the *opa* genes, which accounts for the production of several different *Opa* proteins simultaneously. The relationship between the superficial *Opa* protein composition of a bacterial isolate with invasiveness into the human epithelium has been demonstrated experimentally (110). A variable number of CTCCT pentameric repeat units in the *opa* leader peptide moves the reading of the gene in or out of frame. The sequence is peculiar since a triple-helix conformation is likely to occur and the pentameric repeat region appears to be hypersensitive to single-strand-specific nuclease. The number of repeats varies continuously at low frequency in the *in vivo* situation. Once environmental selection predisposes to survival of one of the "minority SSR types," this type will "translate" its selective advantage into overgrowth of the existing population (110, 177).

As was the case for *H. influenzae*, spacing of the -35 and -10 promoter components in *N. meningitidis* through breathing in a repeat locus appears to be basic to some aspects of gene regulation (198). The strength of the *porA* gene promoter was modulated at three different expression levels, a feature that was linked to the number of guanosines present in a contiguous stretch in between the -10 and -35 regions. The *porA* gene encodes the class I outer membrane protein (OMP) and is supposed to be an interesting vaccine target. However, the amount of this class I OMP exposed on the cell surface varies widely as a consequence of promoter modulation. In some cases, OMP nonproducers were encountered. This implies that the OMP is not essential for survival, which may limit its potential as a vaccine candidate.

**Mycobacteria**

*M. tuberculosis* is very well known for the presence of various repetitive DNA elements dispersed throughout the entire genome. These elements have been used for strain tracking in a very efficient manner (202). Most of the elements are insertion sequences, but directly repeated GC-rich elements (36 bp per repetitive unit) have also been identified and used for molecular identification (76, 94). In addition a polymorphic GC-rich repetitive sequence called PGRS (157) and a major polymorphic tandem repeat sequence (75) were discovered. These repeats display a large degree of sequence heterogeneity. More simple SSR-like sequence motifs have also been described for *M. tuberculosis* through hybridization with a probe consisting of five GTG units (214). Although detection of restriction fragment length polymorphisms does not directly highlight locus-specific repeat variability, indirect evidence for VNTR-type repeat number variation is suggested. A (GTG)<sub>5</sub> probe identified highly polymorphic elements, some of which seemed to appear in multiple copies. Stability during serial cultivation was demonstrated. No additional perfectly repetitive DNA sequence motifs were identified by a computerized scan of the full and partial genome sequences as available for both *M. tuberculosis* and *M. leprae*, respectively (163a). Efforts are now being aimed at determining the whole-genome sequence of both *M. tuberculosis* and *M. leprae*. SSRs suited for epidemiological studies and definition of virulence loci or contingency

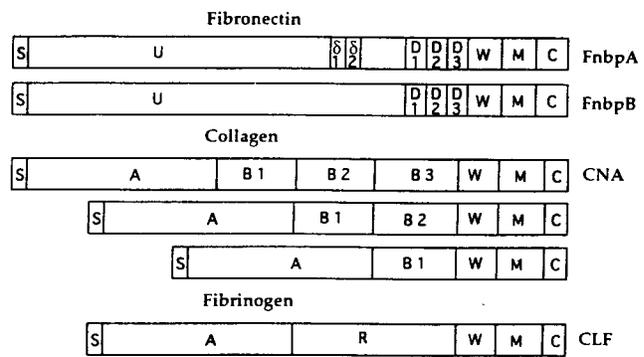


FIG. 5. Schematic models of MSCRAMMs from *S. aureus*. The MSCRAMMs shown are the ones with affinity for fibronectin (FnbpA and FnbpB), collagen (CNA), and fibrinogen (CLF, clumping factor). S, signal sequence; U or A, unique nonrepetitive sequence; δ, upstream repeat sequences; D, B, or R: repeated domains; W, cell wall-spanning domain; M, hydrophobic membrane-spanning domain; C, positively charged carboxy terminus. Reprinted from reference 139 with permission of the authors and the publisher.

loci may thus be identified for these two major bacterial pathogens in the near future (145).

**Staphylococcus aureus**

In *Staphylococcus aureus*, the genes encoding many of the membrane-bound proteins contain contiguous repetitive DNA. These so-called microbial surface components recognizing adhesive matrix molecules (or MSCRAMMs) have similar building-block structures. From the amino to the carboxyl terminus, a signal sequence, a unique protein-dependent domain, several repetitive units, a cell wall-spanning domain, a hydrophobic membrane-spanning domain, and a positively charged terminus can be recognized in all staphylococcal MSCRAMMs (Fig. 5) (for a general description see reference 139). The molecules recognize macromolecular ligands in the extracellular matrix of the host cells. In *S. aureus*, MSCRAMMs specific for fibronectin (49), collagen (74), laminin (117), vitronectin (138), thrombospondin (140), elastin (141), bone sialoprotein (158), and fibrinogen (203) have been identified at the molecular level. In the collagen-binding MSCRAMM gene *cna*, variability of the repeat numbers (from 1 to 3 units) has been documented (182). The clumping-factor *clfA* gene, which encodes a 92-kDa protein encodes a 308-amino-acid domain consisting of 154 Glu-Ser combinations within this protein. This domain is not involved in ligand binding, because a small peptide consisting of Glu-Ser repeats alone does not inhibit the association between MSCRAMM and ligand (53). Apparently, this domain functions as a stalk which enables adequate surface exposure of the active domain. The dipeptides are encoded by an 18-nucleotide DNA repeat which shows some degeneracy (GAY TCN GAY TCN GAY AGY). This sequence heterogeneity increases once the borders of the SSR region are approached (116). Since different degeneracies are observed in the 5' and 3' regions of the SSR, the hypothesis that this particular region evolved by intergenic recombination is clearly substantiated (168). The building-block nature of the MSCRAMMS, together with repeat variability in the stalk domain, once more identifies bacterial candidate virulence genes. Stalk variation may result in differential exposure of the active moieties of the proteins involved, giving rise to differences in immune sensitivity or virulence.

The coagulase protein is a major phenotypic species determinant in *S. aureus*. Within the encoding gene, repeats of 81

nucleotides can be observed (65). These repeats are clearly polymorphic in both number and sequence, as can be visualized by combined PCR and restriction fragment length polymorphism analysis. The assay can be used for epidemiological research (4, 165). Again, it has been demonstrated that the repetitive domains are not required for biological activity. Clotting caused by prothrombin binding occurs independently of the repeats (97).

PCR assays aiming at the amplification of the repeat locus in the staphylococcal protein A gene provide another example of how repetitive DNA can be used to track DNA variability and to delineate bacterial epidemiological behavior (4, 56, 57). However, some debate on the clinical and epidemiological validity of this type of variability mapping has been initiated (195). The fact that various protein A gene polymorphisms are encountered among strains with similar overall genotypes indicates that the speed with which the repeats evolve does not accurately reflect the speed of overall genome evolution. Detailed DNA sequencing studies could in the end reveal the conditions under which repeat variation occurs, allowing the consequences of the variability to be monitored in more detail.

### Streptococci

*S. pneumoniae* is a major human pathogen. One of its major surface-exposed determinants is the pneumococcal surface protein PspA. PspA is built from four distinct structural domains: an N-terminal  $\alpha$ -helical coiled-coil region, a proline-rich region, an apparently defective membrane anchor at the C terminus, and, between the proline-rich region and anchor domain, a repetitive moiety comprising 10 highly conserved 20-amino-acid repeats (219). Homology searches revealed that similar repeats were encountered in species such as *S. mutans*, *S. downei*, and *Clostridium difficile*. Repeat evolution at this locus in these bacterial species was considered to be related to the loss or gain of mechanisms for protein attachment to the bacterial cell.

Group A streptococci harbor so-called M proteins, which confer the ability to resist phagocytosis by polymorphonuclear leukocytes. Numerous serogroups differing in M-protein antigenicity have been described, and the diversity of the M proteins has been demonstrated frequently. Expression of different M-protein types has been coupled to pathogenicity (6), and genetic variation in the encoding *emm* genes was used for molecular identification of streptococci. Repetitive DNA built from imperfect units was found within the *emm* gene.

The alpha C protein-encoding gene, containing repetitive DNA, was identified among group B streptococci. Apparently, antibody-mediated killing of group B streptococci is affected by the number of repeats in the alpha C protein (108). These repeats are long, with 82 amino acids encoded per unit. The unit contains an epitope that is protective against antibodies. Spontaneous repeat deletions occur at a frequency of  $6 \times 10^{-4}$  per generation. Antigenic variation is influenced by repeat variability. Under immunological pressure, the variants are selected and appear frequently. An inverse relationship between the number of repeats and the immunogenicity of the protein was recently defined (66). With increasing numbers of repeats, the titers of specific antibodies elicited appeared to decrease significantly. These observations have important implications for the development of vaccines aimed at the alpha C protein moiety (67): the immune response elicited by the vaccine could be evaded by the pathogen by simply altering the repeat composition in the alpha C protein gene. The studies of streptococci emphasize the relevance of repeat variability as a means for immune evasion.

### *Enterococcus faecalis*

A classical example of the physiological relevance of repetitive DNA was described for *Streptococcus faecalis* (currently *Enterococcus faecalis*) 21 years ago (217). Within plasmid pAM $\alpha$ 1, the tetracycline resistance determinant was amplified by culturing strains for prolonged periods in the presence of sublethal concentrations of tetracycline. This led to the presence of multiple, tandemly repeated copies of a 2.65-MDa DNA element on a single plasmid molecule. Cells harboring the amplified resistance determinant have increased tetracycline resistance. *E. faecalis* contains another very interesting form of repeat variation (15). In the process of plasmid transfer from a donor to a recipient bacterial cell, a small peptide sex pheromone plays an important role (26, 39, 216). The control of the physiological response to the pheromone is subject to a phase variation process involving changes in a repetitive DNA region (210). This so-called iteron region consists of 12- and 13-bp direct repeats, separated by a 78-bp spacer region, thus essentially presenting another example of an SSR region. Phase variation, read from colony morphology differences, involves structural changes in this SSR-containing (TAGTARRR) unit (72). This particular class of degenerate repeats seems to be involved in a complicated process that governs not only plasmid transfer but also vegetative plasmid replication (27, 124). It has been suggested that specific proteins that could bind either to the iterons or to genes involved in the pheromone response play a significant regulatory role. When there is an extended number of units in the SSR, the protein factor preferentially binds to these moieties, thereby enabling transcription of conjugation control genes. The presence of iterons at or near the replication origin of other plasmid species suggests broad-spectrum involvement of these elements in replication and copy number control of plasmids in general (98).

### Mycoplasmas

When the whole genome sequence of *M. genitalium* (55) is scanned for the occurrence of SSRs with unit lengths ranging from 2 to 8 nucleotides, only 3-nucleotide examples are found (Table 2). In addition, single examples of short 5- and 6-nucleotide repeats are found (3 and 4 units in length, respectively). It is stimulating to see that several hits are within the MgPa operon, for which multiple copies have been shown to exist per genome (144). The encoded MgPa adhesin molecule is a highly immunogenic protein, which appears to be essential for efficient adhesion to the host epithelium. *M. pneumoniae* M129 was also sequenced, and, similar to *M. genitalium*, limited numbers of short SSRs were detected by computer searching (Table 2). The observation that two of five SSRs are found within adhesin genes supports the notion that SSRs also occur in virulence genes in mycoplasmas. By comparison with *H. influenzae*, another distribution of SSR type motifs is encountered, indicating species dependence of the most prevalent repeat types.

In the size-variable membrane surface lipoprotein genes (*vlpA*, *vlpB*, and *vlpC*) of *M. hyorhinis*, two different repeat polymorphisms are involved in antigen variation (218). Loss or gain of repeats within the coding sequence results in protein size variation. In addition, gene expression is switched on or off through repeat variation in the promoter region. A tract of contiguous adenosine residues varies the spacing between the -35 and -10 boxes and as such affects transcription efficiency. The two levels of expression giving rise to phenotypic adaptability are the result of a unique combination of basic genetic phenomena. In the infectious agent *M. fermentans*, a ho-

mopolymeric adenosine tract was observed as an intragenic variable element (183). By alterations in the number of adenosines, strains were either capable of or prevented from producing a substrate-binding lipoprotein. The gene for this protein is part of a four-gene operon encoding an ATP binding cassette-type transport complex. In *M. bovis*, several intragenic repeats were identified in membrane surface lipoproteins (Vsp) as well (107). Phenotypic switching of these antigens can be explained by the presence of DNA polymorphisms in the region encoding the VspA gene. These findings suggest that recombination plays a major role in antigenic variation. Although precise mechanisms have not yet been elucidated, it is quite likely that SSM contributes to increased recombination frequencies. At the protein level, *M. synoviae* was shown to express a phase-variable hemagglutinin (133). Taking the relatively large number of functional SSRs already reported for mycoplasmas into account, it is expected that expression of this protein is also under the control of an SSR. Finally, it was shown that in *Ureaplasma urealyticum*, a *Mycoplasma*-like microorganism, small repeat units within the so called MB antigen encode serovar specificity and are associated with antigen size variation (222). A major part of the MB gene is repetitive: 18-bp repeats can form up to two-thirds of the entire coding region. *Mycoplasma* genomes are rich in repeats, and these affect virulence and antigenic properties.

#### *Escherichia coli*

It has been well documented that *E. coli* populations respond adequately to conditions of nonlethal selection (for a detailed review, see reference 51). Various selectable phenotypes arose as a consequence of limited or specialized growth medium supply, and probable mutational events were identified. Detailed studies focusing on the molecular basis of these phenomena have been published. Two elegant exponents of this type of study, appearing in the same issue of *Science* in 1994 (52, 156), focused on reversible frameshift mutations in *E. coli*. Small variations in a homopolymeric tract were identified upon nutritional deprivation. SSM was proposed as the basic mechanism for this type of adaptive behavior. Interestingly, both studies showed that strains with mutations in DNA mismatch repair activity were more prone to variation than were wild-type strains.

Only very recently, the entire genome sequence for *E. coli* K-12 was published (8). Computerized searches for SSRs were quite disappointing, however. Apart from the previously identified homopolymeric tracts, only a few dinucleotide SSRs with a maximum of 5 repeat units per locus were encountered. Analysis of SSR polymorphisms in relation to bacterial virulence does not seem to be a promising investment of time and effort. The molecular basis for the observed adaptive mutations producing resistance to ciprofloxacin in *E. coli* (149) may not be related to SSR-type modification as described for other species in the above sections.

#### Other Bacterial Species

Many other examples of repetitive DNA elements, as for instance those detected in the *Coxiella burnetii* genome (84) or IS elements crossing the species borders in enterococci (185), have been described. These moieties frequently represent dispersed bacterial insertion elements which may be even longer than 1,000 bp. A genuine SSR in *Bacillus anthracis* is built from 12 nucleotide units and was identified by the previously described random amplification of polymorphic DNA identification method for mildly variable regions (3). The repeat number clearly varied among strains. The SSR appeared to be embed-

ded in a gene showing significant homology to a microfilarial sheath protein. This suggests the potential surface exposure of the *B. anthracis* homolog and could indicate that the bacterial gene product might be antigenic in nature and that the repeats may be involved in the generation of antigenic variation. Five different *B. anthracis* alleles were distributed in separate geographic regions (87).

The rickettsia *Anaplasma marginale* causes hemoparasitic disease in cattle, is tick borne, and is economically highly relevant. A subunit of the major surface protein MSP1 exhibits extreme size polymorphism. This can be attributed again to an SSR region within the encoding gene (2). The surface protein carries a neutralization-sensitive epitope that could be traced back to the region encoded by the repeats. It is, however, still not known how a surface-expressed, neutralization-sensitive epitope remains constant despite conceivably constant immune pressure.

In *Listeria monocytogenes*, expression of virulence factors is regulated by the virulence regulator protein PrfA (21). Several PrfA-associated proteins have been characterized (105), and molecular analysis of one of these elements revealed homology to internalin proteins. The homology was partly in the leucine-rich repeat region (37). The function of the relatively long SSR within this secreted internalin-related protein, IrpA, is unclear. Gene deletion showed that IrpA is not involved in invasion or intracellular survival. Since deletion mutants showed attenuated virulence, it is currently suggested that the protein is involved in some way in the dissemination of *L. monocytogenes* infection. Similar leucine-rich repeats have been detected in other species as well (99, 109), and it is interesting that the peptide repeats are thought to be crucial in protein-protein interactions.

In general, the examples mentioned in this section emphasize the importance of SSR elements in many aspects of adaptive bacterial behavior. SSRs enable bacteria to respond to diverse environmental factors, and many of them are clearly related to bacterial pathogenesis. Some of the SSRs seem to play an essential role in controlling surface exposition of active protein domains and antigenic variation.

#### SSRs IN PATHOGENIC PROTOZOA AND LOWER EUKARYOTES

The genome of the lower, single-celled eukaryotes was shown to contain repetitive DNA moieties by hybridization (120, 189, 193) or by PCR-mediated detection of local DNA length variability (reference 150 and references therein). In various eukaryotic microorganisms, the presence of variable repeats was demonstrated with a single set of PCR primers. This particular assay is based on characterization of a DNA element derived from *Trichomonas vaginalis* but can be used for genetic typing of species as diverse as the intestinal parasite *Giardia lamblia* and the amoeba *Paramecium tetraurelia*. The complexity of the banding patterns that were generated indicates the dispersed nature of the repetitive element involved. No cross-reactivity of the PCR primers with human DNA was observed.

Reports on contiguous repeats in other protozoa are relatively sparse. Obvious direct repeats include the terminal elements of the chromosomes called telomeres (50), and different classes of subtelomeric repeats, consisting of large units mainly, have also been described (38). Telomeric repeats have been identified in many species, including *Leishmania donovani* (41). Papers on the presence of direct repeats in the circumsporozoite protein gene of several species of the malaria parasite *Plasmodium* date back to the early 1980s (reference

137 and references therein). The most interesting phenotypic feature of these repeats is their presentation at the polypeptide level. It is thought that the peptide repeat units assume a tertiary structure that shields the remainder of the protein from the host defense system. Malaria parasites contain numerous other genes coding for variant surface antigens, some of which are even presented on the surface of the infected erythrocyte. At present, there are no experimental clues whether SSRs play a role in regulation of immune system evasion in other antigens exposed during the life cycle of the parasite (11).

Fungi form a taxonomic group that contains many medically relevant species. *Aspergillus* spp., *Candida* spp., and *Cryptococcus neoformans* are regularly encountered in neutropenic patients and can cause devastating disease in these individuals.

For the aspergilli, repetitive DNA has been used for genetic typing. A number of DNA fragments containing nonribosomal repetitive sequences were successfully cloned from genomic DNA of *Aspergillus fumigatus* (62). By using some of these clones as the basis for the development of DNA probes, genetic polymorphisms were detected. Eight of these probes gave rise to over 15 different hybridization signals per lane in the Southern blot. Patterns were stable over many laboratory generations, varied among unrelated strains, and could be analyzed by computer imaging and subsequent phylogenetic analysis. Most of the probes were strictly species specific, a phenomenon encountered once more when repetitive DNA from *A. flavus* was cloned (115). Little is known about the sequence, expression, or function of these moieties in fungal genomes. Since SSRs identified in bacteria are often involved in gene regulation, the same may be true for fungi. The ubiquity of SSRs in fungi is nicely demonstrated by the *S. cerevisiae* data obtained by analysis of the whole genome sequence and shown in Table 1. Tracking down the variability and putative functioning of SSRs within the genes of this microorganism may be of importance to the brewing industry. Since the environment in which the yeast cells reside changes in composition during fermentation, it could be speculated that changes in SSR may be involved in the continuous adaptation of *S. cerevisiae* strains to the environmental requirements.

The first species-specific DNA repeat probes suitable for the detection of genetic polymorphisms among strains of the pathogenic yeast *Candida albicans* were described in detail in the late 1980s (162). The so-called 27A element represented a family of mobile structures undergoing high-frequency reorganization. Besides the 27A moieties and the ribosomal repeats, other repeats were identified (163). Among these were the telomeric simple repeats and more enigmatic repeats such as Ca3 and Ca7 (173–175). The telomeric Ca7 repeat consists of multiple and contiguous copies of a 23-bp sequence which evolves at high speed (159). When all available *Candida* DNA sequences were surveyed for the occurrence of potential SSRs, nine such loci were identified within open reading frames (43). Three of these SSRs contained perfectly repeated units only. By PCR-mediated amplification of the repeats, all potential SSRs were shown to be polymorphic among strains. A set of physically linked SSRs present in the *ERK1* protein kinase gene of *C. albicans* provided an interesting example of a highly polymorphic complex repeat region. Most of the *Candida* repeats seemed to encode poly(G) tracts, which suggested involvement in transcription regulation (58). By analogy, it was recently shown that expansion of the poly(G)-encoding SSR in the human gene involved in spinocerebellar ataxia (*SCA1*) leads to alterations in the nuclear localization of the gene product (169). The function of the *C. albicans* poly(G) tracts is the subject of further studies (44a).

In *Cryptococcus neoformans*, repetitive DNA has also been used for strain delineation. Both hybridization-based (23, 180) and PCR-mediated (121) assays with simple sequence primers and probes have been designed and used for epidemiological studies. Both *C. albicans* and *C. neoformans* can be highly pathogenic microorganisms, and many of their virulence determinants have been characterized (123, 134, 172). Whether repetitive DNA in fungi fulfills a role similar to that observed in several prokaryotic species (see below) is unknown. However, by analogy to the functional implications of SSR modification as observed in humans and bacteria, SSRs may also be involved in the modulation of gene expression and phenotype in fungi.

A general strategy for molecular typing of fungi that is based on the presence of SSRs is simple and straightforward: probing DNA restriction digests with simple repeat motifs such as (CT)<sub>10</sub>, (GTG)<sub>5</sub>, or (GACA)<sub>6</sub> provides the fungal epidemiologist with DNA fingerprinting data that are extremely useful for molecular tracking of an as yet unknown spectrum of species. Both contiguous and dispersed repeats show a high degree of interstrain variability suited for the unequivocal identification of genetic relatedness. As described for prokaryotes in the previous sections, SSRs are essential genomic elements in lower eukaryotes. SSRs can be used for strain identification but seem to be involved in several aspects of gene expression and protein functioning as well.

#### SSRs AND MICROBIAL EVOLUTION

Evolutionary processes have been studied in detail for several bacterial species. A recent report postulates that bacterial evolution may not be a continuous process but more of a succession of temporally spaced major events (34, 40). These events cause a nongradual sequence of adaptations to a given environment. So-called pathogenicity islands are an obvious example of genetically unstable elements that are present or absent in a given prokaryotic cell in near-random fashion (68, 118). Novel genes are also acquired by transduction mediated by bacteriophages, by the introduction of plasmids or transposons, by conjugative DNA transfer between cells, or by natural DNA-mediated transformation. Microbial variability due to switching of gene expression modulated by activation of silent genes as a consequence of genetic rearrangements (10) is another example of microbial evolution where major changes are induced in very short periods. The speed of species evolution is determined in part by environmental influences (147). Variation in SSRs involves changes directly at the level of the DNA molecule itself. These changes may alter the genetic repertoire of a given (micro)organism and may result in evolution of the species. Apparently, as has been determined previously, bacterial species take successful advantage of this mode of high-speed evolution to gain selective profits (126, 142, 199, 200).

Different regions of the prokaryotic genome evolve in different manners, depending on sequence composition and reliable recognition by DNA-modifying enzymes. DNA damage is an important source of mutation in addition to the occurrence of spontaneous errors during replication. The occurrence of mutations is not equally likely for any position in a given genome. During the replication of DNA from enteric bacteria, an excess of C-to-T changes was detected in the coding strand (54). Due to their specific structure, SSRs may be highly liable to this type of effect.

Assessing genetic relatedness is complicated. Association between genetic loci reveals linkage or clonality, whereas a lack of association is an indicator of extensive "genome reshuffling"

between strains of a given species, also known as panmicticism. Apparently, bacterial population structures vary with the reproductive strategy of the species (170). Consequently, different bacterial species respond differently to environmental selection. As well as mutations, genetic exchange also plays a significant role in bacterial evolution (29). The two most important consequences of genetic exchange are the creation of new allele combinations in populations of a single species and transfer of adaptations across microbial taxa. Whether this type of "natural transformation" is sufficiently frequent and adequate to drive population diversification is the subject of continuous debate. It has been postulated that adaptive evolution in genes encoding bacterial factors that interact with an "unpredictive environment" may be beneficial to a bacterial cell and, in the end, to bacterial populations (126, 142). Variation in these regions, which became known as contingency loci or hypermutable sites, enables genetic flexibility while maintaining overall genome integrity (125). Variation in the number of repeats in a given SSR is generally supposed to happen by chance. Whether targeted variation in SSR loci exists is currently uncertain but is considered to be very likely by several authors (51, 125). However, no gene products that specifically drive the variation in the number of repeat units have been identified. Moreover, obtaining experimental support for the occurrence of forced mutation prior to the selection event may prove to be very difficult, if not impossible (102). Variability in SSRs may provide selective advantages to the individual cell, which has a significant impact on microbial population biology.

#### SSRs AND MICROBIAL PATHOGENESIS

Bacterial colonization is the first step in a process that may eventually lead to infectious disease (7). This process is initially steered by general physicochemical interactions. In later stages, the stability of the interaction between microbe and host will be determined largely by interactions between specific adhesins and receptors. Bacteria have developed many specific surface structures that permit them to bind firmly to epithelial cells. Pili are an important class of structurally distinct elements in this respect (16, 221). After initial colonization of epithelial cells, invasion of host cells may take place, which in turn may initiate a cascade of events, eventually leading to systemic disease (46, 47). Disease-specific clinical phenomena such as inflammation, tissue damage, and immune response have mutual effects on both the host and the invading pathogen. SSR variability, as described for numerous microorganisms in this review, has clear implications for virulence (126, 142, 151, 212). The contingency genes containing SSRs show high mutation rates, allowing the bacterium to act swiftly upon deleterious environmental conditions (126). Model studies with plasmids containing cloned satellite DNA revealed that variation in the size of the repetitive domains could be detected even among bacteria subcultured from a single colony (187). This variability appeared to be *recA* independent and was probably due to unequal intramolecular recombination in replicating DNA molecules (analogous to sister chromatid exchange in eukaryotes) or, more likely, to SSM. Locating new repeats by GenBank screening and determining whether these genes are located within or in the vicinity of genes provides a general approach to the identification of new virulence-related genetic loci.

In eukaryotes, stretches of certain types of repetitive DNA have been located in the 5'- or 3'-flanking regions of genes. These repeats may be involved in nucleosome organization, recombination, or regulation of gene expression or gene prod-

uct activity (187). Studies based on computer analysis of microbial whole genome sequences revealed overrepresentation of several oligonucleotide motifs. These motifs were reminiscent of the presence of repetitive moieties consisting of uptake signal sequences, intergenic dyad sequences, and multiple tetranucleotide iterations (95, 96). Again, most if not all of these latter elements were located within potential virulence genes (83). Site-directed gene inactivation confirmed this supposition by showing a clear reduction in virulence in the baby-rat model. The existence of similar repeats has been shown for several *Neisseria* species, *Haemophilus parainfluenzae*, and *Moraxella catarrhalis* (142). The apparent variability of 3-, 5-, and 6-nucleotide SSRs for strains of *H. influenzae* has also been demonstrated (196). More detailed analysis of the 5-nucleotide repeat variability, which has clear effects on the continuity of reading frames, is an interesting option. It was demonstrated that one of the *H. influenzae* 5-nucleotide SSRs is located in a gene encoding one of the enzymes involved in restriction modification (unpublished data). The importance of this finding was emphasized by the discovery of another large 5-nucleotide SSR in a similar gene detected in *Pasteurella haemolytica* (80). The repeat number variation in SSRs seems to be intimately related to modulation of the expression of virulence factors.

#### SSRs IN MOLECULAR EPIDEMIOLOGY

Numerous molecular techniques have been used for bacterial DNA typing, and several recent reviews are available (88, 114, 190, 205). The availability of these techniques has accelerated research in bacterial population genetics over the past years and has allowed the identification of expansion of certain bacterial clones, some of which attained global spread (128). In recent years, DNA amplification techniques have been used with increasing frequency as the technique of choice for genotyping bacterial isolates or strains. PCR can also be used to simply monitor SSR variation. The design of locus-specific PCR primers allows high-speed development of multiple assays suitable for the study of molecular evolution, which has immediate implications for the determination of epidemiological relationships as well. SSR sizing tests present a valuable addition to the spectrum of typing procedures that are currently available in medical microbiology (Fig. 6). The SSR sizes are expressed as simple figures, a result that may be more reproducible than the data generated by other typing procedures (31, 192). A major drawback of this technique for studying epidemiology might be that regions under environmental pressure behave as hypervariable targets. This may restrict the general application of these regions for molecular typing of bacterial strains. Moreover, SSRs involved in gene (in)activation present the host organism with two possibilities only: a gene is either switched on or switched off. In a gene containing a 4-nucleotide SSR, the gene is not only switched on when 3 units are present but also if 6, 9, 12, 15, etc., are present. Different SSR genotypes behave identically at the phenotype level, and large differences in repeat number may still be functionally neutral. Successful use of PCR-mediated SSR amplification followed by amplicon size determination to analyze the spread of microbial pathogens has been reported for *H. influenzae* (197) and *Candida albicans* (13). These studies used automated DNA sequencing for the determination of allelic polymorphisms and for precise size determination. Genetic markers that were stable upon cultivation of strains and showed adequate resolution were obtained. SSRs lend themselves to the development of novel assays suited for strain identification and definition of strain relatedness.

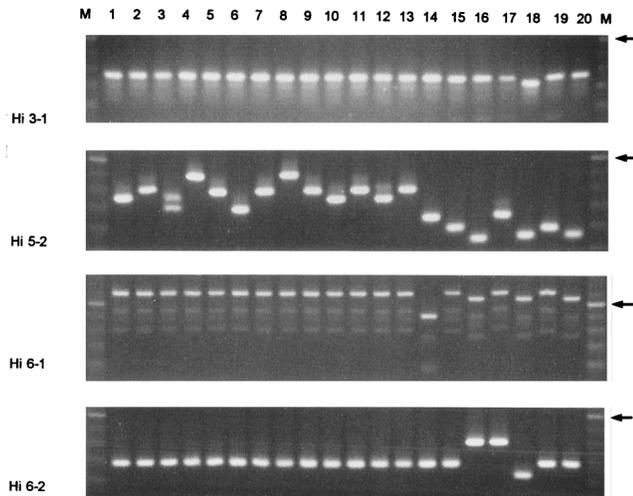


FIG. 6. Epidemiology of *H. influenzae* infections. During an outbreak, strains 1 to 13 were isolated from different patients and compared to seven nonrelated clinical isolates (strains 14 to 20) on the basis of SSR polymorphisms. Four different SSRs were analyzed for the occurrence of length variability, and assays 3-1, 6-1, and 6-2 correctly identified the epidemic isolates as identical. The controls clearly differed in some instances. Interestingly, assay 5-2 also revealed major polymorphisms among the epidemic isolates, possibly identifying a “contingency locus” that is tailoring colonization or infection of a range of individual human hosts. Lanes M contain molecular size markers; the arrows on the right identify a 100-bp DNA fragment. Reprinted from reference 196 with permission of the American Society for Microbiology.

### CONCLUDING REMARKS AND PROSPECTS

Infection processes require that the bacteria adapt to several different host environments. Initial colonization, crossing epithelial and endothelial barriers, survival in circulation, and translocation across, for instance, the blood-brain barrier are all processes that require specific virulence traits (see, e.g., reference 152). The possibility of varying pathogenicity factors to meet these requirements could possibly be achieved through SSR modulation, as has been described for a multitude of different genes. Although it is generally assumed that variation occurs randomly, unknown regulatory mechanisms may still exist. As a result, variation through SSM or recombination processes allows regulatory or adaptive functions to be specifically activated or repressed.

Large-scale nucleotide sequence data are beginning to be available for several of the medically relevant microorganisms. This allows the search for potentially novel SSRs. Although hypervariability of these regions has not yet been documented, the fact that even the small genome of *Mycoplasma genitalium* appears to contain SSR-type DNA suggests that candidate virulence genes or targets for molecular typing purposes could be identified for many other prokaryotes once larger areas of their genomes are identified in detail. Since the pharmaceutical industry, together with granting agencies and biotechnology firms, has invested over a billion dollars in numerous bacterial genome-sequencing projects, additional full-chromosome sequences will become available in the immediate future (30). Currently, over 40 different genome-sequencing projects covering both eubacteria (from *Actinobacillus actinomycetemcomitans* to *Vibrio cholerae*) and archaea (from *Archaeoglobus fulgidus* to *Thermoplasma acidophilum*) are in progress (179). SSR analysis may in the future be applied to microorganisms that are either fastidious or noncultivable, delivering diagnostic information as well as data on the genotype of the strains involved. The efficiency of this type of analysis can be greatly

enhanced once multiplex PCR assays are operational. This would enable the clinical microbiologist to obtain diagnostic, epidemiological, and pathogenicity-related data for various microorganisms by SSR typing of DNA prepared from a single clinical specimen.

It is interesting that only a limited number of studies have been undertaken to precisely establish the tertiary structure-function relation in SSR-type DNA. One study describes the influence of neighboring hairpins on the SSM frequency (155); however, there are no other studies correlating DNA structure with the frequency of SSM. From this perspective, it would be worthwhile analyzing the “efficiency of SSM” against the genetic background and the SSR sequence motifs. This could provide information about the minimum length requirements and the precise influence of base composition. Furthermore, studies like these could provide further insight into the observed interspecies differences in the occurrence and variation of SSRs and their contribution to evolutionary fitness. Novel functional aspects are surfacing continuously. It was observed that in different species of the fruit fly *Drosophila*, length variation in a repeat within the clock gene *period* correlated with the ability of the flies to maintain a correct circadian rhythm but was not influenced by variation in the environmental temperature (161). The number of discoveries correlating repeat flexibility with highly complex phenotypes will undoubtedly grow in the coming years.

Monitoring SSRs enables the study of molecular processes involved in microbial pathogenicity. Regulation of virulence-associated genes, a critical factor in infectious disease progression, can be determined in detail with the help of animal model studies involving site-directed mutants of the pathogen. The really short SSRs seem to be involved mainly in regulation of the expression of genes, whereas the somewhat longer repeat moieties seem to have other functions. The latter structures appear to be mainly involved in implementing size variation in cell wall- or membrane-associated proteins, which may cause enhanced or diminished exposure of active protein domains on bacterial surfaces. In addition, SSR evolution may be a useful feature for monitoring short-term variability in the genome of a large number of medically important microorganisms. The analysis of SSR composition in clinical isolates may in the end be a useful prognosticator of a patient’s risk of developing severe infections.

### ACKNOWLEDGMENTS

Peter W. M. Hermans (Laboratory of Paediatrics, Erasmus University, Rotterdam, The Netherlands) and Wil H. F. Goessens (Department of Medical Microbiology & Infectious Diseases, Erasmus Medical Center Rotterdam) are thanked for critically reviewing the manuscript and for stimulating discussions. Willem van Leeuwen is acknowledged for performing the computer analyses described in the tables.

### REFERENCES

- Aaltonen, L. A., P. Peltomaki, F. S. Leach, P. Sistonen, L. Pylkkanen, J. Mecklin, H. Jarvinen, S. M. Powell, J. Jen, S. R. Hamilton, G. M. Petersen, K. W. Kinzler, B. Vogelstein, and A. de la Chapelle. 1993. Clues to the pathogenesis of familial colorectal cancer. *Science* **260**:812–816.
- Allred, D. R., T. C. McGuire, G. H. Palmer, S. R. Leib, T. M. Harkins, T. F. McElwain, and A. F. Barbet. 1990. Molecular basis for surface antigen size polymorphisms and conservation of a neutralization sensitive epitope in *Anaplasma marginale*. *Proc. Natl. Acad. Sci. USA* **87**:3220–3224.
- Andersen, G. L., J. M. Simchock, and K. H. Wilson. 1996. Identification of a region of variability among *Bacillus anthracis* strains and related species. *J. Bacteriol.* **178**:377–384.
- Arbeit, R. D. 1996. Laboratory procedures for the epidemiologic analysis of staphylococci, p. 253–286. In G. L. Archer and K. B. Grassley (ed.), *The biology of staphylococci*. Churchill Livingstone, New York, N.Y.

5. Asayama, M., M. Kabasawa, I. Takahashi, T. Aida, and M. Shirai. 1996. Highly repetitive sequences and characteristics of genomic DNA in unicellular cyanobacterial strains. *FEMS Microbiol. Lett.* **137**:175–181.
6. Bessen, D., K. F. Jones, and V. A. Fischetti. 1989. Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. *J. Exp. Med.* **169**:269–283.
7. Bisno, A. L. 1995. Molecular aspects of bacterial colonisation. *Infect. Control Hosp. Epidemiol.* **16**:648–657.
8. Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–1462.
9. Boccard, F., and P. Prentki. 1993. Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3' end of transcription units. *EMBO J.* **12**:5019–5027.
10. Borst, P. 1991. Molecular genetics of antigenic variation. *Trends Biochem. Sci.* **33**:A29–A33.
11. Borst, P., W. Bitter, R. McCulloch, F. van Leeuwen, and G. Rudenki. 1995. Antigenic variation in malaria. *Cell* **82**:1–4.
12. Brendler, T. G., A. L. Abeles, L. D. Reaves, and S. J. Austin. 1997. The iteron bases and spacers of the P1 replication origin contain information that specifies the formation of a complex structure involved in initiation. *Mol. Microbiol.* **23**:559–567.
13. Bretagne, S., J. M. Costa, C. Besmond, R. Carsique, and R. Calderone. 1997. Microsatellite polymorphism in the promoter sequence of the elongation factor 3 gene of *Candida albicans* as the basis for a typing system. *J. Clin. Microbiol.* **35**:1777–1780.
14. Britten, R. J., and D. E. Kohne. 1968. Repeated sequences in DNA. *Science* **161**:529–540.
15. Brutlag, D., F. Y. An, K. E. Weaver, and D. B. Clewell. 1995. Phase variation of *Enterococcus faecalis* pAD1 conjugation functions relates to changes in iteron sequence regions. *J. Bacteriol.* **177**:5453–5459.
16. Bullitt, E., and L. Makowski. 1995. Structural polymorphism of bacterial adhesion pili. *Nature* **373**:164–167.
17. Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. M. Geoghagen, J. F. Weidman, J. L. Fuhrmann, D. Nguyen, T. R. Utterback, J. M. Kelley, J. D. Peterson, P. W. Sadow, M. C. Hanna, M. D. Cotton, K. M. Roberts, M. A. Hurst, B. P. Kaine, M. Borodovsky, H. P. Klenk, C. M. Fraser, H. O. Smith, C. Woese, and J. C. Venter. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**:1058–1068.
18. Burch, C. L., R. J. Danaher, and D. C. Stein. 1997. Antigenic variation in *Neisseria gonorrhoeae*: production of multiple lipooligosaccharides. *J. Bacteriol.* **179**:982–986.
19. Campagnari, A. A., S. M. Spinola, and A. J. Lesse. 1990. Lipopolysaccharide epitopes shared among gram negative non enteric mucosal pathogens. *Microb. Pathog.* **8**:353–362.
20. Caskey, C. T., A. Pizzuti, Y. Fu, R. G. Fenwick, Jr., and D. L. Nelson. 1992. Triplet repeat mutations in human disease. *Science* **256**:784–789.
21. Chakraborty, T., M. Leimeister, E. Domann, M. Hartl, W. Goebel, T. Nichterlein, and S. Notermans. 1992. Coordinate regulation of virulence genes in *Listeria monocytogenes* requires the product of the *prfA* gene. *J. Bacteriol.* **174**:568–574.
22. Chamberlain, J. S., R. A. Gibbs, J. E. Ranier, P. N. Nguyen, and C. T. Caskey. 1988. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Res.* **16**:11141–11156.
23. Chen, F., B. P. Currie, L. Chen, S. G. Spitzer, E. D. Spitzer, and A. Casadevall. 1995. Genetic relatedness of *Cryptococcus neoformans* clinical isolates grouped with the repetitive DNA probe CNRE-1. *J. Clin. Microbiol.* **33**:2818–2822.
24. Chiurazzi, P., L. Kozak, and G. Neri. 1994. Unstable triplets and their mutational mechanism: size reduction of the CGG repeat versus germline mosaicism in the fragile X syndrome. *Am. J. Med. Genet.* **51**:517–521.
25. Cifarelli, R. A., M. Gallitelli, and F. Cellini. 1995. Random amplified hybridization microsatellites (RAHM): isolation of a new class of microsatellite containing DNA clones. *Nucleic Acids Res.* **23**:3802–3803.
26. Clewell, D. R. 1993. Bacterial sex pheromone induced plasmid transfer. *Cell* **73**:9–12.
27. Clewell, D. R. 1993. Sex pheromones and the plasmid encoded mating response in *Enterococcus faecalis*, p. 349–367. *In* D. B. Clewell (ed.), *Bacterial conjugation*. Plenum Press, New York, N.Y.
28. Coggins, L. W., and M. O'Prey. 1989. DNA tertiary structures formed *in vitro* by misaligned hybridization of multiple tandem repeat sequences. *Nucleic Acids Res.* **17**:7417–7426.
29. Cohan, F. M. 1996. The role of genetic exchange in bacterial evolution. *ASM News* **62**:631–636.
30. Cohen, J. 1997. The genomics gamble. *Science* **275**:767–772.
31. Cookson, B. D., P. Aparicio, A. Deplano, M. J. Struelens, R. Goering, and R. Marples. 1996. Inter-centre comparison of pulsed-field gel electrophoresis for the typing of methicillin resistant *Staphylococcus aureus*. *J. Med. Microbiol.* **44**:179–184.
32. Correia, F. F., S. Inouye, and M. Inouye. 1986. A 26 basepair repetitive sequence specific for *Neisseria gonorrhoeae* and *N. meningitidis* genomic DNA. *J. Bacteriol.* **167**:1009–1015.
33. Cox, R., and S. M. Mirkin. 1997. Characteristic enrichment of DNA repeats in different genomes. *Proc. Natl. Acad. Sci. USA* **94**:5237–5242.
34. Coyne, J. A., and B. Charlesworth. 1996. Mechanisms of punctuated evolution. *Science* **274**:1748–1749.
35. De Bruijn, F. J. 1992. Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergenic consensus) sequences and the PCR to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Appl. Environ. Microbiol.* **58**:2180–2187.
36. Dempsey, J. A. F., W. Litaker, A. Madhure, T. L. Snodgrass, and J. G. Cannon. 1991. Physical map of the chromosome of *Neisseria gonorrhoeae* FA1090 with location of genetic markers including *opa* and *pil* genes. *J. Bacteriol.* **173**:5476–5486.
37. Doman, E., S. Zechel, A. Lingnau, T. Hain, A. Darji, T. Nichterlein, J. Wehland, and T. Chakraborty. 1997. Identification and characterization of a novel PrfA-regulated gene in *Listeria monocytogenes* whose product, IrpA, is highly homologous to internalin proteins, which contain leucine-rich repeats. *Infect. Immun.* **65**:101–109.
38. Dore, E., T. Pace, M. Ponzi, L. Picci, and C. Frontali. 1990. Organization of subtelomeric repeats in *Plasmodium berghei*. *Mol. Cell. Biol.* **10**:2423–2427.
39. Dunny, G. M. 1995. Pheromone inducible conjugation in *Enterococcus faecalis*: interbacterial and host-parasite chemical communication. *J. Bacteriol.* **177**:871–876.
40. Elena, S. F., V. S. Cooper, and R. E. Lenski. 1996. Punctuated evolution caused by selection of rare beneficial mutations. *Science* **272**:1802–1804.
41. Ellis, J., and J. Crampton. 1988. Characterization of a simple highly repetitive DNA sequence from the parasite *Leishmania donovani*. *Mol. Biochem. Parasitol.* **29**:9–17.
42. Ertugrul, N., M. C. Rodriguez-Barradas, D. M. Musher, M. A. K. Ryan, C. S. Agin, S. J. Murphy, M. Shayegani, and D. A. Watson. 1997. BOX PCR-based DNA analysis of nonserotypable *Streptococcus pneumoniae* implicated in outbreaks of conjunctivitis. *J. Infect. Dis.* **176**:1401–1405.
43. Field, D., L. Eggert, D. Metzgar, R. Rose, and C. Wills. 1996. Use of polymorphic short and clustered coding-region microsatellites to distinguish strains of *Candida albicans*. *FEMS Immunol. Med. Microbiol.* **15**:73–79.
44. Field, D., and C. Wills. 1996. Long, polymorphic microsatellites in simple organisms. *Proc. R. Soc. London Ser. B* **263**:209–251.
- 44a. Field, D., and C. Wills. Personal communication.
45. Filutowicz, M., S. Dellis, I. Levchenko, M. Urh, F. Wu, and D. York. 1994. Regulation of replication of an iteron-containing DNA molecule. *Prog. Nucleic Acid Res. Mol. Biol.* **48**:239–273.
46. Finlay, B. B., and S. Falkow. 1989. Common themes in microbial pathogenicity. *Microbiol. Rev.* **53**:210–230.
47. Finlay, B. B., and S. Falkow. 1997. Common themes in microbial pathogenicity revisited. *Microbiol. Mol. Biol. Rev.* **61**:136–169.
48. Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, E. R. Kerlavage, C. J. Bult, J. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gueh, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**:496–512.
49. Flock, J. I., G. Fröman, K. Jönsson, B. Guss, and C. Signas. 1987. Cloning and expression of the gene for a fibronectin binding protein from *S. aureus*. *EMBO J.* **6**:2351–2357.
50. Foote, S. J., and D. J. Kemp. 1989. Chromosomes of malaria parasites. *Trends Genet.* **5**:337–342.
51. Foster, P. L. 1993. Adaptive mutation: uses of adversity. *Annu. Rev. Microbiol.* **47**:467–504.
52. Foster, P. L., and J. M. Trimarchi. 1994. Adaptive reversion of a frameshift mutation in *Escherichia coli* by simple base deletions in homopolymeric runs. *Science* **265**:407–409.
53. Foster, T. J., and D. McDevitt. 1994. Surface associated proteins of *S. aureus*: their possible roles in virulence. *FEMS Microbiol. Lett.* **118**:199–206.
54. Francino, M. P., L. Chao, M. A. Riley, and H. Ochman. 1996. Asymmetries generated by transcriptional-coupled repair in enterobacterial genes. *Science* **272**:107–109.
55. Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. Tomb, B. A. Dougherty, K. F. Bott, P. Hu, T. S. Lucier, S. N. Peterson, H. O.

- Smith, C. A. Hutchinson III, and J. C. Venter. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**:397–403.
56. Frenay, H. M. E., A. E. Bunschoten, L. M. Schouls, W. J. van Leeuwen, C. M. J. E. Vandenbroucke-Grauls, J. Verhoef, and F. R. Mooi. 1996. Molecular typing of methicillin resistant *Staphylococcus aureus* on the basis of protein A gene polymorphisms. *Eur. J. Clin. Microbiol. Infect. Dis.* **15**:60–64.
57. Frenay, H. M. E., J. P. G. Theelen, L. M. Schouls, C. M. J. E. Vandenbroucke-Grauls, J. Verhoef, W. J. van Leeuwen, and F. R. Mooi. 1994. Discrimination of epidemic and non-epidemic methicillin resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *J. Clin. Microbiol.* **32**:846–847.
58. Gerber, H., K. Seipel, O. Georgier, M. Hofferer, M. Flag, S. Rencobi, and W. Scheffner. 1994. Transcriptional activation modified by homopolymeric glutamine and proline stretches. *Science* **263**:808–811.
59. Gibbs, C., B. Reimann, and E. Shultz. 1989. Reassortment of pilin genes in *N. gonorrhoeae* occurs by two distinct mechanisms. *Nature* **338**:651–652.
60. Giesendorf, B. A. J., A. van Belkum, A. Koeken, H. Stegeman, M. H. C. Henkens, J. van der Plas, H. Goossens, H. G. M. Niesters, and W. G. V. Quint. 1993. Development of species-specific DNA probes for *Campylobacter jejuni*, *Campylobacter coli*, and *Campylobacter lari* by PCR fingerprinting. *J. Clin. Microbiol.* **31**:1541–1546.
61. Gilson, E., S. Bachelier, S. Perrin, D. Perrin, P. A. Grimont, F. Grimont, and M. Hofnung. 1990. Palindromic unit highly repetitive DNA sequences exhibit species specificity within enterobacteriaceae. *Res. Microbiol.* **141**:1103–1116.
62. Girardin, H., J. Latge, T. Srikantha, B. Morrow, and D. R. Soll. 1993. Development of DNA probes for fingerprinting *Aspergillus fumigatus*. *J. Clin. Microbiol.* **31**:1547–1554.
63. Giroux, S., J. Beaudet, and R. Cedergren. 1988. Highly repetitive tRNA(Pro)-tRNA(His) gene cluster from *Photobacterium phosphoreum*. *J. Bacteriol.* **170**:5601–5606.
64. Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldman, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. 1996. Life with 6000 genes. *Science* **274**:546–567.
65. Goh, S. H., S. K. Byrne, and A. W. Chow. 1992. Molecular typing of *Staphylococcus aureus* on the basis of coagulase gene polymorphisms. *J. Clin. Microbiol.* **30**:1642–1645.
66. Gravekamp, C., D. L. Kasper, J. L. Michel, D. E. Kling, V. Carey, and L. C. Madoff. 1997. Immunogenicity and protective efficacy of the alpha C protein of group B streptococci are inversely related to the number of repeats. *Infect. Immun.* **65**:5216–5221.
67. Gravekamp, C., D. S. Horensky, J. L. Michel, and L. C. Madoff. 1996. Variation in repeat number within the alpha C protein of group B streptococci alters antigenicity and protective epitopes. *Infect. Immun.* **64**:3576–3583.
68. Groisman, E. A., and H. Ochman. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* **87**:791–794.
69. Haas, R., and T. F. Meyer. 1986. The repertoire of silent pilus genes in *N. gonorrhoeae*: evidence for gene conversion. *Cell* **44**:107–115.
70. Hauge, X. Y., and M. Litt. 1993. A study of the origin of “shadow bands” seen when typing dinucleotide repeat polymorphisms by the PCR. *Nucleic Acids Res.* **21**:411–415.
71. Heath, D. D., G. K. Iwama, and R. H. Devlin. 1993. PCR primed with VNTR core sequences yields species specific patterns and hypervariable probes. *Nucleic Acids Res.* **21**:5782–5785.
72. Heath, D. G., F. Y. An, K. E. Weaver, and D. B. Clewell. 1995. Phase variation of *Enterococcus faecalis* pAD1 conjugation functions relates to changes in iteron sequence region. *J. Bacteriol.* **177**:5453–5459.
73. Henderson, S. T., and T. D. Petes. 1992. Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**:2749–2757.
74. Hermann, M., S. J. Suchard, L. A. Boxer, F. A. Waldvogel, and P. D. Lew. 1991. Thrombospondin binds to *Staphylococcus aureus* and promotes staphylococcal adherence to surfaces. *Infect. Immun.* **59**:279–288.
75. Hermans, P. W. M., D. van Soelingen, and J. D. A. van Embden. 1992. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *M. kansasii* and *M. goodii*. *J. Bacteriol.* **174**:4157–4165.
76. Hermans, P. W. M., D. van Soelingen, E. M. Bik, P. E. W. de Haas, J. W. Dale, and J. D. A. van Embden. 1991. The insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot spot integration region for insertion elements in *M. tuberculosis* complex strains. *Infect. Immun.* **59**:2695–2705.
77. Higgins, C. F., G. F. L. Ames, W. M. Barnes, J. M. Clement, and M. Hofnung. 1982. A novel intergenic regulatory element of prokaryotic operons. *Nature* **298**:760–762.
78. High, N. J., M. E. Deadman, and E. R. Moxon. 1993. The role of the repetitive DNA motif (5'-CAAT-3') in the variable expression of the *H. influenzae* lipopolysaccharide epitope alpha-Gal(1-4)beta-Gal. *Mol. Microbiol.* **9**:1275–1282.
79. High, N. J., M. E. Deadman, D. W. Hood, and E. R. Moxon. 1996. The identification of a novel gene required for lipopolysaccharide biosynthesis by *Haemophilus influenzae* RM7004, using transposon Tn916 mutagenesis. *FEMS Microbiol. Lett.* **145**:325–331.
80. Highlander, S. K., and O. Garza. 1996. The restriction modification system of *Pasteurella haemolytica* is a member of a new family of type I enzymes. *Gene* **178**:89–96.
81. Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkil, B. Li, and R. Herrmann. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**:4421–4449.
82. Hite, J. M., K. A. Eckert, and K. C. Cheng. 1996. Factors affecting the fidelity of DNA synthesis during PCR amplification of d(C-A)<sub>n</sub> · d(G-T)<sub>n</sub> microsatellite repeats. *Nucleic Acids Res.* **24**:2429–2434.
83. Hood, D. W., M. E. Deadman, M. P. Jennings, M. Bisercic, R. D. Fleischmann, J. C. Venter, and E. R. Moxon. 1996. DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. USA* **93**:11121–11125.
84. Hoover, T. A., M. H. Vodkin, and J. C. Williams. 1992. A *Coxiella burnetii* repeated DNA element resembling a bacterial insertion sequence. *J. Bacteriol.* **174**:5540–5548.
85. Hulton, C. S., C. F. Higgins, and P. M. Sharp. 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.
86. Inzana, T. J., J. Hensley, J. McQuiston, A. J. Lesse, A. A. Campagnari, S. M. Boyle, and M. A. Apicella. 1997. Phase variation and conservation of lipooligosaccharide epitopes in *Haemophilus somnus*. *Infect. Immun.* **65**:4675–4681.
87. Jackson, P. J., E. A. Walthers, A. S. Kalif, K. L. Richmond, D. M. Adair, K. H. Hill, C. R. Kuske, G. L. Andersen, K. H. Wilson, M. E. Hugh-Jones, and P. Keim. 1997. Characterization of the variable number of tandem repeats in *vvfA* from different *Bacillus anthracis* isolates. *Appl. Environ. Microbiol.* **63**:1400–1405.
88. Janssen, P., R. Coopman, G. Huys, J. Swings, M. Bleeker, P. Vos, M. Zabeau, and K. Kersters. 1996. Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology* **142**:1881–1893.
89. Jeffreys, A. J., A. MacLeod, K. Tamaki, D. L. Neil, and D. G. Monckton. 1991. Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354**:204–209.
90. Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Hypervariable “minisatellite” regions in human DNA. *Nature* **314**:67–73.
91. Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Individual specific fingerprints of human DNA. *Nature* **316**:76–79.
92. Jeffreys, A. J., V. Wilson, S. L. Thein, D. J. Weatherall, and B. A. J. Ponder. 1986. DNA fingerprints and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet.* **39**:11–24.
93. Jin, H., Z. Ren, J. M. Poszgay, C. Elkins, P. W. Whitby, D. J. Morton, and T. L. Stull. 1996. Cloning of a DNA fragment encoding a heme-repressible hemoglobin binding outer membrane protein from *Haemophilus influenzae*. *Infect. Immun.* **64**:3134–3141.
94. Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soelingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. D. A. van Embden. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**:907–914.
95. Karlin, S., J. Mrazek, and A. M. Campbell. 1997. Compositional biases of bacterial genomes and evolutionary implications. *J. Bacteriol.* **179**:3899–3913.
96. Karlin, S., J. Mrazek, and A. M. Campbell. 1996. Frequent oligonucleotides and peptides of the *Haemophilus influenzae* genome. *Nucleic Acids Res.* **24**:4263–4272.
97. Kawabata, S., T. Miyata, Y. Yoshizawa, H. Igarashi, and S. Iwanaga. 1989. The amino acid sequence of the procoagulant- and prothrombin-binding domain isolated from staphylocoagulase. *J. Biol. Chem.* **261**:527–531.
98. Kittell, B. L., and D. R. Helinski. 1993. Plasmid incompatibility and replication control, p. 223–242. *In* D. B. Clewell (Ed.), *Bacterial conjugation*. Plenum Press, New York, N.Y.
99. Kobe, B., and J. Deisenhofer. 1994. The leucine-rich repeat: a versatile binding motif. *Trends Biochem. Sci.* **19**:415–421.
100. Koeth, T., J. Versalovic, and J. R. Lupski. 1995. Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. *PCR Methods Applic.* **5**:408–418.
101. Leach, F. S., N. C. Nicolaides, N. Papadopoulos, B. Liu, J. Jen, R. Parsons, P. Peltomaki, P. Sistonen, L. A. Aaltonen, M. Nystrom, X. Y. Guan, G. J. Zhang, P. S. Meltzer, J. W. Yu, F. T. Kao, D. A. Chen, D. M. Cersosaletti, R. E. K. Fournier, S. Todd, T. Lewis, R. J. Leach, S. L. Naylor, J. Green, J. Jass, P. Watson, H. T. Lynch, J. M. Trent, A. de la Chapelle, K. W. Kinzler, and B. Vogelstein. 1993. Mutations of a MutS homolog in hereditary non-polyposis colorectal cancer. *Cell* **75**:1215–1225.
102. Lenski, R. E., and P. D. Sniegowski. 1995. Directed mutations slip sliding away? *Curr. Biol.* **5**:97–99.
103. Levinson, G., and G. A. Gutman. 1987. High frequency of short frameshifts

- in poly-CA/GT tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res.* **15**:5323–5338.
104. Levinson, G., and G. A. Gutman. 1987. Slipped-strand mispairing: mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**:203–221.
  105. Lingnau, A., T. Chakraborty, K. Niebuhr, E. Domann, and J. Wehland. 1996. Identification and purification of a novel internalin-related protein in *Listeria monocytogenes* and *Listeria ivanovii*. *Infect. Immun.* **64**:1002–1006.
  106. Lupski, J. R., and G. M. Weinstock. 1992. Short interspersed repetitive DNA sequences in prokaryotic genomes. *J. Bacteriol.* **174**:4525–4529.
  107. Lysnyansky, I., R. Rosengarten, and D. Yogeve. 1996. Phenotypic switching of variable surface lipoproteins in *Mycoplasma bovis* involves high frequency chromosomal rearrangements. *J. Bacteriol.* **178**:5395–5401.
  108. Madoff, L. C., J. L. Michel, E. W. Gong, D. E. Kling, and D. L. Kasper. 1996. Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. *Proc. Natl. Acad. Sci. USA* **93**:4131–4136.
  109. Makhov, A. M., J. H. Hannah, M. J. Brennan, B. L. Trus, E. Kocsis, J. F. Conway, P. T. Wingfield, M. N. Simon, and A. C. Steven. 1994. Filamentous hemagglutinin of *Bordetella pertussis*. A bacterial adhesin formed as a 50 nm monomeric rigid rod based on a 19-residue repeat motif rich in beta strands and turns. *J. Mol. Biol.* **241**:110–124.
  110. Makino, S., J. P. M. van Putten, and T. F. Meyer. 1991. Phase variation of the opacity outer membrane protein controls invasion by *N. gonorrhoeae* into human epithelial cells. *EMBO J.* **10**:1307–1315.
  111. Mangiarini, L., K. Sathasivam, M. Seller, B. Cozens, A. Harper, C. Hetherington, M. Lawton, Y. Trotter, H. Lehrach, S. W. Davies, and G. P. Bates. 1996. Exon 1 of the *HD* gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87**:493–506.
  112. Marshall, D. G., D. C. Coleman, D. J. Sullivan, H. Xia, C. A. O'Morain, and C. J. Smyth. 1996. Genomic DNA fingerprinting of clinical isolates of *Helicobacter pylori* using short oligonucleotide probes containing repetitive sequences. *J. Appl. Bacteriol.* **81**:509–517.
  113. Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, D. A. Morrison, G. J. Boulnois, and J. P. Claverys. 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. *Nucleic Acids Res.* **20**:3479–3483.
  114. Maslow, J. N., M. E. Mulligan, and R. D. Arbeit. 1993. Molecular epidemiology: application of contemporary techniques to the typing of microorganisms. *Clin. Infect. Dis.* **17**:153–164.
  115. McAlpin, C. E., and B. Mannarelli. 1995. Construction and characterization of a DNA probe for distinguishing strains of *Aspergillus fumigatus*. *Appl. Environ. Microbiol.* **61**:1068–1072.
  116. McDevitt, D., and T. J. Foster. 1995. Variation in the size of the repeat region of the fibrinogen receptor (clumping factor) of *Staphylococcus aureus* strains. *Microbiology* **141**:937–943.
  117. McDevitt, D., P. Francois, P. Vaudaux, and T. J. Foster. 1994. Molecular characterization of the clumping factor (fibrinogen receptor) of *S. aureus*. *Mol. Microbiol.* **11**:237–248.
  118. Mecsas, J., and E. J. Strauss. 1996. Molecular mechanisms of bacterial virulence: type III secretion and pathogenicity islands. *Emerg. Infect. Dis.* **2**:271–288.
  119. Meyer, T. F., C. P. Gibbs, and R. Haas. 1990. Variation and control of protein expression in *Neisseria*. *Annu. Rev. Microbiol.* **44**:451–477.
  120. Meyer, W., A. Koch, C. Niemann, B. Beyermann, J. T. Eppelen, and T. Boerner. 1991. Differentiation of species and strains among filamentous fungi by DNA fingerprinting. *Curr. Genet.* **19**:239–242.
  121. Meyer, W., T. G. Mitchell, E. Z. Freedman, and R. Vilgalys. 1993. Hybridization probes for conventional DNA fingerprinting used as single primers in the polymerase chain reaction to distinguish strains of *Cryptococcus neoformans*. *J. Clin. Microbiol.* **31**:2274–2280.
  122. Mitas, M. 1997. Trinucleotide repeats associated with human disease. *Nucleic Acids Res.* **25**:2245–2253.
  123. Mitchell, T. G., and J. R. Perfect. 1995. Cryptococcosis in the era of AIDS: 100 years after the discovery of *Cryptococcus neoformans*. *Clin. Microbiol. Rev.* **8**:515–548.
  124. Motalebi-Veshareh, M., D. Baizer, E. Lanka, G. Jagura-Burdzy, and C. M. Thomas. 1992. Conjugative transfer functions of broad host range plasmid RK2 are coregulated with vegetative replication. *Mol. Microbiol.* **6**:907–920.
  125. Moxon, E. R., and P. B. Rainey. 1995. Pathogenic bacteria: the wisdom of their genes, p. 255–268. *In* B. A. M. Van der Zeijst, L. Van Alphen, W. P. M. Hoekstra, and J. D. A. van Embden (ed.), *Ecology of pathogenic bacteria*. Royal Dutch Academy of Sciences, second series, no. 96. Royal Dutch Academy of Sciences, Amsterdam, The Netherlands.
  126. Moxon, E. R., P. B. Rainey, M. A. Nowak, and R. E. Lenski. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**:24–33.
  127. Murphy, G. L., T. D. Connell, D. S. Barritt, M. Koohey, and J. G. Cannon. 1989. Phase variation of gonococcal protein II: regulation of gene expression by slipped strand mispairing of a repetitive DNA sequence. *Cell* **56**:539–547.
  128. Musser, J. M. 1996. Molecular population genetic analysis of emerged bacterial pathogens: selected insights. *Emerg. Infect. Dis.* **2**:1–17.
  129. Nadir, E., H. Margalit, T. Gallily, and S. S. Ben-Sasson. 1996. Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proc. Natl. Acad. Sci. USA* **93**:6470–6475.
  130. Nakamura, Y., M. Carlson, K. Krapcho, M. Kanamori, and R. White. 1988. New approach for isolation of VNTR markers. *Am. J. Hum. Genet.* **43**:854–859.
  131. Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, and R. White. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**:1616–1622.
  132. Naveilhan, P., C. Baudet, W. Jabbour, and D. Wion. 1994. A theory that may explain the Hayflick limit: a means to delete one copy of a repeating sequence during each cell cycle in certain human cells such as fibroblasts. *Mech. Ageing Dev.* **75**:205–213.
  133. Noormohammadi, A. H., P. F. Markham, K. G. Whithear, I. D. Walker, V. A. Gurevich, D. H. Ley, and G. F. Browning. 1997. *Mycoplasma synoviae* has two distinct phase-variable major membrane antigens, one of which is a putative hemagglutinin. *Infect. Immun.* **65**:2542–2547.
  134. Odds, F. C. 1994. *Candida* species and virulence. *ASM News* **60**:313–318.
  135. Oppenheim, A. B., K. E. Rudd, I. Mendelson, and D. Teff. 1993. Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli*. *Mol. Microbiol.* **10**:113–122.
  136. Orti, G., D. E. Pearse, and J. C. Avise. 1997. Phylogenetic assessment of length variation at a microsatellite locus. *Proc. Natl. Acad. Sci. USA* **94**:10745–10749.
  137. Ozaki, L. S., P. Svec, R. S. Nussenzweig, V. Nussenzweig, and G. N. Godson. 1983. Structure of the *Plasmodium knowlesi* gene coding for the circumsporozoite protein. *Cell* **34**:815–822.
  138. Park, P. W., D. D. Roberts, L. E. Grosso, W. C. Parks, and J. Rosenbloom. 1991. Binding of elastin to *S. aureus*. *J. Biol. Chem.* **266**:23399–23406.
  139. Patti, J. M., B. L. Allen, M. J. McGavin, and M. Höök. 1994. MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu. Rev. Microbiol.* **48**:585–617.
  140. Patti, J. M., K. Jönsson, B. Guss, L. M. Switalski, and K. M. Wiberg. 1992. Molecular characterization and expression of a gene encoding a *S. aureus* collagen adhesin. *J. Biol. Chem.* **267**:4766–4772.
  141. Paulsson, M., O. D. Liang, F. Ascencio, and T. Wadström. 1992. Vitronectin-binding surface proteins of *S. aureus*. *Zentralbl. Bakteriol.* **277**:54–64.
  142. Peak, I. R. A., M. P. Jennings, D. W. Hood, M. Bisceric, and E. R. Moxon. 1996. Tetrameric repeat units associated with virulence factor phase variation in *Haemophilus* also occur in *Neisseria* spp. and *Moraxella catarrhalis*. *FEMS Microbiol. Lett.* **137**:109–114.
  143. Peixuan, Z., H. Xujiang, and X. Li. 1995. Typing *N. meningitidis* by analysis of RFLPs in the gene encoding the class 1 outer membrane protein: application to assessment of epidemics throughout the last 4 decades in China. *J. Clin. Microbiol.* **33**:458–462.
  144. Peterson, S. N., C. C. Bailey, J. S. Jensen, M. B. Borre, E. S. King, K. F. Bott, and C. A. Hutchison III. 1995. Characterization of repetitive DNA in the *Mycoplasma genitalium* genome: possible role in the generation of antigenic variation. *Proc. Natl. Acad. Sci. USA* **92**:11829–11833.
  145. Philipp, W. J., S. Poulet, K. Eiglmeier, L. Pascopella, V. Balasubramanian, B. Heym, S. Bergh, B. R. Bloom, W. R. Jacobs, Jr., and S. T. Cole. 1996. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proc. Natl. Acad. Sci. USA* **93**:3132–3137.
  146. Poh, C. L., V. Ramachandran, and J. W. Tapsall. 1996. Genetic diversity of *N. gonorrhoeae* IB-2 and IB-6 isolates revealed by whole cell repetitive element sequence based PCR. *J. Clin. Microbiol.* **34**:292–295.
  147. Reznick, D. N., F. H. Shaw, F. H. Rodd, and R. G. Shaw. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275**:1934–1937.
  148. Richardson, T., S. Cato, J. Ramser, G. Kahl, and K. Weising. 1995. Hybridization of microsatellites to RAPD: a new source of polymorphic markers. *Nucleic Acids Res.* **23**:3798–3799.
  149. Riesenfeld, C., M. Everett, L. J. V. Piddock, and B. G. Hall. 1997. Adaptive mutations produce resistance to ciprofloxacin. *Antimicrob. Agents Chemother.* **41**:2059–2060.
  150. Riley, D. E., M. Samadpour, and J. N. Krieger. 1991. Detection of variable DNA repeats in diverse eukaryotic microorganisms by a single set of polymerase chain reaction primers. *J. Clin. Microbiol.* **29**:2746–2751.
  151. Robertson, B. D., and T. F. Meyer. 1992. Genetic variation in pathogenic bacteria. *Trends Genet.* **8**:422–427.
  152. Roche, R. J., and E. R. Moxon. 1995. Phenotypic variation in *H. influenzae*: the interrelationship of colony opacity, capsule and lipopolysaccharide. *Microb. Pathog.* **18**:129–140.
  153. Roche, R. J., N. J. High, and E. R. Moxon. 1994. Phase variation in *H. influenzae* lipopolysaccharide: characterization of lipopolysaccharide from individual colonies. *FEMS Microbiol. Lett.* **120**:279–284.
  154. Rosche, W. A., A. Jaworski, S. Kang, S. F. Kramer, J. E. Larsson, D. P. Geidroc, R. D. Wells, and R. R. Sinden. 1996. Single-stranded DNA-binding

- protein enhances the stability of CTG triplet repeats in *Escherichia coli*. J. Bacteriol. 178:5042–5044.
155. Rosche, W. A., T. Q. Trinh, and R. R. Sinden. 1995. Differential DNA secondary-structure-mediated deletion mutation in the leading and lagging strands. J. Bacteriol. 177:4385–4391.
  156. Rosenberg, S. M., S. Longrich, P. Gee, and R. S. Harris. 1994. Adaptive mutation by deletions in small mononucleotide repeats. Science 265:405–407.
  157. Ross, C., K. Raios, K. Jackson, and B. Dwyer. 1992. Molecular cloning of a highly repeated element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. J. Clin. Microbiol. 30:942–946.
  158. Ryden, C., I. Maxe, A. Ljungh, A. Franzen, and D. Heinegard. 1987. Selective binding of BSP to *S. aureus* in osteomyelitis. Lancet ii:515–518.
  159. Sadhu, C., M. J. McEachern, E. P. Rustchenko-Bulgac, J. Schmid, D. R. Soll, and J. B. Hicks. 1991. Telomeric and dispersed repeat sequences in *Candida* yeast and their use in strain identification. J. Bacteriol. 173:842–850.
  160. SanMiguel, P., A. Tikhonov, Y. K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P. S. Springer, K. J. Edwards, M. Lee, Z. Avramova, and J. L. Bennetzen. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768.
  161. Sawyer, L. A., J. M. Hennesy, A. A. Peixoto, E. Rosato, H. Parkinson, R. Costa, and C. P. Kyriacou. 1997. Natural variation in a *Drosophila* clock gene and temperature compensation. Science 278:2117–2120.
  162. Scherer, S., and D. A. Stevens. 1988. A *Candida albicans* dispersed repeated gene family and its epidemiologic applications. Proc. Natl. Acad. Sci. USA 85:1452–1456.
  163. Scherer, S., and P. T. Magee. 1990. Genetics of *Candida albicans*. Microbiol. Rev. 54:226–241.
  - 163a. Scherer, S., and A. van Belkum. Unpublished data.
  164. Schlotterer, C., B. Amos, and D. Tautz. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature 354:63–65.
  165. Schwarzkopf, A., and H. Karch. 1994. Genetic variation in *Staphylococcus aureus* coagulase genes: potential and limits for use as epidemiological marker. J. Clin. Microbiol. 32:2407–2412.
  166. Seifert, H. S., R. Ajioka, and C. Marchal. 1988. DNA transformation leads to pilin antigenic variation in *N. gonorrhoeae*. Nature 336:392–395.
  167. Sharples, G. J., and R. G. Lloyd. 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. Nucleic Acids Res. 18:6503–6508.
  168. Shields, D. C., D. McDevitt, and T. J. Foster. 1995. Evidence against concerted evolution in a tandem repeat array in the clumping factor gene of *Staphylococcus aureus*. Mol. Biol. Evol. 12:963–965.
  169. Skinner, P. J., B. T. Koshy, C. J. Cummings, I. A. Klement, K. Helin, A. Servadio, H. Y. Zoghbi, and H. T. Orr. 1997. Ataxin-1 with an expanded glutamine tract alters nuclear matrix associated structures. Nature 389:971–974.
  170. Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA 90:4384–4388.
  171. Sobel, J. D., J. Vazquez, M. Lynch, C. Meriwether, and M. J. Zervos. 1993. Vaginitis due to *S. cerevisiae*: epidemiology, clinical aspects and therapy. Clin. Infect. Dis. 16:93–99.
  172. Soll, D. R. 1992. High frequency switching in *Candida albicans*. Clin. Microbiol. Rev. 5:183–203.
  173. Soll, D. R., C. J. Langtimm, J. McDowell, J. Hicks, and R. Galask. 1987. High frequency switching in *Candida* strains isolated from vaginitis patients. J. Clin. Microbiol. 25:1611–1622.
  174. Soll, D. R., M. Staebell, C. Langtimm, M. Pfaller, J. Hicks, and T. V. Gopala Rao. 1988. Multiple *Candida* strains in the course of a systemic infection. J. Clin. Microbiol. 26:1448–1459.
  175. Soll, D. R., R. Galask, S. Isley, T. V. Gopala Rao, D. Stone, J. Hicks, K. Mac, and C. Hanna. 1989. Switching of *Candida albicans* during successive episodes of recurrent vaginitis. J. Clin. Microbiol. 27:681–690.
  176. Stern, A., and T. F. Meyer. 1987. Common mechanism controlling phase and antigenic variation in pathogenic neisseriae. Mol. Microbiol. 1:5–12.
  177. Stern, A., M. Brown, P. Nickel, and T. F. Meyer. 1986. Opacity genes in *N. gonorrhoeae*: control of phase and antigenic variation. Cell 47:61–71.
  178. Strand, S., T. A. Prolla, R. M. Liskay, and T. D. Petes. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365:274–276.
  179. Strauss, E. J., and S. Falkow. 1997. Microbial pathogenesis: genomics and beyond. Science 276:707–712.
  180. Sullivan, D., K. Haynes, G. Moran, D. Shanley, and D. Coleman. 1996. Persistence, replacement and microevolution of *Cryptococcus neoformans* strains in recurrent meningitis in AIDS patients. J. Clin. Microbiol. 34:1739–1744.
  181. Swaminathan, B., G. M. Matar, M. W. Reeves, L. M. Graves, G. Ajello, W. F. Bibb, L. O. Helsel, M. Morales, H. Dronavalli, M. El-Swift, W. DeWitt, and S. B. Hunter. 1996. Molecular subtyping of *Neisseria meningitidis* serogroup B: comparison of five methods. J. Clin. Microbiol. 34:1468–1473.
  182. Switalski, L. M., J. M. Patti, W. Butcher, A. G. Cristina, P. Speziale, et al. 1993. A collagen receptor from *S. aureus* strains isolated from patients with septic arthritis mediates adhesion to cartilage. Mol. Microbiol. 7:99–107.
  183. Theiss, P., and K. S. Wise. 1997. Localized frameshift mutation generates selective, high-frequency phase variation of a surface lipoprotein encoded by a mycoplasma ABC transporter operon. J. Bacteriol. 179:4013–4022.
  184. Thibodeau, S. N., G. Bren, and D. Schaid. 1993. Microsatellite instability in cancer of the proximal colon. Science 260:816–819.
  185. Thorisdottir, A. S., L. L. Carias, S. H. Marshall, M. Green, M. J. Zervos, C. Giorgio, L. A. Mermel, J. M. Boyce, A. A. Medeiros, H. Fraimow, and L. B. Rice. 1994. IS6770, and enterococcal insertion-like sequence useful for determining the clonal relationship of clinical enterococcal isolates. J. Infect. Dis. 170:1539–1548.
  186. Tinsley, C. R., and X. Nassif. 1996. Analysis of the genetic differences between *N. meningitidis* and *N. gonorrhoeae*: two closely related bacteria expressing two different pathogenicities. Proc. Natl. Acad. Sci. USA 93:11109–11114.
  187. Tripathi, J., and S. K. Brahmachari. 1977. Synthesis of hybrid bacterial plasmids containing highly repeated satellite DNA. Cell 10:509–519.
  188. Turner, B. J., J. F. Elder, T. F. Laughlin, and W. P. Davis. 1990. Genetic variation in clonal vertebrates detected by simple-sequence DNA fingerprinting. Proc. Natl. Acad. Sci. USA 87:5653–5657.
  189. Uitterlinden, A. G., P. E. Slagboom, T. E. Johnson, and J. Vijg. 1989. The *Caenorhabditis elegans* genome contains monomorphic minisatellites and simple sequences. Nucleic Acids Res. 17:9527–9530.
  190. Van Belkum, A. 1994. DNA fingerprinting of medically important microorganisms by use of PCR. Clin. Microbiol. Rev. 7:174–184.
  191. Van Belkum, A., B. Duim, A. Regelink, L. Moeller, W. Quint, and L. van Alphen. 1994. Genomic DNA fingerprinting of clinical *Haemophilus influenzae* isolates by polymerase chain reaction amplification: comparison with major outer membrane protein and restriction fragment length polymorphism analysis. J. Med. Microbiol. 41:63–68.
  192. Van Belkum, A., J. Kluytmans, W. van Leeuwen, R. Bax, W. Quint, E. Peters, A. Fluit, C. Vandembrouck-Grauls, A. van den Brule, H. Koeleman, W. Melchers, J. Meis, A. Elaichouni, M. Vanechoutte, F. Moonens, N. Maes, M. Struelens, F. Tenover, and H. A. Verbrugh. 1995. Multicenter evaluation of arbitrarily primed PCR for typing of *Staphylococcus aureus* strains. J. Clin. Microbiol. 33:1537–1547.
  193. Van Belkum, A., J. Ramesar, G. Trommelen, and A. G. Uitterlinden. 1992. Mini- and micro-satellites in the genome of rodent malaria parasites. Gene 118:81–86.
  194. Van Belkum, A., M. Sluiter, R. de Groot, H. A. Verbrugh, and P. Hermans. 1996. Novel BOX repeat PCR assay for high-resolution typing of *Streptococcus pneumoniae* strains. J. Clin. Microbiol. 34:1176–1179.
  195. Van Belkum, A., N. Riewerts Eriksen, M. Sijmons, W. van Leeuwen, M. Vandenberg, J. Kluytmans, F. Espersen, and H. A. Verbrugh. 1996. Are variable repeats in the *spa* gene suitable targets for epidemiological studies of MRSA strains? Eur. J. Clin. Microbiol. Infect. Dis. 15:768–769.
  196. Van Belkum, A., S. Scherer, W. van Leeuwen, D. Willemse, L. van Alphen, and H. A. Verbrugh. 1997. Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*. Infect. Immun. 65:5017–5027.
  197. Van Belkum, A., W. J. G. Melchers, C. IJsseldijk, L. Nohlmans, H. A. Verbrugh, and J. F. G. M. Meis. 1997. Outbreak of amoxicillin-resistant *Haemophilus influenzae* type b: variable number of tandem repeats as novel molecular markers. J. Clin. Microbiol. 35:1517–1520.
  - 197a. Van Belkum, A., et al. Unpublished data.
  198. Van der Ende, A., C. T. Hopman, S. Zaai, B. B. Essink, B. Berkhout, and J. Dankert. 1995. Variable expression of class I outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing of the –10 and –35 regions of the promoter. J. Bacteriol. 177:2475–2480.
  199. Van Ham, S. M., L. van Alphen, F. R. Mooi, and J. P. M. van Putten. 1993. Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. Cell 73:1187–1196.
  200. Van Ham, S. M., L. van Alphen, F. R. Mooi, and J. P. M. van Putten. 1994. The fimbrial gene cluster of *H. influenzae* type b. Mol. Microbiol. 13:673–684.
  201. Van Putten, J. P. M. 1993. Phase variation of lipopolysaccharide directs interconversion of invasive and immuno-resistant phenotypes of *N. gonorrhoeae*. EMBO J. 12:4043–4051.
  202. Van Soelingen, D., P. E. W. de Haas, P. W. M. Hermans, P. M. A. Groenen, and J. D. A. van Embden. 1993. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 31:1987–1995.
  203. Vercelotti, G. M., J. B. McCarthy, P. Lindholm, P. R. Peterson, H. S. Jacobs, et al. 1985. Extracellular matrix proteins (fibronectin, laminin and type IV collagen) bind and aggregate bacteria. Am. J. Pathol. 120:13–21.
  204. Versalovic, J., and J. R. Lupski. 1995. DNA fingerprinting of *Neisseria* strains by rep-PCR. Methods Mol. Cell. Biol. 5:96–104.
  205. Versalovic, J., C. R. Woods, G. R. Georghiou, R. J. Hamill, and J. R. Lupski. 1993. DNA-based identification and epidemiologic typing of bacterial pathogens. Arch. Pathol. Lab. Med. 117:1088–1098.
  206. Versalovic, J., M. Schneider, F. J. de Bruijn, and J. R. Lupski. 1994.

- Genomic fingerprinting of bacteria using repetitive sequence based PCR. *Methods Mol. Cell. Biol.* **5**:25–40.
207. **Versalovic, J., T. Koeth, and J. R. Lupski.** 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res.* **19**:406–409.
  208. **Virji, M., J. N. Weiser, A. A. Lindberg, and E. R. Moxon.** 1990. Antigenic similarities in lipopolysaccharides of *Haemophilus* and *Neisseria* and expression of a digalactoside structure also present on human cells. *Microb. Pathog.* **9**:441–450.
  209. **Vogt, P.** 1990. Potential genetic functions of tandemly repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". *Hum. Genet.* **84**:301–336.
  210. **Weaver, K. E., D. B. Clewell, and F. Y. An.** 1993. Identification, characterization and nucleotide sequence of a region of *Enterococcus faecalis* pheromone responsive plasmid pAD1 capable of autonomous replication. *J. Bacteriol.* **175**:1900–1909.
  211. **Weiser, J. N., D. J. Maskell, P. D. Butler, A. A. Lindberg, and E. R. Moxon.** 1990. Characterization of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide. *J. Bacteriol.* **172**:3304–3309.
  212. **Weiser, J. N., J. M. Love, and E. R. Moxon.** 1989. The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. *Cell* **59**:657–665.
  213. **Welsh, J., and M. McClelland.** 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18**:7213–7218.
  214. **Wiid, I. J. F., C. Weryly, N. Beyers, P. Donald, and P. D. van Helden.** 1994. Oligonucleotide (GTG)<sub>5</sub> as a marker for *Mycobacterium tuberculosis* strain identification. *J. Clin. Microbiol.* **32**:1318–1321.
  215. **Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey.** 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**:6531–6535.
  216. **Wirth, R.** 1994. The sex pheromone system of *Enterococcus faecalis*: more than just a plasmid collection system? *Eur. J. Biochem.* **222**:235–246.
  217. **Yagi, Y., and D. B. Clewell.** 1977. Identification and characterisation of a small sequence located at two sites on the amplifiable tetracycline resistance plasmid pAM-alpha-1 in *Streptococcus faecalis*. *J. Bacteriol.* **129**:400–406.
  218. **Yogev, D., R. Rosengarten, R. Watson, and K. S. Wise.** 1991. Molecular basis of *Mycoplasma* surface antigenic variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5' regulatory sequences. *EMBO J.* **10**:4069–4079.
  219. **Yother, J., and D. E. Briles.** 1992. Structural properties and evolutionary relationships of PspA, a surface protein of *Streptococcus pneumoniae*, as revealed by sequence analysis. *J. Bacteriol.* **174**:601–609.
  220. **Zerva, L., R. J. Hollis, and M. A. Pfaller.** 1996. In vitro susceptibility testing and DNA typing of *Saccharomyces cerevisiae* clinical isolates. *J. Clin. Microbiol.* **34**:3031–3034.
  221. **Zhang, J. P., and S. Normark.** 1996. Induction of gene expression in *Escherichia coli* after pilus mediated adherence. *Science* **273**:1234–1236.
  222. **Zheng, X., L. Teng, H. L. Watson, J. I. Glass, A. Blanchard, and G. H. Cassell.** 1995. Small repeating units within the *Ureaplasma urealyticum* MB antigen gene encode serovar specificity and are associated with antigen size variation. *Infect. Immun.* **63**:891–898.