Variable Number of Tandem Repeats in Clinical Strains of *Haemophilus influenzae*

ALEX VAN BELKUM,[1]* STEWART SCHERER,[2] WILLEM VAN LEEUWEN,[1]
DIANA WILLEMSE,[1] LOEK VAN ALPHEN,[3] AND HENRI VERBRUGH[1]

*Department of Bacteriology, University Hospital Rotterdam, 3015 GD Rotterdam,[1] and Department of Medical
Microbiology, Academic Medical Center, 1105 Amsterdam,[2] The Netherlands, and Department of
Microbiology, University of Minnesota Medical School, Minneapolis, Minnesota 5545-0312[3]*

An algorithm capable of identifying short repeat motifs was developed and used to screen the whole genome sequence available for *Haemophilus influenzae*, since some of these repeats have been shown to affect bacterial virulence. Various di- to hexanucleotide repeats were identified, confirming and extending previous findings on the existence of variable-number-of-tandem-repeat loci (VNTRs). Repeats with units of 7 or 8 nucleotides were not encountered. For all of the 3- to 6-nucleotide repeats in the *H. influenzae* chromosome, PCR tests capable of detecting allelic polymorphisms were designed. Fourteen of 18 of the potential VNTRs were indeed highly polymorphic when different strains were screened. Two of the potential VNTRs appeared to be short and homogeneous in length; another one may be specific for the *H. influenzae* Rd strain only. One of the primer sets generated fingerprint-type DNA banding patterns. The various repeat types differed with respect to intrinsic stability as well. It was noted for separate colonies derived from a single clinical specimen or strains passaged for several weeks on chocolate agar plates that the lengths of the VNTRs did not change. When several strains from different patients infected during an outbreak of lung disease were analyzed, increased but limited variation was encountered in all VNTR sites analyzed. One of the 5-nucleotide VNTRs proved to be hyper-variable. This variability may reflect the molecular basis of a mechanism used by *H. influenzae* bacteria to successfully colonize and infect different human individuals.

Short tandemly repeated sequences occur in several to thousands of copies dispersed through the genomes of many, if not all, higher eukaryotes. These sequence elements show hyper-variability among human individuals, and individual-specific DNA fingerprints can be generated by visualizing these elements with the help of molecular techniques. These molecular loci themselves became known as variable-number-of-tandem-repeat regions (VNTRs) and are now well-established targets for pedigree analysis (16). Characteristic prokaryotic repeats such as the enterobacterial repetitive intergenic consensus (ERIC) sequences and the repetitive extragenic palindrome sequence motif have been found in microbial species as diverse as *Enterobacteriaceae* and cyanobacteria (2, 42). All of these DNA elements, also including the pneumococcal BOX repeat (24), are not organized in tandem generally but have been used for molecular typing purposes as successfully as the VNTRs mentioned above (19, 38). More simple VNTR-like bacterial sequence motifs have been detected in *Mycobacterium tuberculosis* by hybridization with a probe consisting of five GTG units (41, 44). In other bacterial species, including *Staphylococcus aureus*, many of the genes encoding membrane-bound proteins have been shown to contain contiguous repetitive DNA. The staphylococcal coagulase and protein A genes are clearly polymorphic in their repetitive regions (11–13). Relatively little is known about the presence, putative function, and possible (high-speed?) evolution of both the repeat sequence and unit number of VNTR-like DNA in prokaryotes, but one of the exceptions is the well-documented repeat variability related to phenotypic switching in *Haemophilus influenzae* (43).

Within the *lic1* gene of *H. influenzae*, variable numbers of the tetranucleotide unit 5′-CAAT-3′ were detected. Variation in the overall number of units moves upstream-located ATG codons in or out of frame, which affects protein synthesis and structure. Additional switching signals involving CAAT repeats were identified in the genes *lic2* and *lic3*. The *lic2* and *lic3* gene products are also involved in lipopolysaccharide (LPS) biosynthesis (14, 43). Another *H. influenzae* VNTR locus is found between the genes *hifA* and *hifB*, which encode fimbrial subunits (39). Reversible phase variation is enabled in this particular case by changes in the number of TA repeats which space the −35 and −10 boxes of the dual promoter controlling *hifA* and *hifB*. Interestingly, detailed genetic analysis of the entire fimbrial gene cluster also revealed tandemly repeated repetitive extragenic palindrome units (40). The basic link between VNTR polymorphism and expression of pathogenesis-related genes as described above for *H. influenzae* was identified in *Neisseria meningitidis* and *Neisseria gonorrhoeae* as well (34). It has been postulated that this efficient form of adaptive evolution, especially in those genes encoding bacterial factors that interact with an "unpredictable environment," may be highly beneficial to bacterial populations (25, 26). The existence of slipped-strand mispairing as a basic molecular mechanism of DNA variation in VNTRs is generally accepted (21, 43).

The availability of the full genome sequence of *H. influenzae* (9) enables computerized analyses of the presence of repetitive DNA with VNTR-like potential (18). Recent molecular analysis of only the tetranucleotide VNTRs, revealed the involvement of repeat number variation not only in the *lic* genes but also in repeats located in some putatively iron-regulated genes (15). Apparently, *H. influenzae* is capable of translating VNTR variability into adaptive virulence. We here demonstrate VNTR variability in a large number of clinical isolates of *H. influenzae*. It is shown that in addition to the tetranucleotide VNTRs,

---

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

other repeat species, consisting of 3 to 6 nucleotide units, can also be variable in length. Moreover, we here demonstrate the variation of repeats during growth of *H. influenzae* in different human hosts. The possible implications of these findings for the study of bacterial genome evolution and pathogenesis with the help of PCR tests aimed at VNTR loci are discussed.

MATERIALS AND METHODS

Bacterial strains. *H. influenzae* strains were obtained from different sources. Ten invasive strains, isolated from the blood of 10 different, unrelated patients, were selected at random from the collection of bacterial blood isolates present in the Department of Bacteriology, University Hospital Rotterdam, Rotterdam, The Netherlands. Strains were isolated from patients in 10 different departments at different times (at least 1-month intervals between isolation dates). For reproducibility studies, two laboratory strains of *H. influenzae* (strains 714 and 965) were kept on solid chocolate medium for 16 consecutive days. Every day a fresh plate was inoculated in a serial fashion. Ten colonies were selected from each primary culture plate, which was inoculated directly from an archival glycerol stock. These 20 colonies were cultured as described above, and samples were taken every other day and stored at −80°C in glycerol-containing medium. In addition, 10 phenotypically characteristic *H. influenzae* colonies were picked from primary cultures of sputum samples of four different patients nursed in different parts of the hospital. Separate colonies taken from the primary, diagnostic chocolate agar plate (Oxoid, Mannheim, Germany) were cultured only once. Twenty strains isolated from different patients during an outbreak in a geriatric nursing home were also analyzed. These strains were analyzed before by genotypic and phenotypic procedures (35, 36). Seven strains chosen at random from the collection of the Microbiology Department of the Academic Medical Center (AMC) (Amsterdam, The Netherlands) were added to the set. Strains were coded so that analyses were performed in a blinded fashion. Single strains of the following species were included for species specificity studies: *Haemophilus segnis*, *Haemophilus parainfluenzae*, *Haemophilus haemolyticus*, *Haemophilus parahaemolyticus*, *Haemophilus aegyptus*, *Haemophilus aphrophilus*, and *Haemophilus haemoglobinophilus*. Table 1 gives some additional characteristics of the strains analyzed in the present study. Prior to DNA isolation, all strains were cultured overnight on chocolate agar plates (Sanofi Pasteur, Marnes-la-Coquette, France) in an atmosphere containing 5% CO_2.

Computer-aided VNTR searches. The whole genome sequence as determined for *H. influenzae* (GenBank accession number L42023) was screened with a newly developed algorithm. The program was written in the C language to scan for short (2- to 8-nucleotide) repeated sequences. To be reported, the minimum lengths of the sequences built of shorter repeats were as follows: 2-mer, 10 bases; 3-mer, 18 bases; 4-mer, 20 bases; 5-mer, 18 bases; 6-mer, 20 bases; 7-mer, 21 bases; and 8-mer, 24 bases. The minimum lengths were chosen empirically in order to avoid large numbers of extremely short repeats to be identified. Repeats built of multimers of shorter repeats were reported as the shorter repeat only. The program does not report repeats that are less than 30 nucleotides from a sequence terminus. Complete genome sequences can be scanned, but also specific collections of separate GenBank entries can be analyzed in a single run. The program is available through the World Wide Web (HTTP://ALCES.MED .UMN.EDU/webrepeat.htlm).

DNA purification. After overnight growth on chocolate agar plates, bacterial cells were collected and immediately dissolved in a 4 M guanidinium isothiocyanate-containing lysis buffer (2). DNA was purified by affinity chromatography with Celite (Sigma, Bornem, Belgium), and the purified DNA was stored at −20°C at a concentration of 50 ng/μl.

PCR analysis and gel electrophoresis. PCR was performed in 50-μl incubation volumes in Biomed (Theres, Germany) PCR machines. Reaction mixtures were composed as described earlier, with all ingredients matching the optimal reaction parameters for the Super Taq DNA polymerase (Sphaero Q, Leiden, The Netherlands) (36, 37). Twenty-nucleotide PCR primers were designed on the basis of sequences bordering the repetitive DNA (see Results and Table 2), and approximately 25 ng of template DNA was included per assay. A general program consisting of 40 cycles of 1 min at 94°C, 1 min at 52°C, and 1 min at 74°C was used for amplification. Prior to cycling, a 4-min predenaturation step at 94°C was included. Post-PCR gel electrophoresis was carried out in 3% Genetic Technology Grade agarose or 2 to 4% Metaphor agarose (FMC Bioproducts, Maastricht, The Netherlands). Gels were run in 0.5× Tris-borate-EDTA buffer at a constant voltage of 100 V (31) and photographed with a charged-coupled device camera equipped with a Fujinon zoom lens. Data were collected and thermoprinted by using the Visionary Photo Analyst system (FotoDyne; Progress Control, Waalwijk, The Netherlands). The lengths of the PCR products were determined by comparison with small-molecular-size markers (10-bp ladder; Gibco BRL Life Sciences, Breda, The Netherlands).

RAPD analysis. Overall genetic homo- or heterogeneity for the *H. influenzae* strains was determined by random amplification of polymorphic DNA (RAPD) as described before (36, 37). For the generation of the amplimer patterns, a combination of the primers ERIC1 and ERIC2 (42), in a single incubation mixture, was employed. Amplification products were analyzed by electrophoresis on 1.5% agarose gels (Hispanagar; Sphaero Q) and documented by Polaroid

photography. Single-band differences were considered decisive for the definition of separate genotypes.

RESULTS

Tracking potential VNTRs. The result of the computer-aided searches are summarized in Table 2. All 23 potential VNTR loci comprising repeat units ranging from 2 to 6 bases in length were given a recognizable code (Hi 2-1 through Hi 6-3) and are defined by the position in the *H. influenzae* genome, the unit sequence, and the number of units in the *H. influenzae* Rd strain (9). The results given in Table 2 are in agreement with those recently reported for the tetranucleotide repeats (15). In the case of the *H. influenzae* genome, five dinucleotide repeats were identified. None of these repeats exceeded a length of five sequence units, and three of them were inside an open reading frame. The TA repeat present between the *hifA* and *hifB* genes in *H. influenzae* AM20 and AM30 (39) is not present in the genome of the fimbria-deficient Rd strain (9). A single, relatively long extragenic 3-nucleotide repeat was identified, as were two and three 5- and 6-nucleotide repeats, respectively. Seven- or 8-nucleotide repeats were not identified.

The majority of potential VNTRs comprised repeats with a 4-nucleotide-long unit. Twelve loci were identified, varying from 6 to 32 units in overall length. The unit sequences appeared to be quite diverse, although a single motif (TTGG) occurred five times (Hi 4-4, Hi 4-5, Hi 4-7, Hi 4-9, and Hi 4-12) (Table 2). In all of these cases, except for Hi 4-9, the sequences immediately bordering the repeats show homology (Table 2). This may be an indication of locus multiplication (15).

Repeat variability. Primers for tracking repeat variability were designed for all of the potential VNTRs of *H. influenzae* (Table 2). The 20-nucleotide primers were selected on the basis of positional criteria only, their location being 5 nucleotides upstream and downstream of potential VNTRs. Only in the case of repeat locus Hi 6-1 was additional spacing introduced, due to the presence of single, noncontiguous copies of the unit sequence in the region surrounding the actual VNTR (data not shown). Our main research focus was with the tetranucleotide repeats, because it is the largest class of repeat types and because there is a link between repeat variability and pathogenicity (14, 15, 25, 26, 43).

All strains subjected to VNTR analyses of diverse kinds were characterized by RAPD. Table 1 shows the overall results, indicating the unique character of many strains but also the close genetic relatedness (if not identity) of the outbreak-related *H. influenzae* strains. When combined with DNAs from 10 independent isolates of *H. influenzae*, different DNA products were synthesized by using the primers designed for the amplification of the tetranucleotide repeats (Fig. 1). The average lengths of the repeats and the observed variation in length are given in Table 3. The variability in the length of the repeats is quite high for this class of repeats. Sometimes multiple DNA fragments were produced upon amplification of a single DNA template. For the Hi 4-5 VNTR, for instance, two widely differing amplicons were generated upon PCR; the two fragments differed by 28 repeat units in length ($n = 2$ and $n = 30$ [Table 3]). The precise nature of these fragments needs to be determined, although VNTR duplication or heterogeneity of a bacterial population are possible explanations for this phenomenon. The primer combination for amplification of Hi 4-7 produced no DNA product at an annealing temperature of neither 42 or 52°C. This is probably due to an absence of primer binding sites in the templates. Lowering the annealing temperature had no effect. Interestingly, the primer sequences

TABLE 1. *H. influenzae* strains used for VNTR analysis

Strain(s)	Source			MOMP type[b]	RAPD type[c]
	Location[a]	Material	Date isolated (day-mo-yr)		
Unrelated clinical isolates from Rotterdam					
1	IC1	Blood	4-5-95	ND[d]	I
2	IC4	Blood	3-5-95	ND	II
3	NQAS	Blood	3-4-95	ND	III
4	IC1	CSF[e]	22-10-94	ND	IV
5	Hema	Blood	26-12-94	ND	V
6	CF	Sputum	4-1-95	ND	VI
7	Bact	Nasal swab	16-1-95	ND	VII
8	Int Med 1	Sputum	24-1-95	ND	VIII
9	Gen Surg	Sputum	24-1-95	ND	IX
10	CF	Sputum	30-1-95	ND	X
Primary sputum isolates from Rotterdam					
Patient I, A to J	CF	Sputum	5-5-95	ND	I
Patient II					
A to D	CF	Sputum	5-5-95	ND	II
E	CF	Sputum	5-5-95	ND	III
F to J	CF	Sputum	5-5-95	ND	II
Patient III, A to J	CF	Sputum	5-5-95	ND	IV
Patient IV					
A to F	CF	Sputum	5-5-95	ND	V
G	CF	Sputum	5-5-95	ND	VI
H to J	CF	Sputum	5-5-95	ND	V
Laboratory strains used for passaging (Rotterdam)[f]					
714	Int Med 2	Blood	11-6-96	ND	I
965	IC1	CSF	23-7-96	ND	II
Outbreak-related strains from Amsterdam					
A880021 to -29	Nur	Sputum	1989	A	II
A880031 to -34	Nur	Sputum	1989	A	II
Incidental isolates from Amsterdam[g]					
A880035	Unknown	Sputum	Unknown	B	VII
A880036	Unknown	Sputum	Unknown	C	I
A880037	Unknown	Sputum	Unknown	D	VIII
A880038	Unknown	Sputum	Unknown	E	IX
A880039	Unknown	Sputum	Unknown	F	X
A880040	Unknown	Sputum	Unknown	G	XI
A880041	Unknown	Sputum	Unknown	H	XII

[a] IC, intensive care unit; NQAS, National Quality Assurance System (control strain); Hema, hematology; CF, clinic for cystic fibrosis patients; Bact, member of the personnel of the Bacteriology Department; Int Med, internal medicine; Gen Surg, general surgery; Nur, nursing home.

[b] MOMP typing was performed for the Amsterdam strains only.

[c] The RAPD types were determined per group of strains; type I in one group is different from type I in another group. There was no overlap between groups with respect to the RAPD-determined genotypes. The strains from Amsterdam were analyzed in a blinded and randomized order.

[d] ND, not done.

[e] CSF, cerebrospinal fluid.

[f] Strains 714 and 965 were retrieved from glycerol stocks and were subsequently passaged by serial transfer for 16 consecutive days. All isolates analyzed appeared to be of the indicated RAPD type.

[g] The missing RAPD types (III, IV, V, and VI) for this group of strains were assigned to control strains from Rotterdam that were included when RAPD was performed for this group of isolates. Additional information on both the strains from Amsterdam and the isolates of the non-*H. influenzae Haemophilus* spp. can be found in references 35 and 36.

bordering Hi 4-7 showed a limited set of point mutations compared to the sequences bordering Hi 4-4 or Hi 4-5 (Table 2). None of the primer combinations gave rise to identical amplicon patterns. Table 3 shows that changes in the size of a given repeat did not coincide with similar changes in other repeats. This demonstrates independent evolution of repeats present within a single genomic molecule. The identity of the 5′ primer for repeats Hi 4-4 and Hi 4-5 combined with different 3′ primers (Table 2) points to earlier crossover events. The 3′ primer for the Hi 4-12 repeat displays sequence homology with the 5′ primers for Hi 4-4 and Hi 4-5 as well. One of the primer combinations gave rise to a single size of fragment. Interestingly, this lack of variability coincided with the shortest repeat length ($n = 6$) found in the fully sequenced Rd type strain (Hi 4-6) (Table 2). All other VNTR regions showed amplification fragment length variability. Five of the VNTR analyses pro-

TABLE 2. Characteristics of potential VNTR sequences in the genome of *H. influenzae* Rd

Repeat unit length (nucleotides)[a]	Repeat code	Repeat position[b]	Unit sequence	Unit no.	Gene product homolog[c]	Genus[c]	Function[c]	Primers for VNTR amplification[d] 5'	3'
2	Hi 2-1	9612–9622	AT	5	NA[e]	NA	NA	None	None
	Hi 2-2	514282–514292	GC	5	NA	NA	NA	None	None
	Hi 2-3	789050–789060	AC	5	NA	NA	NA	None	None
	Hi 2-4	798986–798996	TG	5	NA	NA	NA	None	None
	Hi 2-5	1105917–1105927	TC	5	NA	NA	NA	None	None
3	Hi 3-1	291617–291644	ATT	9	NA	NA	NA	CAAATGATTATAAATAAACC	TAATTAAAAAGAGGAGGAATG
4	Hi 4-1	288751–288839	GTCT	22	LgtC	*Neisseria*	Glycosyltransferase	TCATTGTCTGACTGACAGTC	CGTTAATCAGAGATAATTTC
	Hi 4-2	379523–379651	CAAT	32	Lic3	*Haemophilus*	LPS biosynthesis	ATTACCTGCAATAATGACAG	TATTCAATGAACGGTAGAAT
	Hi 4-3	570800–570892	CAAT	23	Lic2	*Haemophilus*	LPS biosynthesis	CCTCTTATATTATGTAATAT	TTTAGTTTCTTTAATGCGTA
	Hi 4-4	677133–677217	TTGG	21	Hemoglobin receptor	*Neisseria*	Iron binding	CTAGTTGTTCAGAAACATTA	TAAATGCAAGCATAGCCTAT
	Hi 4-5	705897–705979	TTGG	20	Hemoglobin receptor	*Neisseria*	Iron binding	CTAGTTGTTCAGAAACATTA	GGCAGGTGTTGCTTATGCAG
	Hi 4-6	731230–731254	TTTA	6	32.9-kDa protein	*Bacillus*	Unknown	ATAATTGGTGAACCTAAAAT	TTTCCAGTCATAAAAATACC
	Hi 4-7	760526–760674	TTGG	37	Hemoglobin receptor	*Neisseria*	Iron binding	CTAATTGTTCAGAAACATTA	TAAATGCAAGCA**CA**GTCTAT
	Hi 4-8	1122922–1123002	TGAC	20	Methyltransferase	*Neisseria*	Restriction modification	TTTTCGGGTTAATTTGGTGT	AATTTATTTTAAACGCAATA
	Hi 4-9	1481220–1481284	TTGG	16	No homology	*Bacillus*		AAAATGAAAAGGATCTATAC	ACTACCGCAACGGTTTTATT
	Hi 4-10	1543152–1543252	TTGC	25	YadA	*Yersinia*	Adhesin	GACAGATGAAAAGAAAAGAT	TATAATATGTTTTATTACAA
	Hi 4-11	1608031–1608099	CAAT	17	Lic1	*Haemophilus*	LPS biosynthesis	TAAAAATGAATACAAAAATG	AAGTTTTAACAAATCCTACA
	Hi 4-12	1633205–1633281	TTGG	19	Hemoglobin receptor	*Neisseria*	Iron binding	AACGGCAAGTGTTGCTTATG	CTAGTTGTTCAGAA**GC**ATTA
5	Hi 5-1	922080–922139	TTATC	12	NA	NA	NA	ACAGACAGGCTATGGCAATG	TATAAAGCACTAGAACAATG
	Hi 5-2	1368890–1368910	GTCTC	4	NA	NA	NA	GTGATTTTTATCGACAATCT	TACAGAGGGCATAATTTATG
6	Hi 6-1	283097–283115	CTGGCT	4	NA	NA	NA	TCTACAATTTCTTGTTTTTC	ATGGTGTTGGAAGAACCTGC
	Hi 6-2	296053–296071	GGCAAT	3	NA	NA	NA	AGATTTAGAGAGAATCAGTG	CGTCTTTTAGTTTACGGGTA
	Hi 6-3	548966–548989	TTAAAA	3	NA	NA	NA	TGACATAAATCTATCCTCTTG	TAGGTATAATACGAAAAGTT

[a] Repeats with a unit length of 7 or 8 nucleotides were not found.
[b] Repeat positions are identified on the basis of sequences presented in reference 15.
[c] Data are derived from reference 9.
[d] Primer sequences are given from the 5' to the 3' end. The nucleotides shown in boldface are those in the Hi 4-7 and Hi 4-12 sets differing from the homologous ones in the Hi 4-4 and the Hi 4-5 sets. Underlined primer sequences are homologous and may define repeat locus multiplication (see text for details). Primers for the Hi 6-1 repeat are located more distantly from the core units (see Results).
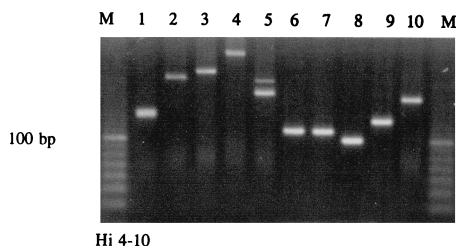[e] NA, not applicable.

FIG. 1. Polymorphisms in the Hi 4-10 VNTR as determined for 10 unrelated strains of *H. influenzae*. Lane numbers correspond to the strain numbering as used in Table 3. Lanes M display 10-nucleotide ladder patterns; the intensely staining bands represent a 100-nucleotide fragment (Boehringer, Mannheim, Germany). Note that only strains 6 and 7 are identical with respect to this specific repeat (16 repeat units [Table 3]).

vided clear length polymorphisms (Hi 4-3, Hi 4-5, Hi 4-10, Hi 4-11, and Hi 4-12), and these PCRs were studied further for species specificity and evolutionary concordance.

Figure 2 displays the data obtained when DNAs from the same, nonrelated strains were used for amplification with the primers specific for the repeats harboring units of 3, 5, or 6 nucleotides. A degree of variability similar to that observed with the tetranucleotide motifs was noted, although locus Hi 6-3 appeared to be monomorphic. With the primers designed for the amplification of Hi 5-1, RAPD-type DNA fingerprints were obtained.

Species specificity. When Hi 4-3, Hi 4-5, Hi 4-10, Hi 4-11, and Hi 4-12 were amplified for the *Haemophilus* species other than *H. influenzae*, no DNA fragment was synthesized, indicating the species specificity of the primer sites. For the other motifs some cross-reactivity was observed. Hi 3-1, Hi 6-1, and Hi 6-2 amplification led to the synthesis of single, identically sized fragments for all of the different species in the panel. Hi 5-2 amplification detected limited polymorphism among the different species.

Stability of repeat numbers. A selection of the tetranucleotide VNTRs (Hi 4-3, 4-5, 4-10, 4-11, and 4-12) was amplified by using DNAs isolated from the serial samples obtained during long-term cultivation of laboratory strains 714 and 965. Analysis of DNA from primary colonies derived directly from the glycerol stocks revealed that these archival stocks contained cells harboring identical repeats. Moreover, repeat lengths measured at day 1 of cultivation were identical to those measured after 16 days of continued subcultivation. This shows that routine cultivation did not produce detectable repeat variation.

Repeats Hi 4-3, 4-5, 4-10, 4-11, and 4-12 were also amplified by using as templates DNAs isolated from 10 isolates from each of four different patients (Table 1, strains A to J for patients I to IV). An overview of repeat length polymorphisms is provided in Table 4, whereas an example of the experimental data is given in Fig. 3. From the RAPD analyses it became clear that patients II and IV carried mixtures of strains. Strain E from patient II showed a different RAPD pattern but appeared to be similar to the strains from the same patient when the VNTRs were considered. Most of the VNTRs were genetically stable. Reproducible multiplicity of some of the VNTR alleles was detected (Hi 4-5 for patients I and II and Hi 4-12 for patient I). Overall, most of the strains derived from colonies from the primary culture from individual patients showed identical numbers of distinct VNTRs. When the VNTRs consisting of 3-, 5-, or 6-nucleotide units were amplified for a selection of the same strains as used in the experiments described above, identical conclusions could be drawn. Length conservation was documented for Hi 3-1, Hi 5-2, Hi 6-1, and Hi 6-3, indicating short-term genetic stability.

Evolutionary concordance. Epidemiologically related strains were studied with RAPD and major outer membrane protein (MOMP) profiling (35, 36). MOMP and RAPD data confirm the genetic relatedness of the first 13, epidemiologically related isolates (Tables 1 and 5) (A880021 through A880034, RAPD type II). All of the other strains could be discriminated, although the differences as measured with RAPD are restricted to one band only (Fig. 4). Hi 3-1, Hi 4-3, Hi 4-5, Hi 4-10, Hi 4-11, Hi 4-12, Hi 5-2, Hi 6-1, and Hi 6-2 were amplified by using as templates DNAs isolated from the outbreak strains. An overview of repeat length polymorphisms for the tetranucleotide repeats is provided in Table 5. Compared to the data presented in Fig. 3 and Table 4, increased heterogeneity was identified among the VNTRs. Multiple bands were observed with Hi 4-12. It is reassuring that when the average repeat lengths are calculated (Table 5), epidemic strains appear as a homogeneous collection. The length variation was relatively small (maximum of 12.0% of the average), whereas for the nonrelated strains this value is much greater (minimum value of 42.2%). Consequently, a set of strains that are genetically closely related could be defined as a group showing a

TABLE 3. Repeat polymorphisms in individual tetranucleotide VNTR loci determined for 10 unrelated strains of *H. influenzae*[a]

Repeat code	No. of repeat units for strain:										Avg no. of units ± SD
	1	2	3	4	5	6	7	8	9	10	
Hi 4-1	10	38	37	18	28	24, 30, 36	26	24, 30, 36	26	7	26 ± 10
Hi 4-2	21, 36	28	18	22, 26	18	15	12, 30	20, 22	24	17, 21	22 ± 6
Hi 4-3	27	17	17	12	24, 26	21	23	20	14	36	22 ± 7
Hi 4-4	20	20	34	43	22	20	39	22	23	26	27 ± 9
Hi 4-5	10	2, 30	38	8	37	MB[b]	40	44	10	MB	24 ± 17
Hi 4-6	7	7	7	7	7	7	7	7	7	7	7
Hi 4-8	17	3	2	12	3	3	20	3	30	26	12 ± 11
Hi 4-9	9	6, 7	7	7, 14	7, 14	7, 14	7, 14	14	6	14	10 ± 4
Hi 4-10	24	43	50	>60	33, 40	16	16	13	21	31	32 ± 15
Hi 4-11	30	32	19	18	19	7	22	3	12	3	17 ± 10
Hi 4-12	19	32	36	8	34	25	40	8	19	16	24 ± 11

[a] Clinical data and origins of the bacterial isolates are given in Table 1. The assay involving repeat Hi 4-7 (see Table 1) was unsuccessful in that no DNA fragments were synthesized. In some cases, of multiple DNA fragments were present. Amplification of Hi 4-4 and Hi 4-5 gave rise to coamplification of small amounts of several other products. This also occurred with amplification of Hi 4-8. In this case, high-molecular-weight amplicons were synthesized. See Fig. 1 for an example of the experimental results.
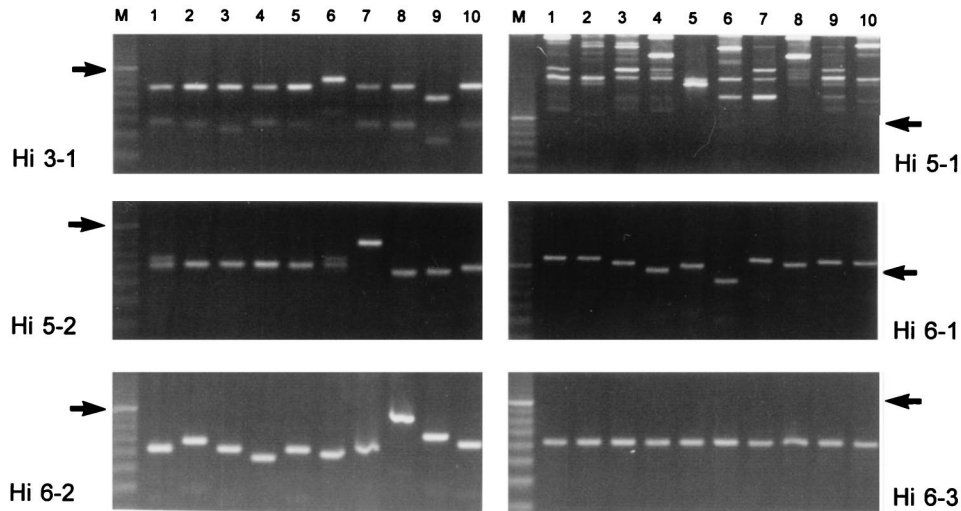[b] MB, multiple (more than three) bands.

FIG. 2. Polymorphisms in the Hi 3-1, 5-1, 5-2, 6-1, 6-2, and 6-3 VNTRs as determined for 10 unrelated strains of *H. influenzae*. Lane numbers correspond to the strain numbering as used in Table 3. Lanes M display 10-nucleotide ladder patterns (Boehringer); the intensely staining bands represent a 100-nucleotide fragment (arrows).

maximum of 10% length variability, when measured as the average variation in at least five VNTR loci. However, more data on the response to environmental conditions of the VNTRs should be available before definite criteria can be defined. It is still possible that strains change their VNTR profiles drastically under unfavorable conditions (25).

Figure 5 summarizes the experimental data obtained for repeats consisting of 3, 5, or 6 nucleotide units. Three of four loci seem to behave as stable genetic markers: among the epidemic isolates, Hi 3-1, Hi 6-1, and Hi 6-2 are identical. However, several of the nonrelated strains cannot be discriminated from the outbreak cluster as well. For Hi 5-2, five different amplicon lengths can be determined among the epidemic isolates, indicating the hypervariable nature of this region.

In order to analyze the data for the tetranucleotide repeats in more mathematical detail, we devised a program to produce a distance matrix for the data shown in Table 5. For reasons of simplicity, the program uses the largest size in case of multiple bands. When differences in allele size are taken into account, the distance matrix appears as shown in Table 6. Comparison of any of the strains with all of the other strains in the collection shows that the genetically related, MOMP- and RAPD-identical strains are closest with respect to the calculated distance: the maximum score is 18 (strain 3 versus 4). The minimal score among the incidental isolates is 19 (strain 4 versus 17).

DISCUSSION

Occurrence of VNTRs in bacterial DNA. Bacterial genomes may contain repetitive DNA moieties which show significant homology, both in nature and behavior, to the VNTR-type DNA stretches encountered in the genomes of lower and higher eukaryotes (1). A systematic, computerized search for the presence of longer arrays of short repeat units in the genomes of simple organisms was performed recently (8). In DNA sequences of slime molds, fungi, protists, prokaryotes, viruses, intracellular organelles, and plasmids, 78 putative sites were discovered. In addition to the *H. influenzae* repeats, of these 78 repetitive DNA domains only 3 originated from prokaryotes. A $(CTA)_9$ domain was discovered in the genome of *Neisseria meningitidis*, whereas $(CTA)_{11}$ and $(CTT)_{21}$ were dis-

covered in sequences derived from *Mycoplasma genitalium* and *Mycobacterium leprae*, respectively. It was suggested, but not experimentally demonstrated, that these bacterial loci may be as prone to genetic variability as those encountered in eukaryotes. We here present evidence that in the case of *H. influenzae*, a large fraction of VNTR regions comprising 3- to 6-nucleotide repeat units, as identified here and previously (15, 18), are highly polymorphic in nature. Sequence determinations revealed that in the case of variability in the Hi 5-2 locus, multiplication or deletion of GTCTC units could be demonstrated. Also for the other VNTRs, the sequence motifs were shown to be identical to those previously reported (15). On the other hand, cloning VNTR PCR products regularly resulted in the acquisition of clones with the insert being shorter than the original PCR product (results not shown). The straightforward 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.

TABLE 4. Typing of primary isolates of *H. influenzae* from individual patients[a]

Patient	Strain(s)	RAPD type	No. of repeats				
			Hi 4-3	Hi 4-5	Hi 4-10	Hi 4-11	Hi 4-12
I	A to J	I	23	18, 22, 23	13	22	19, 23, 24
II	A to D	II	19	28, 29	14	33	28
	E	**III**	19	28, 29	14	33	28
	F to J	II	19	28, 29	14	33	28
III	A to J	IV	29	29	11	33	28
IV	A to F	V	23	44	12	18	44
	G	**VI**	17	44	2	10	19
	H to J	V	23	44	12	18	16

[a] Boldface identifies cases in which RAPD typing does not correspond to VNTR typing. Note that strain lettering (A through J) corresponds to different **colonies** selected from primary culture plates, corresponding to the identification as shown in Fig. 3.
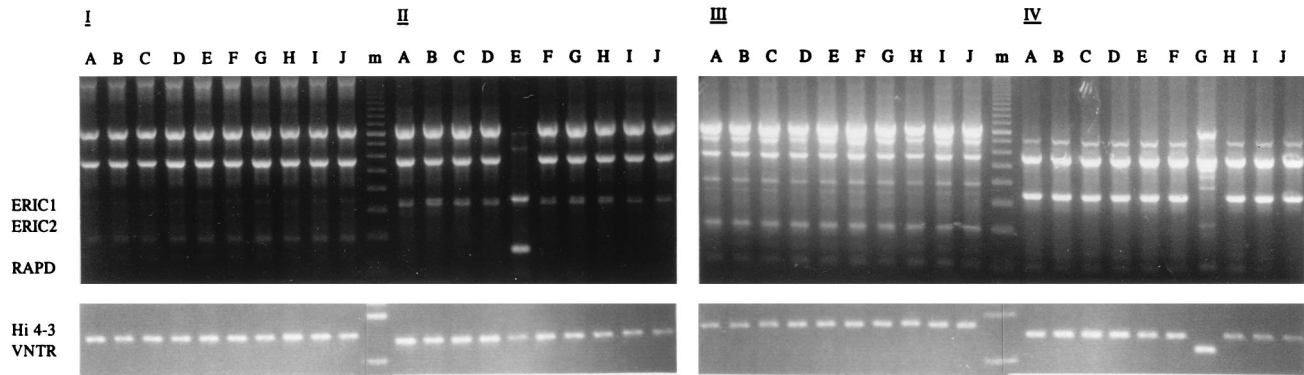
FIG. 3. Genetic polymorphisms in strains of *H. influenzae* derived from individual patients (I through IV). The strains are designated A to J as in Table 4. The upper panel displays the results obtained with RAPD, showing homogeneity for the populations from patients I and III. Single aberrations are detected for patients II and IV. Note that the analysis of the Hi 4-3 VNTR confirms the RAPD data except for strain E from patient II. Lanes m display 10-nucleotide ladder patterns; the intensely staining bands represent a 100-nucleotide fragment (Boehringer). For determination of the sizes of the RAPD products, a 100-bp ladder (Pharmacia, Gouda, The Netherlands) was used.

Other (archaea)bacterial full genome sequences have been determined. The *M. genitalium* genome (10) contained a smaller number of potential VNTRs ($n = 11$) (results not shown) than *H. influenzae*, but the genome is also about half as long. Interestingly, in this case the trinucleotide repeats represented a clear majority. Five of the 11 repeats were built from the repeat unit TAG. As was the case for *H. influenzae*, virulence genes were identified at the sites of repetitive DNA (27). Single 5- and 6-nucleotide repeats were identified, but these were built from relatively low numbers of units. No repeats consisting of 7- or 8-nucleotide units were detected. Only short dinucleotide VNTR candidates were detected in the genome of the methanogenic archaeon *Methanococcus jannaschii* (4). All of these 35 loci, 2 of which were located on one of the two extrachromosomal plasmids, were short (≤ 6 dinucleotide units) and indicative of the nonuniversal character of the VNTR-type domains. Apparently, species exist that can survive without large numbers of variable repeats. Moreover, the species specificity studies we performed show that the tetranucleotide variability may be a peculiarity of *H. influenzae*, since the primers we used did not cross-hybridize with DNAs from closely related species such as *H. parainfluenzae*. The existence of potential VNTRs, however, has been shown for several *Neisseria* species, *H. parainfluenzae*, and *Moraxella catharralis* by means of hybridization analysis (26).

VNTR function. In eukaryotes, repetitive DNA may be involved in nucleosome organization, recombination, or regulation of gene expression or gene product activity (5). For several

TABLE 5. Typing of isolates of *H. influenzae* from patients involved in an outbreak of infection[a]

Strain origin	Patient	Strain	MOMP type	RAPD type	No. of repeats				
					Hi 4-3	Hi 4-5	Hi 4-10	Hi 4-11	Hi 4-12
Outbreak	1	A880021	A	II	23	21	19	43	23
	2	A880022	A	II	22	20	19	44	21
	3	A880023	A	II	20	18	18	44	19, 21
	4	A880024	A	II	22	26	18	40	16, 25
	5	A880025	A	II	21	20	18	44	21
	6	A880026	A	II	22	20	20	44	21
	7	A880027	A	II	21	22	18	44	16, 23
	8	A880028	A	II	23	23	19	44	21, 25
	9	A880029	A	II	22	20	19	44	21
	10	A880031	A	II	22	21	17	44	22, 23
	11	A880032	A	II	23	22	18	42	18, 23
	12	A880033	A	II	22	24	19	44	23, 26
	13	A880034	A	II	22	20	18	44	22, 23
Avg repeat length ± SD					**21.9 ± 0.9**	**21.3 ± 2.1**	**18.5 ± 0.8**	**43.5 ± 1.2**	**21.6 ± 2.6**
Incidental	14	A880035	B	VII	12	12	2	32	13
	15	A880036	C	I	14	8	19	20	9
	16	A880037	D	VIII	7	4, 20	2	11	5, 22, 29
	17	A880038	E	IX	25	18, 20, 30	14	38	31
	18	A880039	F	X	14	21, 25	9	30	23, 25
	19	A880040	G	XI	23	26, 30	2	>50	25, 30
	20	A880041	H	XII	8	17	22	20	17
Avg repeat length ± SD					**14.7 ± 6.9**	**19.3 ± 8.2**	**10.0 ± 8.5**	**28.7 ± 13.0**	**20.6 ± 8.7**

[a] RAPD types I and II are identical to those in Table 4. Not all the differences between repeat lengths among various strains are in multiples of three units. This implies that, potentially, reading frames may be interrupted. Strains 1 to 13 have been described in detail before (35, 36).
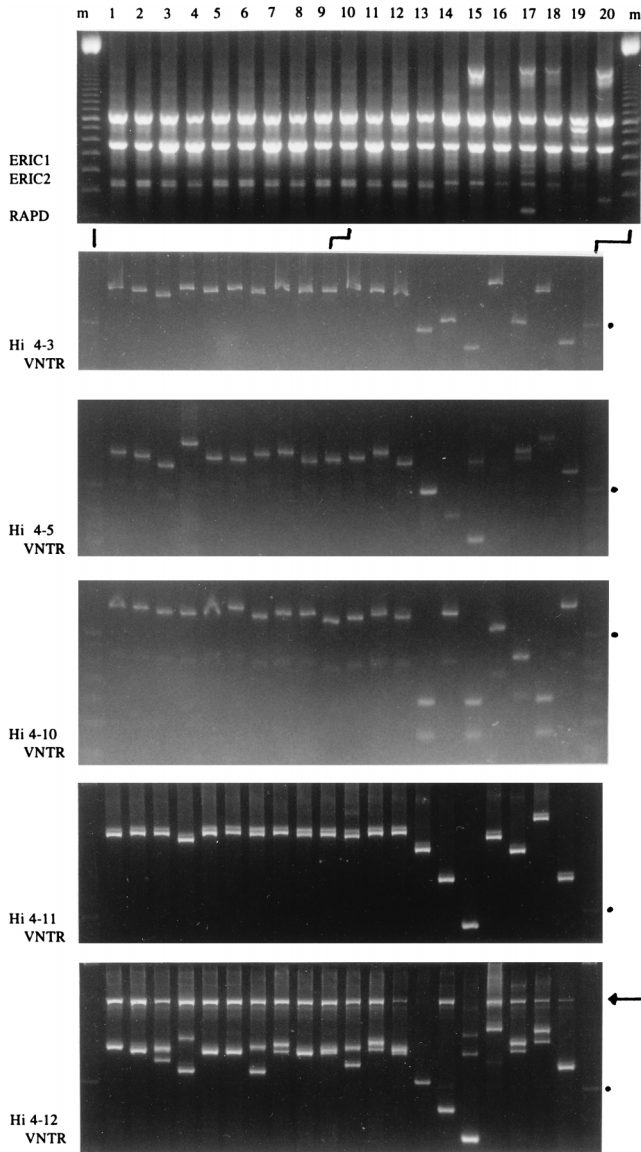
FIG. 4. Analysis of outbreak-related strains of *H. influenzae* by RAPD and VNTR sizing studies. The strains are identified as in Table 5. Strains 1 to 13 are epidemiologically linked; strains 14 to 20 are random isolates from the AMC collection (see Table 1 for additional information). Lanes m display 10-nucleotide ladder patterns in the VNTR panels; the intensely staining bands represent a 100-nucleotide fragment (black dots on the right) (Boehringer Mannheim). For determination of the sizes of the RAPD products, a 100-bp ladder (intense band at 800 bp) (Pharmacia) was used. The arrow on the right identifies a nonspecifically coamplified DNA fragment of unknown origin.

that repeat variability once more appeared to be intimately related to adaptive virulence (15). Overall, VNTR regions may be involved in the bacterial response to the physiological environment imposed by the host. The linking of repeat variation in the alpha C protein gene of streptococci and antibody-mediated killing provides another interesting example (23). Genes like these have been termed contingency genes, which show high mutation rates, allowing the bacterium to swiftly act in response to deleterious environmental conditions (25). In addition, model studies employing 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 (30). For *Enterococcus faecalis* it was recently demonstrated that phase variation of pAD1 conjugation functions relates to physical changes in so-called iterons, repetitive sequences as found in plasmids (3). Apparently, variation in these 8-bp repeat regions can lead to discrete physiological changes.

Multiplicity of VNTRs. Multiple bands in the PCR assays described in the present communication can be caused by erroneous amplification processes, such as mispriming, secondary priming, or the presence of multiple alleles of a given sequence domain. There are reasons that indicate locus multiplication rather than polymerase slippage. First, polymerase slippage is generally demonstrated by DNA ladder-type patterns. In our case, bands were discrete and differed by multiples of repeat units, not by single units. Second, individual colonies derived from single patients showed a consistent picture: when double bands were observed, all of the DNA extracts derived from individual colonies give rise to similar amplification patterns. Third, all of the experiments have been demonstrated to be reproducible, which was confirmed by multiplex PCRs (results not shown). The fact that strains that are epidemiologically related and consequently genetically closely linked differ in a relatively small number of repeat numbers (Table 4) underscores the likelihood of slipped-strand mispairing events (21, 39).

Microbial evolution. A recent report postulates that bacterial evolution may be not a continuous process but a succession
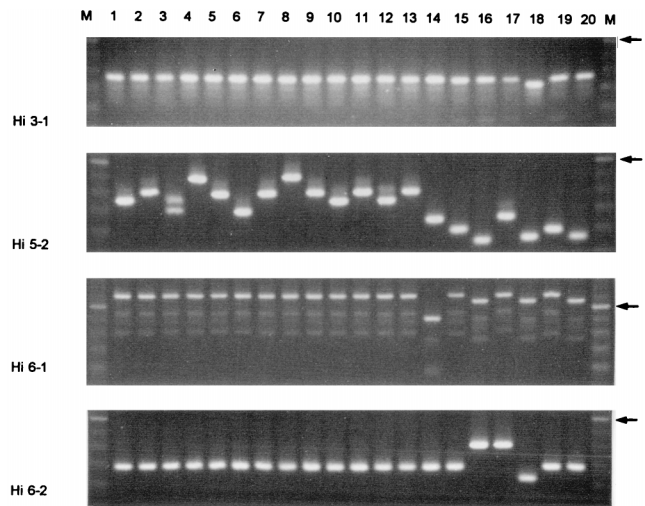

FIG. 5. Analysis of outbreak-related strains of *H. influenzae* by VNTR sizing studies. The strains are identified as in Table 5. Strains 1 to 13 are epidemiologically linked; strains 14 to 20 are random isolates from the AMC collection (Table 1). Lanes M display 10-nucleotide ladder patterns in the VNTR panels; the intensely staining bands represent a 100-nucleotide fragment (arrows on the right) (Boehringer).

species of prokaryotes, dinucleotide VNTRs were shown to be involved in, for instance, promoter activity modulation (39). Although several short dinucleotide repeats were identified (Table 2), it was decided not to analyze them. Their locations with respect to genes were not the same as for the fimbrial genes (39, 40); possible involvement in promoter spacing seems to be unlikely. Interestingly, all of the *H. influenzae* tetranucleotide repeats are within genes, and changing the repeat length has documented effects on virulence (25, 26, 43). The relationship between repeat variability and LPS phase variation has been elucidated previously (43). For a set of putatively iron-regulated genes, it was recently demonstrated

TABLE 6. Distance matrix calculated for the strains involved in the outbreak of *H. influenzae* infections in the geriatric nursing home and for nonrelated control strains[a]

Strain	Distance score with strain:																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	5	10	12	7	8	5	5	5	4	3	8	4	58	59	72	29	38	40	51
2		0	5	15	2	5	6	8	0	5	8	9	3	55	56	73	34	41	43	48
3			0	18	3	10	7	13	5	8	11	14	6	50	53	72	37	40	46	45
4				0	15	12	11	9	15	12	9	8	12	60	63	70	19	28	36	55
5					0	7	4	10	2	5	8	11	3	53	56	71	34	39	43	48
6						0	11	13	5	10	9	14	8	52	53	70	31	38	48	43
7							0	6	6	3	4	7	3	57	60	71	30	35	39	52
8								0	8	7	6	3	7	63	64	73	26	35	35	56
9									0	5	8	9	3	55	56	73	34	41	43	48
10										0	5	8	2	56	61	70	29	36	38	53
11											0	9	5	57	60	71	26	35	39	52
12												0	8	64	65	72	25	34	34	57
13													0	56	59	70	31	38	40	51
14														0	39	50	67	36	64	45
15															0	65	78	53	99	26
16																0	69	42	66	45
17																	0	35	27	70
18																		0	46	45
19																			0	91
20																				0

[a] Numbering of the strains is as in Fig. 4. Strains 1 to 13 are epidemiologically related; strains 14 to 20 are incidental isolates.

of temporally spaced major events (7). These events cause a nongradual sequence of adaptation to a given environment. The extreme diversity in repeat length and the diversity among strains with respect to the relative lengths of the individual repeat loci (Tables 3 and 4) may be indicative of another, somewhat more contiguous form of DNA variation. Different regions of the genome (coupled to a great diversity of functions) seem to evolve in different manners, undoubtedly dependent on sequence composition and DNA polymerase reliability. In this context, it is important to realize that several of the VNTR loci described in the present communication will be under environmentally induced selection. Repeats in two strains can be identical because the strains are identical by descent or because the strains are under the same selective pressure. Nothing is currently known about the validities of these alternatives, and it should be emphasized that especially the latter option may be a caveat for epidemiological studies focusing on the detection of VNTR variation. Furthermore, it is important to realize that VNTRs greatly differing in length can still give rise to the same phenotype. When a gene contains a tetranucleotide VNTR and is switched on when seven copies of the repeat unit are present, the gene will be expressed in a comparable fashion when $7 \pm 3n$ copies of the element are present. From a practical perspective, the *H. influenzae* VNTR regions are candidate targets for molecular identification and, consequently, subspecies identification of bacterial clones or strains. The tests described in the present communication present a valuable addition to the spectrum of techniques suited for characterization of the stability of the *H. influenzae* genome among related or unrelated clinical isolates (17, 20, 22, 32, 33, 36). Most promising is the possibility that specific PCR tests like the ones described here may prove to be reproducible in different labs, which would be an important advance beyond both RAPD and pulsed-field gel electrophoresis, two commonly applied identification techniques that have problems with intercenter reproducibility (6, 37).

Concluding remarks and future prospects. The type of analysis described in the present communication can be extended to any microorganism for which extensive nucleotide sequence data are available. The fact that the small genome of *M. genitalium* also appears to contain VNTR-type DNA suggests that for other prokaryotes suitable targets for a digital typing approach could be identified once larger proportions of their genomes are sequenced. As an interesting additional option, digital procedures for the study of pathogenicity-related genomic regions may in the future be applied to those microorganisms that are either fastidious or noncultivable. In view of the current effort aimed at determining the whole genome sequences of both *M. tuberculosis* and *M. leprae*, these two major bacterial pathogens may become amenable to this approach soon (28). Once the DNA sequences of the VNTR-bordering PCR primers have been determined and found to be species specific, both diagnostic information and data on the evolutionary status of the strains involved can be obtained by performing a single PCR test. The efficiency of this type of analysis can be greatly enhanced once multiplex PCR assays are designed (work in progress).

The *H. influenzae* infection process requires that the bacteria adapt to several different host environments. Initial colonization of the nasopharynx, crossing of epithelial and endothelial barriers, survival in circulation, and translocation across the brain-blood barrier are all processes requiring specific virulence traits (29). The variation of pathogenicity factors in high frequency in order to meet these requirements can possibly be achieved through (random?) VNTR modulation, as was described previously for several *H. influenzae* virulence factors (see reference 15 for a review). Studies in progress show that clonal strains isolated during persistent infection of cystic fibrosis patients show differing VNTR compositions (results not shown). Besides the variation in VNTR patterns observed among clonal strains isolated from different patients, VNTR variation possibly due to the changing (immunological or physiological) environment within a given individual can also be documented.

We here show the existence of very large differences in VNTR sizes when clinical strains of *H. influenzae* are com-

pared. This implies that the expression of multiple (virulence-related) genes is modulated at very high frequencies and on an independent, individual basis. This is particularly clear for the VNTRs analyzed for the strains involved in an outbreak of infections in a nursing home. In a constant genetic background, VNTRs provide the contingency loci (25) essential for a swift response to the environmental selection imposed by the individual patient. Elucidation of the complex virulence gene expression in a patient-dependent fashion and focusing on gene activation or inactivation by VNTR-induced changes may in the end significantly extend our understanding of the *H. influenzae* infection process.

REFERENCES

1. **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.
2. **Boom, R., C. J. A. Sol, M. M. M. Salimans, C. L. Jansen, P. M. E. Wertheim-van Dillen, and J. van der Noordaa.** 1990. Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. **28:**495–503.
3. **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.
4. **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.
5. **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.
6. **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.
7. **Elena, S. F., V. S. Cooper, and R. E. Lenski.** 1996. Punctuated evolution caused by selection of rare beneficial mutations. Science **272:**1802–1804.
8. **Field, D., and C. Wills.** 1996. Long, polymorphic microsatellites in simple organisms. Proc. R. Soc. London **263:**209–251.
9. **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. Weidmann, 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. Gnehm, 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.
10. **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.
11. **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.
12. **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 nonepidemic methicillin resistant *Staphylo-*

13. **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.
14. **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.
15. **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.
16. **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.
17. **Jordens, J. Z., N. I. Leaves, E. C. Anderson, and M. P. E. Slack.** 1993. Polymerase chain reaction-based strain characterization of nonencapsulate *Haemophilus influenzae.* J. Clin. Microbiol. **31:**2981–2987.
18. **Karlin, S., J. Mrazek, and A. M. Campbell.** 1996. Frequent oligonucleotides and peptides of the *Haemophilus influenzae* genome. Nucleic Acids Res. **24:**4263–4272.
19. **Koeuth, T., J. Versalovic, and J. R. Lupski.** 1995. Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. PCR Methods Appl. **5:**408–418.
20. **Landgraf, I. M., and M. F. P. Vieira.** 1993. Biotypes and serotypes of *Haemophilus influenzae* from patients with meningitis in the city of Sao Paulo, Brazil. J. Clin. Microbiol. **31:**743–745.
21. **Levinson, G., and G. A. Gutman.** 1987. Slipped-strand mispairing: mechanism for DNA sequence evolution. Mol. Biol. Evol. **4:**203–221.
22. **Levy, J., G. Verhaegen, P. de Mol, M. Couturier, D. Dekegel, and J. P. Butzler.** 1993. Molecular characterisation of resistance plasmids in epidemiologically unrelated strains of multiresistant *Haemophilus influenzae.* J. Infect. Dis. **168:**177–187.
23. **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.
24. **Martin, B., O. Humbert, M. Camara, E. Guuenzi, 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.
25. **Moxon, E. R., P. B. Rainey, M. A. Nowak, and R. E. Lenski.** 1995. Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr. Biol. **4:**24–33.
26. **Peak, I. R. A., M. P. Jennings, D. W. Hood, M. Bisercic, 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.
27. **Peterson, S. N., C. C. Bailey, J. S. Jensen, M. B. Borre, E. S. King, K. F. Bott, and C. A. Hutchinson 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.
28. **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.
29. **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.
30. **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.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Samuelson, A., A. Freijd, J. Jonasson, and A. A. Lindberg.** 1995. Turnover of nonencapsulated *Haemophilus influenzae* in the nasopharynges of otitis-prone children. J. Clin. Microbiol. **33:**2027–2031.
33. **Smith-Vaughan, H. C., K. S. Sriprakash, J. D. Matthews, and D. J. Kemp.** 1995. Long PCR-ribotyping of nontypeable *Haemophilus influenzae.* J. Clin. Microbiol. **33:**1192–1195.
34. **Stern, A., and T. F. Meyer.** 1987. Common mechanism controlling phase and antigenic variation in pathogenic neisseriae. Mol. Microbiol. **1:**5–12.
35. **Sturm, A. W., R. Mosteret, P. J. E. Rouing, B. van Klingeren, and L. van Alphen.** 1990. Outbreak of multiresistant nonencapsulated *Haemophilus influenzae* infections in a pulmonary rehabilitation center. Lancet **335:**214–216.
36. **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.

37. **Van Belkum, A., J. Kluytmans, W. van Leeuwen, R. Bax, W. Quint, E. Peters, A. Fluit, C. Vandenbroucke-Grauls, A. van en Brule, H. Koeleman, W. Melchers, J. Meis, A. Elaichouni, M. Vaneechoutte, 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.

38. **Van Belkum, A., M. Sluijter, 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.

39. **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.

40. **Van Ham, S. M., L. van Alphen, F. R. Mooi, and J. P. M. van Putten.** 1994.

41. **Van Soolingen, 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.

42. **Versalovic, J., T. Koeuth, 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.

43. **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.

44. **Wiid, I. J. F., C. Werely, N. Beyers, P. Donald, and P. D. van Helden.** 1994. Oligonucleotide $(GTG)_5$ as a marker for *Mycobacterium tuberculosis* strain identification. J. Clin. Microbiol. **32:**1318–1321.

The fimbrial gene cluster of *H. influenzae* type b. Mol. Microbiol. **13:**673–684.

Editor: J. G. Cannon