INFECTION AND IMMUNITY, Feb. 1998, p. 656–663 Vol. 66, No. 2
0019-9567/98/$04.00+0

Comparative Analysis of *Haemophilus influenzae hifA* (Pilin) Genes

DANIEL L. CLEMANS,[1]* CARL F. MARRS,[2] MAYURI PATEL,[1] MICHELLE DUNCAN,[1] AND JANET R. GILSDORF[1]

Department of Pediatrics and Communicable Diseases, University of Michigan Medical School,[1] and Department of Epidemiology, University of Michigan School of Public Health,[2] Ann Arbor, Michigan

Received 29 August 1997/Returned for modification 28 October 1997/Accepted 26 November 1997

Adherence of *Haemophilus influenzae* to epithelial cells plays a central role in colonization and is the first step in infection with this organism. Pili, which are large polymorphic surface proteins, have been shown to mediate the binding of *H. influenzae* to cells of the human respiratory tract. Earlier experiments have demonstrated that the major epitopes of *H. influenzae* pili are highly conformational and immunologically heterogenous; their subunit pilins are, however, immunologically homogenous. To define the extent of structural variation in pilins, which polymerize to form pili, the pilin genes (*hifA*) of 26 type a to f and 16 nontypeable strains of *H. influenzae* were amplified by PCR and subjected to restriction fragment length polymorphism (RFLP) analysis with *Alu*I and *Rsa*I. Six different RFLP patterns were identified. Four further RFLP patterns were identified from published *hifA* sequences from five nontypeable *H. influenzae* strains. Two patterns contained only nontypeable isolates; one of these contained *H. influenzae* biotype aegyptius strains F3031 and F3037. Another pattern contained predominantly *H. influenzae* type f strains. All other patterns were displayed by a variety of capsular and noncapsular types. Sequence analysis of selected *hifA* genes confirmed the 10 RFLP patterns and showed strong identity among representatives displaying the same RFLP patterns. In addition, the immunologic reactivity of pili with antipilus antisera correlated with the groupings of strains based on *hifA* RFLP patterns. Those strains that show greater reactivity with antiserum directed against *H. influenzae* type b strain M43 pili tend to fall into one RFLP pattern (pattern 3); while those strains that show equal or greater reactivity with antiserum directed against *H. influenzae* type b strain Eagan pili tend to fall in a different RFLP pattern (pattern 1). Sequence analysis of representative HifA pilins from typeable and nontypeable *H. influenzae* identified several highly conserved regions that play a role in bacterial pilus assembly and other regions with considerable amino acid heterogeneity. These regions of HifA amino acid sequence heterogeneity may explain the immunologic diversity seen in intact pili.

Haemophilus influenzae is a fastidious, gram-negative bacterium that is commonly found as a commensal organism in the human nasopharynx (28). *H. influenzae* is characterized as encapsulated (possessing one of six chemically and immunologically distinct polysaccharide capsules, i.e., types a to f) or nonencapsulated (i.e., nontypeable *H. influenzae*). Invasive infections, such as bacteremia, cellulitis, septic arthritis, and meningitis, occur in nonimmune hosts and are usually caused by organisms possessing the type b capsule. In children, immunocompromised individuals, and individuals with underlying pulmonary disease (e.g., cystic fibrosis, chronic bronchitis, and chronic obstructive pulmonary disease), *H. influenzae* can cause localized respiratory infections, such as otitis media, sinusitis, conjunctivitis, and pneumonia, and acute exacerbations of chronic lung diseases (16, 28, 30, 34).

Colonization of the upper respiratory tract is an essential step in the pathogenesis of *H. influenzae* disease and is a likely target for therapeutic intervention. Both typeable and nontypeable *H. influenzae* organisms have been shown to adhere to cultured epithelial cells and human nasopharyngeal tissues (33). One of the cell surface molecules shown to mediate attachment to epithelial cells is the polymeric hemagglutinating pilus found on both typeable and nontypeable *H. influenzae* (15).

Five genes (*hifA*, *hifB*, *hifC*, *hifD*, and *hifE*) are required for the synthesis of mature *H. influenzae* pili, and they are located on an approximately 6-kb chromosomal locus (15, 26, 40). *hifA* encodes the major pilin subunit and lies on one end of the pilus gene cluster (26, 40). The HifA pilin is approximately 24 kDa and comprises the primary structural component of the shaft of the mature pilus (9, 27, 35). The *hifA* pilin genes of 11 *H. influenzae* strains, including 5 type b strains and 6 nontypeable strains (including 2 *H. influenzae* biotype aegyptius strains), have been cloned and their nucleotide sequences have been determined in earlier studies by several investigators (3, 10, 12, 20, 22, 37, 39, 43).

Immunologic characterizations of intact *H. influenzae* pili and the HifA pilins have been complicated by the fact that intact pili are highly conformational and are immunologically diverse while denatured pilins are immunologically homogeneous (11, 13). Further, polyclonal antisera raised against native pili from type b strains Eagan and M43 bind to homologous piliated type b *H. influenzae* but do not bind to homologous denatured HifA pilins, suggesting that epitopes defined by these sera may be assembled by protein folding or by protein-protein interactions and are not available on denatured pilins (13). Similarily, polyclonal antisera raised against pilins of strains M43 and Eagan do not bind to intact pili of the homologous strains (11, 13).

The two goals of this work were (i) to identify differences in the HifA sequences from several different typeable and nontypeable *H. influenzae* isolates that might explain the pilus

* Corresponding author. Mailing address: Department of Pediatrics and Communicable Diseases, The University of Michigan, 109 S. Observatory St., SPH I/Rm. 2030, Ann Arbor, MI 48109-2029. Phone: (313) 647-3943. Fax: (313) 764-3192. E-mail: dclemans@sph.umich.edu.

immunologic heterogeniety and (ii) to identify sequence similarities that might relate to functional importance in bacterial pilus assembly. To do this analysis, the *hifA* genes from 26 typeable and 16 nontypeable strains were amplified by PCR and subjected to restriction fragment length polymorphism (RFLP) analysis with *Alu*I and *Rsa*I. Six different RFLP patterns were displayed with this analysis, and four more patterns were revealed from the nucleotide sequences of cloned *hifA* genes. Cloning and sequencing of representative *hifA* genes from each of the six RFLP patterns were performed and used for further analysis.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The *H. influenzae* strains used in this study are presented in Table 1. Except for strains AAr108 and AA61, the strains listed were isolated from individuals in a variety of geographical areas over a number of years and thus probably represent different bacterial clones. Strains AAr108 and AA61 were isolated from a mother and her son and may be the same strain. Bacterial strains designated AA and AAr were obtained from the clinical laboratories at the University of Michigan from 1983 to 1988, while strains designated M and Mr were obtained from the clinical laboratories at the University of Minnesota from 1979 to 1982. Bacterial strains were grown on Levinthal agar (37 g of brain heart infusion broth [Difco Laboratories, Detroit, Mich.], 18 g of Bacto agar [Difco], 2,000 μg of NAD [Sigma Chemical Co., St. Louis, Mo.], and 2,000 μg of hemin [Sigma] in 1,000 ml of deionized water) at 37°C with 5% CO_2 for 18 to 24 h (13). The *H. influenzae* strains were classified by using type-specific anticapsular antisera (for types a to f [Difco]) in a slide agglutination test.

Competent *Escherichia coli* DH5α (Gibco BRL, Gaithersburg, Md.) was grown in Luria-Bertani (LB) broth or on LB agar (Gibco BRL) at 37°C for transformation. Transformants were screened on LB agar containing 100 μg of ampicillin (Sigma) per ml and 40 μg of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal [Sigma]) per ml.

Isolation of genomic DNA from *H. influenzae*. Genomic DNA was isolated from *H. influenzae* either by a modification of the Marmur procedure (25, 42) or by the Wizard genomic DNA purification kit (Promega, Madison, Wis.).

Amplification of *hifA* from *H. influenzae* by PCR. PCR was used for the amplification of *hifA* from *H. influenzae* genomic DNA. Primers used were based on the conserved 5′ and 3′ regions of six *hifA* genes from strains Eagan, M43, AM30, 86-1249, 86-0295, and 81-0384 (3, 10, 12, 20, 22, 37, 39, 43). The primer sequences were derived from the 5′ and 3′ regions in the *hifA* gene that show significant nucleotide sequence identity among the six *H. influenzae* pilin sequences. The nucleotide sequences of these primers were 5′-ATGAAAAAAA CACT(AT)CTTGGTAGC-3′ and 5′-TTAT(CT)CGTAAGCAATT(GT)GGA AACC-3′. Fifty nanograms of *H. influenzae* genomic DNA was mixed with 20 pmol of each primer and 45 μl of PCR SuperMix (Gibco BRL) to a final volume of 50 μl and overlayed with mineral oil (Sigma). The final PCR mixture contained 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM (each) deoxynucleoside triphosphates (dATP, dCTP, dGTP, and dTTP), and 1 U of recombinant *Taq* DNA polymerase along with the *H. influenzae* genomic DNA and primers. The published error frequency for *Taq* DNA polymerase ranges from 1.1×10^{-4} errors/bp to 8.9×10^{-5} errors/bp (2, 38). The mixture was first incubated for 1 min at 95°C and then for 35 cycles of 95°C for 1 min, 50°C for 1 min, and 72°C for 2 min, followed by a final elongation step for 3 min at 72°C in a model PTC-100 programmable thermal controller (MJ Research, Inc., Watertown, Mass.). After amplification, samples were separated on 1% agarose gels and bands were visualized after staining with ethidium bromide (Sigma) and illumination by UV light. Molecular weight markers were run to estimate PCR fragment sizes (50- and 100-bp ladders [Gibco BRL]).

RFLP analysis of *hifA* PCR fragments. Amplified *hifA* PCR fragments were digested with restriction endonucleases *Alu*I and *Rsa*I according to the manufacturer's directions (Gibco BRL). Digestion products were resolved on 1% agarose gels and were visualized on a UV transilluminator after ethidium bromide staining. Molecular weight markers were run to estimate *Alu*I and *Rsa*I fragment sizes (50- and 100-bp ladders [Gibco BRL]). *H. influenzae* strains were grouped according to the *hifA Alu*I and *Rsa*I digestion patterns displayed after agarose gel electrophoresis.

Cloning of representative *hifA* genes. Two representative strains from each RFLP group were selected for cloning and sequencing of their *hifA* genes (Table 1). Amplified *hifA* fragments were electrophoresed on a preparative 1% agarose gel and purified with the GeneClean II kit (Bio 101, Inc., La Jolla, Calif.). The purified *hifA* PCR fragments were ligated into the *Srf*I site of pCR-Script Amp SK(+) after generation of blunt ends with a PCR Polishing Kit (Stratagene Cloning Systems, La Jolla, Calif.) and transformed into competent *E. coli* DH5α (Gibco BRL). Transformants were selected on LB agar containing 100 μg of ampicillin per ml and 40 μg of X-Gal per ml. Plasmid DNA from putative transformants was isolated with Qiagen Minipreps (Qiagen, Chatsworth, Calif.). *Bam*HI/*Not*I (Gibco BRL) double digests were used to confirm DNA inserts in the isolated recombinant plasmids.

Sequencing of cloned representative *hifA* PCR fragments. Cloned representative *hifA* genes were sequenced at the University of Michigan Medical School DNA Core Facility with an Applied Biosystems model 373A automated sequencer (Applied Biosystems, Inc., Foster City, Calif.). Sequencing primers were purchased from Stratagene (M13 −20 and reverse primers) and synthesized at the University of Michigan Medical School DNA Core Facility with the *hifA* DNA sequences from the representative cloned genes. DNA and protein sequences were analyzed with Lasergene Biocomputing software for the Macintosh from DNASTAR, Inc. (Madison, Wis.) and the Wisconsin Package, version 9.0, from the Genetics Computer Group (GCG) (Madison, Wis.) (5).

Nucleotide sequence accession numbers. The GenBank accession numbers for the *hifA* DNA sequences and the derived protein sequences determined in this study are as follows: AAr176, AF020908; AAr49, AF020909; 1712MEE (LB5), AF020910; ATCC 9007, AF020911; AAr73, AF020912; AAr160, AF020913; AA18, AF020914; AAr32, AF020915; ATCC 9006, AF020916; AAr157, AF020917.

RESULTS

PCR amplification and RFLP analysis of *hifA* from typeable and nontypeable *H. influenzae*. The *hifA* genes from representative typeable ($n = 26$) and nontypeable ($n = 16$) *H. influenzae* (Table 1) were amplified directly from genomic DNA. The resultant *hifA* PCR products were approximately 650 bp in length (data not shown). *H. influenzae* Rd, which lacks the *hif* gene cluster (8), was used as a negative control and did not yield PCR products with the *hifA* primers.

RFLP analysis was performed on the amplified *hifA* fragments with restriction endonucleases *Alu*I and *Rsa*I. Six different RFLP patterns were displayed with each enzyme by this analysis (groups 1 to 6 [Fig. 1 and Table 1]). Pattern 2 contained all nontypeable isolates ($n = 4$), while pattern 5 contained predominantly type f isolates (7 of 8 isolates). Both *H. influenzae* biotype aegyptius *hifA* sequences displayed the same RFLP patterns (pattern 10 [Table 1]). In considering a possible association of capsular type (types b and f and nontypeable strains) with *hifA* RFLP patterns (patterns 1, 3, and 5), capsular type f was significantly associated with RFLP pattern 5 ($\chi^2 = 22.339$; $P \leq 0.001$); no association was seen between the other capsular types and RFLP patterns ($\chi^2 = 0.02$; $0.80 < P \leq 0.90$).

Eleven different *hifA* genes have been cloned and sequenced previously and are included in Table 1. Five of these *hifA* genes fall into RFLP pattern 3 [M43, AM30 (770235), MinnA, AO2, and 86-1249 (LKP4) (Table 1)], while one falls into pattern 1 [Eagan (E1a; LKP3) (Table 1)]. The *hifA* genes from *H. influenzae* biogroup aegyptius strains F3031 and F3037 display the same RFLP pattern (pattern 10) and are distinct from the six RFLP groups defined in this study (Table 1). The three remaining *hifA* sequences each display new RFLP patterns that are different from the six RFLP patterns identified in this study (patterns 7 to 9 [Table 1]).

Correlation between the *hifA* RFLP grouping of strains and immunoreactivity with antipilus polyclonal antisera. Gilsdorf et al. (13) have shown that polyclonal antisera raised against intact pili from *H. influenzae* type b strains Eagan and M43 each reacted with a different subset of 22 piliated type b isolates. Table 2 presents representative type b strains reactive with each of the antipilus sera and representative nontypeable strains that do not react with either antipilus serum (13). Those type b strains that demonstrate greater reactivity with antiserum directed against strain M43 pili (i.e., 4^+) tend to fall into RFLP pattern 3, while those type b strains that react equally (i.e., 2^+) or greater (2^+ to 4^+) with antiserum directed against strain Eagan pili tend to fall into RFLP pattern 1 ($\chi^2 = 6.875$; $0.02 < P \leq 0.05$). Strain AAr103p$^+$ is the exception, in that it

TABLE 1. *hifA* RFLP comparison of various *H. influenzae* clinical isolates by using *Alu*I and *Rsa*I

hifA RFLP pattern[a]	Strain[b]	Source and/or reference[c]	Capsular type
1	**Eagan (E1a)**	CSF (10)	b
	AAr108	Throat (13)	b
	AA61	Blood (13)	b
	R9	(13)	b
	AAr120	Trachea (this study)	f
	AAr39	Throat (11)	NT[f]
	AAr176[d]	Nasopharynx (11)	NT
	AAr64	Trachea (13)	b (lost capsule)
	AAr169	Sinus (11)	NT
	AAr117	Nasopharynx (this study)	b (lost capsule)
2	Mr31	Middle ear (11)	NT
	AAr49[d]	Throat (11)	NT
	1712MEE (LB5)[d]	Middle ear (11)	NT
	AAr91	Throat (11)	NT
3	M96	CSF (this study)	a
	M43[e]	Throat (12)	b
	AM30 (770235)	CSF (39)	b
	MinnA	CSF (3)	b
	AO2	Throat (22)	b
	AAr122	Nasopharynx (13)	b
	SL2	Throat (13)	b
	AAr7	Nasopharynx (23)	b
	AAr103	Eye (23)	b
	AAr119	Nasopharynx (13)	b
	ATCC 9007[d]	ATCC	c
	AAr45	Nasopharynx (11)	NT
	AAr154	Nasopharynx (11)	NT
	AAr60	Peritonsillar abscess (this study)	NT
	86-1249	Middle ear (1)	NT
4	ATCC 9008	Throat (ATCC)	d
	AAr191	Throat (this study)	e
	AAr73[d]	Nasopharynx (11)	NT
	86-042NP (LB1)	Nasopharynx (11)	NT
	1128MEE (LB2)	Middle ear (11)	NT
	AAr160[d]	Trachea (11)	NT
5	AAr107	Nasopharynx (this study)	e
	AA18[d]	Blood (this study)	f
	AAr9	Throat (this study)	f
	AAr32[d]	Throat (this study)	f
	AAr80	Nasopharynx (this study)	f
	AA12	Blood (this study)	f
	AA58	Blood (this study)	f
	AA195	Blood (this study)	f
6	ATCC 9006[d]	ATCC	a
	AAr157[d]	Nasopharynx (11)	NT
	AAr180	Sputum (11)	NT
7	**86-0295**	Middle ear (1)	NT
8	**81-0384**	Middle ear (1)	NT
9	**M37**	Nasopharynx (3)	NT
10	**F3037**	Blood (37)	NT
	F3031	Blood (43)	NT

[a] *hifA* RFLP fragment sizes (in base pairs) based on DNA sequences. Pattern 1: *Alu*I, 374, 213, 41, 23; *Rsa*I, 411, 117, 86, 37. Pattern 2: *Alu*I, 288, 243, 41, 29, 23; *Rsa*I, 243, 189, 155, 37. Pattern 3: *Alu*I, 365, 159, 95, 23; *Rsa*I, 264, 144, 111, 68, 37, 18. Pattern 4: *Alu*I, 320, 112, 80, 48, 41, 23; *Rsa*I, 501, 86, 37. Pattern 5: *Alu*I, 314, 235, 55, 23, 10; *Rsa*I, 263, 254, 83, 37. Pattern 6: *Alu*I, 262, 146, 100, 54, 41, 23; *Rsa*I, 589, 37. Pattern 7: *Alu*I, 272, 127, 98, 60, 47, 23, 12; *Rsa*I, 339, 228, 72. Pattern 8: *Alu*I, 613, 23; *Rsa*I, 264, 228, 84, 48, 12. Pattern 9: *Alu*I, 380, 236, 23; *Rsa*I, 335, 267, 37. Pattern 10: *Alu*I, 284, 165, 164, 23; *Rsa*I, none.

[b] Strains with published *hifA* sequences are in boldface type. Database accession numbers or references are as follows: Eagan (E1a), M64334; M43, reference 12; AO2, X52419; AM30 (770235), X16991; 86-1249, U19761; 86-0295, U19730; 81-0384, U19795; M37 and MinnA, reference 3; F3037, S59288; F3031, L08606.

[c] CSF, cerebrospinal fluid; ATCC, American Type Culture Collection.

[d] H. influenzae strains whose *hifA* genes were sequenced in this study.

[e] The *hifA* sequences from strains M43, AM30 (770235), and MinnA are identical and are treated as one sequence throughout this article.

[f] NT, nontypeable *H. influenzae.*

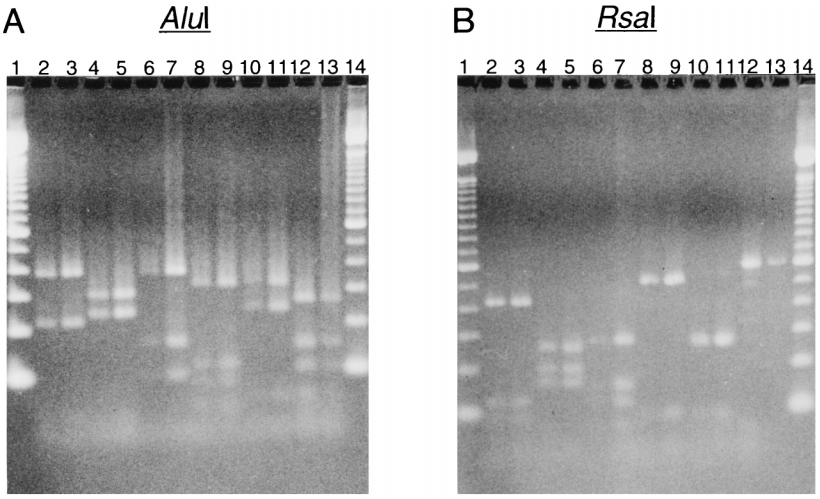

FIG. 1. *Alu*I and *Rsa*I RFLP analysis of PCR amplified *hifA* from genomic DNAs of representative *H. influenzae* strains. (A) *hifA* samples digested with *Alu*I. (B) *hifA* samples digested with *Rsa*I. Lanes: 1 and 14, 100-bp ladder; 2, Eagan (E1a; LKP3); 3, AAr176; 4, 1712MEE (LB5); 5, AAr49; 6, M43; 7, ATCC 9007; 8, AAr73; 9, AAr160; 10, AAr32; 11, AA18; 12, ATCC 9006; 13, AAr157.

shows greater reactivity with anti-Eagan pilus serum and yet falls into RFLP pattern 3 (Table 2).

Table 2 also includes four nontypeable *H. influenzae* strains that display either *hifA* RFLP pattern 1 or pattern 3; these strains demonstrated no reactivity with either antipilus serum (11). Thus, *H. influenzae* type b strains reacted with either antipilus serum while nontypeable *H. influenzae* strains reacted with neither serum, irrespective of *hifA* RFLP type (11, 13, 23).

Sequence analysis of representative *hifA* genes. In order to confirm the validity of the RFLP analysis and to explore sequence differences between each RFLP group, we chose 10 different representative *hifA* genes to clone and sequence from the *H. influenzae* strains in Table 1 to complement the existing *hifA* sequence database. The strains chosen were from several different sources and represent the six RFLP patterns defined in this study.

Analysis of the derived HifA amino acid sequences (Fig. 2) revealed that all 19 representative pilins have a highly conserved 18- to 20-amino-acid leader sequence and contain such pilin signatures as strong C-terminal amino acid homologies, conserved tyrosines and glycines at 2 and 14 residues, respectively, from the C terminus, and a similarly spaced pair of cysteine residues at positions 45 and 85 (Fig. 2). Further analysis of the C-terminal amino acids showed a range of 60 to 100% identity in the terminal 16 residues, with 8 of the 16 residues being absolutely conserved (Fig. 2). These pilin signature sequences are shared among a wide variety of bacterial pilus proteins that are assembled by periplasmic chaperones (17, 21). Recent studies by St. Geme III et al. (36) demonstrated that the biogenesis of *H. influenzae* pili is dependent upon the periplasmic chaperone HifB, which belongs to the PapD family of immunoglobulin-like chaperones (17, 21). Several regions of amino acid identity which are distributed throughout the HifA sequence are evident in the sequence comparison (e.g., residues 29 to 85, 99 to 120, 127 to 136, 145 to 155, 176 to 178, and 191 to 200 [Fig. 2]).

The validity of the RFLP groups for identifying like *hifA* genes is confirmed by comparisons of the amino acid sequence identities of the derived, representative HifA pilins (Table 3). The amino acid identities were stronger within RFLP groups than between groups ($P < 0.0001$ [by two-way analysis of variance). For example, pairwise comparison of HifA sequences within a specific RFLP group yielded between 87 and 100% identities, whereas pairwise comparisons of HifA sequences between members of different RFLP groups showed between 59 and 81% identities. The HifA amino acid sequences of RFLP groups 1 and 3 have between 78 and 81% identities.

A dendrogram (Fig. 3) depicting the groupings of strains based upon the complete pileup of derived amino acid se-

TABLE 2. Reactivity of *H. influenzae* strains with antipilus antisera in a dot blot assay[a]

hifA RFLP pattern	Strain (capsular type)[b]	Hemagglutination titer	Reactivity with antiserum[c]	
			R1 (anti-M43 pili)	R19 (anti-Eagan pili)
1	E1ap$^+$ (b)	1:16	0	4$^+$
	E1ap$^-$ (b)[d]	0		0
	R9p$^+$ (b)	1:16	3$^+$	4$^+$
	AA61p$^+$ (b)	1:16	2$^+$	2$^+$
	AAr108p$^+$ (b)	1:16	0	2$^+$
	AAr176 (NT)	1:16	0	0
	AAr39 (NT)	1:16	0	0
	AAr169 (NT)	1:16	0	0
3	M43p$^+$ (b)	1:16	4$^+$	1$^+$
	M42p$^-$ (b)[e]	0	0	0
	AAr122p$^+$ (b)	1:32	4$^+$	0
	AAr7p$^+$ (b)	1:16	4$^+$	2$^+$
	AAr119p$^+$ (b)	1:32	4$^+$	0
	AAr103p$^+$ (b)	1:32	2$^+$	4$^+$
	SL2p$^+$ (b)	1:32	4$^+$	0
	AAr45 (NT)	1:64	0	0

[a] Table generated with data from Gilsdorf et al. (11, 13).

[b] An enrichment technique that depends upon erythrocyte agglutination by piliated *H. influenzae* type b strains was used to select piliated variants from the clinical isolates. NT, nontypeable *H. influenzae.*

[c] The reactivities of the antibodies with native pili on the surfaces of intact bacteria were assessed visually with a dot blot assay and compared with reactivities of positive (homologous strains) and negative (nonpiliated variants) controls included in each assay by using a grading scale ranging from 0 (identical to negative controls) to 4$^+$ (identical to homologous strains).

[d] Nonpiliated phase variant of E1ap$^+$.

[e] Nonpiliated variant of strain M43p$^+$; used as a negative control.

FIG. 2. Comparison of predicted amino acid sequences of HifA from representative *H. influenzae* strains. Identical residues throughout all HifA sequences are in boldface type and are shown in the bottom line as "Consensus." Arrowheads denote the conserved cysteine residues; asterisks denote the positions of the conserved tyrosine and glycine residues 2 and 14 amino acids, respectively, from the C terminus. Hydrophilic regions I, II, and III are underlined (10), and pilin motifs "Segment S3" and "Segment S6" (17) are in shaded boxes. The comparison was performed with the Pileup program of the Wisconsin Package, version 9.0, from GCG (5).

quences (Fig. 2) further confirms the groupings based on *hifA* RFLP patterns and shows the relationships between these groups.

DISCUSSION

The goal of this work was to compare the *hifA* genes from several independent typeable and nontypeable *H. influenzae* isolates and identify any sequence differences that might explain the pilus immunologic heterogeneity. Further, these *hifA* genes were compared to pilins from other bacteria to identify conserved regions potentially important for pilus assembly.

The relationships of *H. influenzae* strains based upon *hifA* RFLP analysis (Table 1; Fig. 1) were confirmed by analysis of the derived amino acid sequences (Fig. 2 and 3; Table 3). All of the strains contain highly conserved 18- to 20-residue signal sequences (Fig. 2) and several regions of high sequence identity. Of note are the equally spaced pairs of cysteine residues at positions 45 and 85 in all of the HifA pilin sequences (Fig. 2). This Cys-Cys loop is conserved among pilins from several different bacteria, is postulated to play a role in the maintenance of protein structure, and is thought to be a dominant immunogenic epitope in the PapA pilins of uropathogenic *E. coli* (4, 17). The amino acids contained within the region of the Cys-Cys loop (residues 58 to 121 [Fig. 2]) of *H. influenzae* M37 were demonstrated by Palmer and Munson, Jr. (31), to possess a significant part of the epitope defined by the pilin-specific monoclonal antibody 3H12, emphasizing the immunogenic potential of this region in *H. influenzae* pilins.

The region in HifA from amino acids 156 to 205, which is the most variable region within the pilin sequences, is analogous to the variable region found in the PapA pilins of uropathogenic *E. coli* and the pilins from other bacteria and may account for the immunologic diversity in *H. influenzae* pili (4, 17). For PapA, strain-to-strain differences in the variable region and the Cys-Cys loop are thought to constitute the basis for the serological diversity of these pili (4).

In an effort to identify common regions of type b pili that are surface exposed and represent antigenic epitopes, Forney et al. (10) analyzed the hydrophilicity of the pilin proteins expressed by the type b strains M43 and Eagan. They identified three hydrophilic regions within the HifA sequence (regions I, II, and III [Fig. 2]) and proposed that these regions might constitute conserved antigenic epitopes. The present study shows that these three regions are highly conserved within all HifA sequences (Fig. 2), with 10 of 19 residues being absolutely conserved in region I. Further, 10 of 18 and 4 of 12 residues are absolutely conserved in regions II and III, respectively, in the HifA comparison (Fig. 2). Along with the absolutely conserved amino acids in each region, several conserved amino acid substitutions result in high degrees of sequence similarity in these regions. To determine if these regions contain surface-exposed, immunogenic epitopes, Gilsdorf et al. (14) constructed 14- to 15-amino-acid peptides corresponding to regions I, II, and III and raised polyclonal rabbit antisera to these peptides. Sera to these peptides demonstrated poor to no reactivity to native pili and moderate to strong reactivity to denatured pili, suggesting that the epitopes determined by these peptides are not present on assembled pili in a conformation that can be recognized by the antipeptide antibodies (14). These results were supported by the previous observation that antibodies raised against denatured pilin and an internal peptide of strain M43 HifA recognized epitopes on denatured pilins of both type b and nontypeable *H. influenzae* better than native pili on the same strains (11, 13). Therefore, although these regions are

TABLE 3. Amino acid identities of HifA based on DNA sequences from various strains confirm relationships defined by RFLP analysis[a]

hifA RFLP pattern	Strain	% Amino acid identity[b]																		
		Pattern 1		Pattern 2		Pattern 3				Pattern 4		Pattern 5		Pattern 6		Pattern 7	Pattern 8	Pattern 9	Pattern 10	
		Eagan	AAr176	AAr49	LB5	AO2	M43[c]	ATCC 9007	86-1249	AAr73	AAr160	AA18	AAr32	ATCC 9006	AAr157	86-0295	81-0384	M37	F3031	F3037
1	Eagan		**99**	69	69	79	79	81	79	68	68	65	65	67	69	60	62	65	75	75
	AAr176			69	68	78	78	80	78	68	68	65	65	67	69	59	62	64	74	74
2	AAr49				**99**	71	72	71	72	68	68	77	77	69	67	67	62	64	71	71
	LB5					71	72	71	72	67	67	76	76	69	67	67	62	64	72	72
	AO2						**99**	**87**	**99**	68	68	66	66	69	70	66	63	69	72	72
3	M43[c]							**88**	**99**	69	69	66	66	69	70	66	64	69	73	73
	ATCC 9007								**87**	70	70	68	68	72	72	62	62	68	73	73
	86-1249									69	69	66	66	69	70	66	64	69	73	73
4	AAr73										**98**	74	74	80	70	63	71	64	70	70
	AAr160											74	74	80	70	63	71	64	70	70
5	AA18												**100**	77	71	64	63	65	68	68
	AAr32													77	71	64	63	65	68	68
6	ATCC 9006														**89**	64	66	73	73	73
	AAr157															63	66	74	72	72
7	86-0295																62	65	66	66
8	81-0384																	66	66	66
9	M37																		74	74
10	F3031																			**99**
	F3037																			

[a] Amino acid sequence comparisons were performed with the BestFit program of the Wisconsin Package, version 9.0, from GCG (5).
[b] Boldface type indicates percent amino acid identities between different strains within the same RFLP pattern.
[c] See Table 1, footnote e.

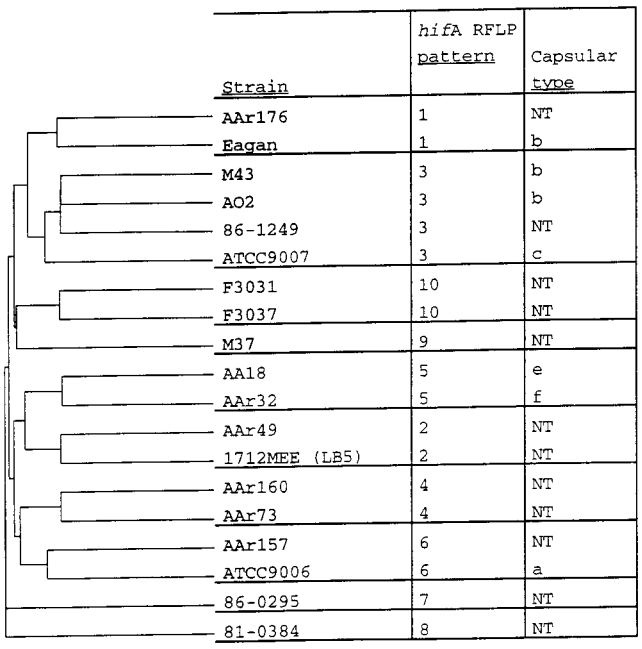

FIG. 3. Dendrogram showing the relationships of the predicted amino acid sequences of HifA from the 19 representative *H. influenzae* strains. The dendrogram was generated with the Pileup program of the Wisconsin Package, version 9.0, from GCG (5). NT, nontypeable *H. influenzae.*

highly conserved on denatured pilins, they are not available for antipeptide or antipilin antibody binding on native pili.

Several conserved features characteristic of pilus proteins assembled by *E. coli* PapD-like molecular chaperones are seen in the C-terminal sequences of the HifA pilins (Fig. 2). For example, the derived HifA sequences have strong amino acid homology to one another in the C terminus (Fig. 2) and contain absolutely conserved tyrosines and glycines at 2 and 14 residues, respectively, from the C terminus (3, 10, 19, 21, 22, 37, 39, 43). These conserved HifA sequence features are shared with the minor pilin HifD and the C terminus of the putative adhesin, HifE (26, 40). Recently, HifB-HifA and HifB-HifD chaperone-pilin complexes have been isolated, demonstrating that the biogenesis of *H. influenzae* pili is dependent upon the periplasmic chaperone HifB (36).

Recently, Girardeau and Bertin (17), using two-dimensional sequence analysis, described other conserved markers of the bacterial pilin family; these features are conserved within the representative HifA sequences (Fig. 2). Among the motifs identified, segments S3 (F*x*l*x*L*xx*C [where *x* is any residue]) and S6 (A*x*[G/N]VGVQ*i* [where *i* is a hydrophobic residue]) were the most conserved. Girardeau and Bertin (17) suggest that the S3 and S6 motifs, along with the conserved Cys-Cys loop and C-terminal homology, play a role in the function or maintenance of the structural integrity of the protein.

The intact pili of *H. influenzae* demonstrate a high degree of immunologic heterogeneity with both polyclonal and monoclonal antipilus sera (1, 11, 13, 23, 31). Brinton et al. (1) used polyclonal anti-LKP pilus sera to differentiate clinical isolates of *H. influenzae* into seven different LKP pilus types (LKP1 to LKP7). Four LKP serotypes are represented in this study [LKP3, Eagan (E1a); LKP4, 86-1249; LKP1, 86-0295; LKP5, 81-0384] and each displays a different RFLP pattern (Table 1).

Polyclonal antipilus sera raised against the pili of type b strains Eagan and M43 each reacted with a different subset of 22 type b *H. influenzae* strains (13, 23). Those type b strains showing equal or greater reactivity with Eagan antipilus serum displayed *hifA* RFLP pattern 1, while those strains showing greater reactivity with M43 antipilus serum displayed *hifA* RFLP pattern 3 (Table 2). Several strains, though, show reactivity with both antipilus sera, suggesting that epitopes defined by these antisera are shared by some strains and not others. These findings support those of Denich et al. (4), who found common immunogenic domains among PapA pilins from different strains of uropathogenic *E. coli.*

The amino acid sequences of representative HifA pilins from RFLP patterns 1 and 3 range between 78 and 81% identity (Table 3). Further, strain AAr103p$^+$ reacts more strongly with Eagan antipilus serum yet displays *hifA* RFLP pattern 3, to which strain M43 belongs (Table 2). This result demonstrates that although the HifA pilin sequence of strain AAr103p$^+$ has an overall amino acid sequence identity closer to that of M43, it shares certain critical antigenic residues with that of Eagan. Few amino acid differences between HifA pilins could explain the varying reactivities with antipilus antisera. For example, *H. influenzae* AAr176 bacteria do not react with Eagan antipilus serum (Table 2) (11). This serum appears to be specific for the Eagan HifA pilin and decorates the entire pilus shaft on whole bacteria subjected to immunoelectron microscopy (9). The derived HifA sequences from strains Eagan and AAr176 are 99% identical and differ by only three amino acids (residues 43, 217, and 221) (Table 3; Fig. 2).

The first amino acid difference (residue 43 [Fig. 2]) lies within the predicted hydrophilic region I originally identified by Forney et al. (10) and near a pair of conserved cysteines that define the Cys-Cys loop that may represent a major surface-exposed antigenic region in HifA. The second pair of amino acid differences between the Eagan and AAr176 HifA sequences (residues 217 and 221) are at the C terminus, a region that tends to be conserved among pilins and is thought to play a role in pilus subunit interactions (17, 21, 26). However, Palmer and Munson, Jr. (31), observed that monoclonal antibody 3H12 reactivity with M37 HifA was enhanced by the addition of M37 C-terminal sequences to the M37-MinnA HifA chimeras, suggesting that this region may itself be antigenic or that amino acids in this region contribute to nonlinearly assembled epitopes.

The rationale for the immunological heterogeneity of *H. influenzae* pili is not well understood. Limited studies have shown that humans can produce serum antibodies directed against *H. influenzae* pili (7, 32). Further, *H. influenzae* has been shown to undergo pilus phase variation and antigenic variation in such cell surface molecules as major outer membrane protein P2, immunoglobulin A1 proteases, and lipopolysaccharide (6, 18, 24, 29, 41). Taken together, the antigenic diversity of *H. influenzae* pili may be due to small changes in immunodominant surface-exposed epitopes in HifA and may play a role, along with phase variation and antigenic drift of other surface molecules, in the organism's ability to evade the host immune system.

ACKNOWLEDGMENTS

We thank Gregory Russell for designing the *hifA* primers used in PCR.

This work was supported by Public Health Service grant AI25630 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

1. **Brinton, C. C., M. J. Carter, D. B. Derber, S. Kar, J. A. Kramarik, A. C.-C. To, S. C.-M. To, and S. W. Wood.** 1989. Design and development of pilus vaccines for *Haemophilus influenzae* diseases. Pediatr. Infect. Dis. J. **8:**S54–S61.

2. **Cariello, N. F., J. A. Swenberg, and T. R. Skopek.** 1991. Fidelity of *Thermococcus litoralis* DNA polymerase (Vent) in PCR determined by denaturing gradient gel electrophoresis. Nucleic Acids Res. **19:**4193–4198.
3. **Coleman, T., S. Grass, and R. Munson, Jr.** 1991. Molecular cloning, expression, and sequence of the pilin gene from nontypeable *Haemophilus influenzae* M37. Infect. Immun. **59:**1716–1722.
4. **Denich, K., L. B. Blyn, A. Craiu, B. A. Braaten, J. Hardy, D. A. Low, and P. D. O'Hanley.** 1991. DNA sequences of three *papA* genes from uropathogenic *Escherichia coli* strains: evidence of structural and serological conservation. Infect. Immun. **59:**3849–3858.
5. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
6. **Duim, B., L. Vogel, W. Puijk, H. M. Jansen, R. H. Meloen, J. Dankert, and L. van Alphen.** 1996. Fine mapping of outer membrane protein P2 antigenic sites which vary during persistent infection by *Haemophilus influenzae*. Infect. Immun. **64:**4673–4679.
7. **Erwin, A. L., G. E. Kenny, A. L. Smith, and T. L. Stull.** 1988. Human antibody response to outer membrane proteins and fimbriae of *Haemophilus influenzae* type b. Can. J. Microbiol. **34:**723–729.
8. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L.-I. 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. Fritch, 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.
9. **Forney, L. J., J. R. Gilsdorf, and D. C. L. Wong.** 1992. Effect of pili-specific antibodies on the adherence of *Haemophilus influenzae* type b to human buccal cells. J. Infect. Dis. **165:**464–470.
10. **Forney, L. J., C. F. Marrs, S. L. Bektesh, and J. R. Gilsdorf.** 1991. Comparison and analysis of the nucleotide sequences of pilin genes from *Haemophilus influenzae* type b strains Eagan and M43. Infect. Immun. **59:**1991–1996.
11. **Gilsdorf, J. R., H. Chang, K. W. McCrea, and L. Bakaletz.** 1992. Comparison of hemagglutinating pili of *Haemophilus influenzae* type b with similar structures of nontypeable *H. influenzae*. Infect. Immun. **60:**374–379.
12. **Gilsdorf, J. R., C. F. Marrs, K. W. McCrea, and L. J. Forney.** 1990. Cloning, expression, and sequence analysis of the *Haemophilus influenzae* type b strain M43p$^+$ pilin gene. Infect. Immun. **58:**1065–1072.
13. **Gilsdorf, J. R., K. W. McCrea, and L. J. Forney.** 1990. Conserved and nonconserved epitopes among *Haemophilus influenzae* type b pili. Infect. Immun. **58:**2252–2257.
14. **Gilsdorf, J. R., L. J. Forney, and K. W. McCrea.** 1993. Reactivity of antibodies against conserved regions of pilins of *Haemophilus influenzae* type b. J. Infect. Dis. **167:**962–965.
15. **Gilsdorf, J. R., K. W. McCrea, and C. F. Marrs.** 1997. Role of pili in *Haemophilus influenzae* adherence and colonization. Infect. Immun. **65:** 2997–3002.
16. **Gilsdorf, J. R.** 1987. *Haemophilus influenzae* non-type b infections in children. Am. J. Dis. Child. **141:**1063–1065.
17. **Girardeau, J.-P., and Y. Bertin.** 1995. Pilins of fimbrial adhesins of different member species of enterobacteriaceae are structurally similar to the C-terminal half of adhesin proteins. FEBS Lett. **357:**103–108.
18. **Groeneveld, K., L. van Alphen, C. Voorter, P. P. Eijk, H. M. Jansen, and H. C. Zanen.** 1989. Antigenic drift of *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease. Infect. Immun. **57:**3038–3044.
19. **Hultgren, S. J., and C. H. Jones.** 1995. Utility of the immunoglobulin-like fold of chaperones in shaping organelles of attachment in pathogenic bacteria. ASM News **61:**457–464.
20. **Kar, S., S. C.-M. To, and C. C. Brinton, Jr.** 1990. Cloning and expression in *Escherichia coli* of LKP pilus genes from a nontypeable *Haemophilus influenzae* strain. Infect. Immun. **58:**903–908.
21. **Kuehn, M. J., D. J. Ogg, J. Kihlberg, L. N. Slonim, K. Flemmer, T. Bergfors, and S. J. Hultgren.** 1993. Structural basis of pilus subunit recognition by the PapD chaperone. Science **262:**1234–1241.
22. **Langermann, S., and A. Wright.** 1990. Molecular analysis of the *Haemophilus influenzae* type b pilin gene. Mol. Microbiol. **4:**221–230.
23. **LiPuma, J. J., and J. R. Gilsdorf.** 1988. Structural and serological relatedness of *Haemophilus influenzae* type b pili. Infect. Immun. **56:**1051–1056.
24. **Lomholt, H., L. van Alphen, and M. Kilian.** 1993. Antigenic variation of immunoglobulin A1 proteases among sequential isolates of *Haemophilus influenzae* from healthy children and patients with chronic obstructive pulmonary disease. Infect. Immun. **61:**4575–4581.
25. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. **3:**208–218.
26. **McCrea K. W., W. J. Watson, J. R. Gilsdorf, and C. F. Marrs.** 1994. Identification of *hifD* and *hifE* in the pilus gene cluster of *Haemophilus influenzae* type b strain Eagan. Infect. Immun. **62:**4922–4928.
27. **McCrea, K. W., W. J. Watson, J. R. Gilsdorf, and C. F. Marrs.** 1997. Identification of two minor subunits in the pilus of *Haemophilus influenzae*. J. Bacteriol. **179:**4227–4231.
28. **Moxon, E. R.** 1986. The carrier state: *Haemophilus influenzae*. J. Antimicrob. Chemother. **18**(Suppl A)**:**17–24.
29. **Moxon, E. R., P. B. Rainey, M. A. Nowak, and R. E. Lenski.** 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr. Biol. **4:**24–33.
30. **Murphy, T., and S. Sethi.** 1992. Bacterial infection in chronic obstructive pulmonary disease. Am. Rev. Respir. Dis. **146:**1067–1083.
31. **Palmer, K. L., and R. S. Munson, Jr.** 1992. Construction of chimaeric genes for mapping a surface-exposed epitope on the pilus of non-typeable *Haemophilus influenzae*. Mol. Microbiol. **6:**2583–2588.
32. **Pichichero, M. E., P. Anderson, M. Loeb, and D. H. Smith.** 1982. Do pili play a role in pathogenicity of *Haemophilus influenzae* type b? Lancet **ii:**960–962.
33. **St. Geme, III, J. W.** 1996. Molecular determinants of the interaction between *Haemophilus influenzae* and human cells. Am. J. Respir. Crit. Care Med. **154:**S192–S196.
34. **St. Geme, III, J. W.** 1993. Nontypeable *Haemophilus influenzae* disease: epidemiology, pathogenesis, and prospects for prevention. Infect. Agents Dis. **2:**1–16.
35. **St. Geme, III, J. W., D. Cutter, and S. J. Barenkamp.** 1996. Characterization of the genetic locus encoding *Haemophilus influenzae* type b surface fibrils. J. Bacteriol. **178:**6281–6287.
36. **St. Geme, III, J. W., J. S. Pinkner, G. P. Krasan, J. Heuser, E. Bullitt, A. L. Smith, and S. J. Hultgren.** 1996. *Haemophilus influenzae* pili are composite structures assembled via the HifB chaperone. Proc. Natl. Acad. Sci. USA **93:**11913–11918.
37. **St. Geme, III, J. W., and S. Falkow.** 1993. Isolation, expression, and nucleotide sequencing of the pilin structural gene of the Brazilian purpuric fever clone of *Haemophilus influenzae* biogroup aegyptius. Infect. Immun. **61:** 2233–2237.
38. **Tindall, K. R., and T. A. Kunkel.** 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry **27:**6008–6013.
39. **van Ham, S. M., F. R. Mooi, M. G. Sindhunata, W. R. Maris, and L. van Alphen.** 1989. Cloning and expression in *Escherichia coli* of *Haemophilus influenzae* fimbrial genes establishes adherence to oropharyngeal epithelial cells. EMBO J. **8:**3535–3540.
40. **van Ham, S. M., L. van Alphen, F. R. Mooi, and J. P. M. van Putten.** 1994. The fimbrial gene cluster of *Haemophilus influenzae* type b. Mol. Microbiol. **13:**673–684.
41. **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.
42. **Watson, W. J., J. R. Gilsdorf, M. A. Tucci, K. W. McCrea, L. J. Forney, and C. F. Marrs.** 1994. Identification of a gene essential for piliation in *Haemophilus influenzae* type b with homology to the pilus assembly platform genes of gram-negative bacteria. Infect. Immun. **62:**468–475.
43. **Whitney, A. M., and M. M. Farley.** 1993. Cloning and sequence analysis of the structural pilin gene of Brazilian purpuric fever-associated *Haemophilus influenzae* biogroup aegyptius. Infect. Immun. **61:**1559–1562.

Editor: P. E. Orndorff